

高等学校计算机硬件技术课程系列教材

计算机硬件技术基础

袁正定 主编

YF102

高等教育出版社

内容提要

本书是在高等教育出版社2001年12月出版的《计算机硬件技术基础》的基础上改编而成。

本书由浅入深地介绍了计算机的组成原理,对组成计算机的各个部件的原理、性能指标和标准进行了详细介绍。在各部分的内容组织上紧跟计算机技术的最新发展,涵盖了计算机主机系统和各种外部设备,包括CPU、存储器、总线、芯片组、接口、外存储器、基本输入/输出设备、显示系统、音频系统、通信设备、工业控制计算机等。

本书由具有丰富教学经验的一线教师编写,内容新颖、概念清晰,配有大量的习题,适合作为高等学校非计算机专业计算机硬件技术课程的教材,也可以作为计算机爱好者的自学用书。

图书在版编目(CIP)数据

计算机硬件技术基础/裴正定主编. —北京:高等教育出版社,2007.8

ISBN 978-7-04-021957-9

I. 计… II. 裴… III. 硬件-高等学校-教材

IV. TP303

中国版本图书馆CIP数据核字(2007)第096836号

策划编辑 孙志丽 责任编辑 郭福生 封面设计 于文燕 责任绘图 尹莉
版式设计 陆瑞红 责任校对 胡晓琪 责任印制 陈伟光

出版发行 高等教育出版社 购书热线 010-58581118
社址 北京市西城区德外大街4号 免费咨询 800-810-0598
邮政编码 100011 网 址 <http://www.hep.edu.cn>
总机 010-58581000 <http://www.hep.com.cn>
网上订购 <http://www.landrao.com>
经销 蓝色畅想图书发行有限公司 <http://www.landrao.com.cn>
印刷 涿州市京南印刷厂 <http://www.widedu.com>
畅想教育

开本 787×1092 1/16 版次 2007年8月第1版
印张 25 印次 2007年8月第1次印刷
字数 610 000 定 价 31.10 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 21957-00

前 言

“计算机硬件技术基础”是计算机技术基础教育中一门重要的课程。为了促进计算机教学在大学教育中的规范化,普及计算机硬件知识,针对大学各个专业计算机硬件教学实际编写了本教材。

计算机硬件的教学,在过去几十年主要是面向计算机专业的学生。这方面的课程主要包括计算机组成原理、接口技术、汇编语言等。对于非计算机专业大学本科生的计算机硬件教学,在内容和深度方面尚未形成一个统一的标准。和计算机专业相比,非计算机专业的计算机硬件课程在教学内容和方法上需要做出重大调整。在本教材的编写过程中主要考虑了以下问题:

- 授课对象为各专业本科生,包括理科和文科类专业的学生。
- 对学生没有计算机专业前修课程的要求。
- 侧重在计算应用过程中的硬件相关技术。
- 侧重介绍标准以及标准的发展过程。
- 重点介绍当前最新的硬件新技术和新标准。

本书可以分为三部分,第一部分包括第1、2章,内容是计算机概述和原理;第二部分包括第3~8章,是本书的主要特色内容;第三部分包括第9~11章,是本书的扩展知识部分。全书共需要授课课时40学时。如果是通常的32学时,建议讲授第1~8章。

本课程近10年的教学实践证明,本教材的内容编排是比较合理的,在教学的过程中也达到了预期的效果。

本书是在高等教育出版社2001年12月出版的《计算机硬件技术基础》的基础上改编而成。

为了增加课程的实践性,第3~7章除了习题之外,还增加了实验部分。这些实验都是比较简单易实现的。

本书由裴正定教授主编。具体分工如下:第1、11章由裴正定教授编写,第2~4章由朱卫东编写,第5~7章由周洪利编写,第8~10章由陈连坤编写。

国防科技大学邹逢兴教授审阅了全书,并提出了许多宝贵的意见,在此表示衷心的感谢!

由于编者水平和知识的局限性以及时间的仓促,书中难免存在错误及不妥之处,敬请广大读者批评指正,提出宝贵意见。

编 者

2007年3月

3.3.6 超线程技术	66
3.3.7 多核心技术	67
3.4 CPU 的封装与接口类型	69
3.4.1 CPU 的封装	69
3.4.2 CPU 的接口	70
3.5 CPU 的内核	72
3.6 典型 CPU 介绍	73
3.6.1 Intel 系列 CPU	73
3.6.2 AMD 系列 CPU	78
习题	83
实验	84

第4章 内部存储器

4.1 存储系统概述	85
4.2 内存的作用及其分类	86
4.2.1 内存的作用	86
4.2.2 内存的分类	87
4.2.3 内存的主要技术指标	87
4.3 半导体存储器的组成及工作 原理	89

4.3.1 随机存储器	89
4.3.2 只读存储器	91
4.3.3 内存的组成	93
4.4 RAM 的基本工作方式	95
4.5 内存模组与基本结构	96
4.5.1 物理存储器	96
4.5.2 逻辑存储器	96
4.5.3 内存条(模组)的结构及工作 原理	97
4.6 主流内存条介绍	100
4.6.1 FPM DRAM	100
4.6.2 EDO DRAM	101
4.6.3 SDRAM	102
4.6.4 DDR SDRAM 内存	103
4.6.5 DDR2 SDRAM 内存	104
4.6.6 Rambus 内存	106
4.6.7 几种常见内存的带宽比较	108
4.6.8 内存接口类型	109

4.7 内存相关技术	110
4.7.1 内存双通道技术	110
4.7.2 内存参数及其优化	112
4.7.3 内存技术规范及标注格式	114
习题	115
实验	116

第5章 总线与芯片组

5.1 总线	117
5.1.1 总线的概念	117
5.1.2 总线的类型和性能	118
5.1.3 PC 机总线的演变	119
5.1.4 ISA 总线	122
5.1.5 PCI 总线	123
5.1.6 PCI Express 总线	126
5.2 芯片组	128
5.2.1 芯片组的概念	128
5.2.2 Intel 公司早期的芯片组	129
5.2.3 用于 Pentium 4 的 8 × x 系列芯 片组	131
5.2.4 Intel 9 × x 系列芯片组	132
5.3 主板	137
5.3.1 主板的构成	137
5.3.2 主板标准	139
习题	141
实验	142

第6章 接口

6.1 接口的基本知识	143
6.1.1 接口的构成	143
6.1.2 接口信号	145
6.1.3 接口的分类	146
6.1.4 接口的操作方式	147
6.2 RS - 232C	150
6.2.1 RS - 232C 串行接口标准	150
6.2.2 RS - 232C 的接口信号	153
6.2.3 RS - 232C 串行接口的实现	154

6.3 并行接口	156
6.3.1 并行接口简介	156
6.3.2 并行接口的工作模式	158
6.3.3 并行接口的使用	160
6.4 USB 接口	160
6.4.1 新型的串行传输技术简介	161
6.4.2 USB 简介	161
6.4.3 USB 的技术特点	162
6.4.4 USB 传输模式	164
6.5 IEEE 1394	165
6.5.1 IEEE 1394 标准	165
6.5.2 IEEE 1394a 和 USB 的比较	166
6.6 IDE 接口	168
6.6.1 IDE 接口简介	168
6.6.2 IDE 接口的版本和发展过程	170
6.7 SATA 接口	172
6.7.1 从 ATA 到 SATA	172
6.7.2 SATA 简介	173
6.8 SCSI 接口	175
6.8.1 SCSI 概述	175
6.8.2 SCSI 电缆和连接器	176
6.8.3 SCSI 的版本	176
6.8.4 SCSI 到 SAS	178
6.9 红外线接口	179
6.9.1 红外线传输的特点	179
6.9.2 IrDA 简述	180
6.10 无线接口	182
6.10.1 蓝牙	182
6.10.2 Wi - Fi	183
习题	184
实验	185

第7章 外部存储器

7.1 硬盘	186
7.1.1 硬盘的结构与工作原理	186
7.1.2 硬盘的性能指标	190
7.1.3 分区和格式化	193

第8章 常用外部设备

8.1 键盘与鼠标	216
8.1.1 键盘	216
8.1.2 鼠标	224
8.2 扫描仪	229
8.3 触摸屏	233
8.4 显示系统	236
8.4.1 CRT 显示器	236
8.4.2 液晶显示器	243
8.4.3 显示适配器	248
8.5 音频设备	256
8.5.1 音频设备的发展历史及用途	256
8.5.2 声卡的工作原理	257
8.5.3 声卡的输入输出接口	261
8.6 打印机	263
8.6.1 针式打印机	263
8.6.2 喷墨打印机	264
8.6.3 激光打印机	265
8.6.4 打印机的接口	266
习题	267

第9章 多媒体设备

9.1 多媒体的基本概念	269
9.1.1 多媒体技术的定义	269

第 1 章 计算机概述	1	2.2.3 组合逻辑电路	32
1.1 计算机发展史	1	2.2.4 时序逻辑电路	36
1.1.1 从电子器件的发展看电子计算机的发展史	1	2.2.5 典型时序电路——触发器	36
1.1.2 摩尔定律	3	2.2.6 计数器	39
1.1.3 计算机发展过程中的重要里程碑	5	2.2.7 寄存器	40
1.2 现代计算机的特点及其类型	8	2.3 计算机的硬件组成	41
1.3 计算机的性能与发展趋势	10	2.3.1 CPU 的组成	41
1.3.1 计算机的运算速度与存储容量	10	2.3.2 指令系统	43
1.3.2 计算机的发展趋势	11	2.3.3 CPU 的工作过程	45
1.4 计算机的基本结构	12	习题	46
1.4.1 冯·诺依曼模型	12	第 3 章 微型计算机的中央处理器	48
1.4.2 计算机系统结构	13	3.1 CPU 概述	48
1.4.3 计算机的模块结构	14	3.2 CPU 的主要技术指标	50
习题	16	3.2.1 位、字节和字长	50
第 2 章 计算机系统组成及工作		3.2.2 时钟频率	50
原理	17	3.2.3 一级和二级高速缓存的容量和速率	51
2.1 二进制数与信息表示	17	3.2.4 CPU 的扩展指令集	52
2.1.1 为什么要使用二进制	17	3.2.5 工作电压	52
2.1.2 信息在计算机内的表示	18	3.2.6 地址总线与数据总线的宽度	52
2.1.3 数值型数据	18	3.2.7 制造工艺	52
2.1.4 非数值型数据	21	3.3 提高 CPU 性能的先导技术	53
2.1.5 数据格式	23	3.3.1 流水线与超标量结构	53
2.1.6 原码、补码和反码	24	3.3.2 高速缓存技术	54
2.1.7 二进制加、减运算	27	3.3.3 NetBurst 架构 Pentium 4 中的高速缓存实现技术	61
2.2 逻辑电路	28	3.3.4 扩展指令集	62
2.2.1 有关逻辑电路的基础知识	28	3.3.5 64 位技术	65
2.2.2 逻辑门电路	31		

9.1.2 多媒体技术的主要特征	270	10.2.1 数据通信的分类	303	第11章 专用计算机与工业控制 系统	356	11.5 工业控制系统	372
9.1.3 多媒体系统	271	10.2.2 通信速率与数据传输方式	304	11.1 嵌入式系统	356	11.5.1 什么是工业控制系统	372
9.1.4 多媒体个人计算机	271	10.2.3 数据的同步方式	304	11.1.1 嵌入式系统的概念及特点	356	11.5.2 分散控制系统的特性	372
9.2 音频卡	272	10.2.4 编码与调制	305	11.1.2 嵌入式系统的结构及分类	357	11.5.3 现场总线控制系统	375
9.2.1 采样频率与采样精度	272	10.2.5 多路复用技术	306	11.2 DSP	360	11.5.4 现场总线控制系统的特性	375
9.2.2 输入输出接口	272	10.2.6 信号带宽、信道带宽与差错控制	308	11.2.1 DSP 的概念	360	11.5.5 现场总线控制网络模型	376
9.2.3 声卡芯片	276	10.3 分布系统与计算机网络	308	11.2.2 DSP 的发展历史	360	11.5.6 控制网络和信息网络的集成	378
9.2.4 外置声卡	276	10.4 虚拟专用网络	313	11.2.3 DSP 的特性	361	11.5.7 现场总线的标准	379
9.2.5 与专业声卡相关的一些概念	278	10.5 调制解调器	315	11.3 SOPC 及其技术	364	习题	380
9.3 摄像头	279	10.5.1 模拟调制解调器	315	11.4 微控制器	367	英文缩略语	381
9.3.1 图像传感器	279	10.5.2 DSL 技术	316	11.4.1 什么是微控制器	367	参考文献	387
9.3.2 数字信号处理器	280	10.5.3 电缆调制解调器	324	11.4.2 微控制器的基本结构	368		
9.3.3 镜头	280	10.5.4 电力线调制解调器	326				
9.3.4 摄像头的主要技术指标	281	10.5.5 ISDN 和 ISDN 调制解调器	330				
9.4 多功能显示卡	282	10.5.6 CDMA 和 GSM/GPRS 调制解调器	334				
9.4.1 多功能显示卡的发展	282						
9.4.2 多功能显示卡的常见功能与应用	284	10.6 网络设备	334				
9.5 视频卡	286	10.6.1 网络适配器	335				
9.5.1 视频卡的基本概念	286	10.6.2 无线网卡	337				
9.5.2 视频卡的基本类型	287	10.7 中继器	340				
9.5.3 视频卡的主要性能指标	287	10.8 网桥	340				
9.6 VGA 设备	289	10.9 集线器	340				
9.7 投影机	292	10.10 交换机	341				
9.7.1 投影机的工作原理	292	10.11 路由器	343				
9.7.2 投影机接口的分类及用途	297	10.12 宽带路由器	344				
9.7.3 投影机的主要性能指标	298	10.13 集线器、交换机与路由器的区别	344				
9.7.4 便携式投影机	299	10.14 网关	346				
习题	300	10.15 网络打印机	348				
第10章 计算机通信和通信设备	301	10.15.1 网络打印机的基本概念	348				
10.1 计算机通信	301	10.15.2 网络打印机的应用	348				
10.1.1 计算机技术与通信技术的融合	301	10.16 VoIP	350				
10.1.2 通信系统的组成	302	10.16.1 VoIP 的基本概念	350				
10.2 数字通信的基本概念	303	10.16.2 VoIP 的实现	352				
		10.16.3 VoIP 的应用方案	354				
		习题	355				

- 半导体晶体管:体积小、功耗低、可靠性高。包括双极型晶体管和场效应晶体管。
- 集成电路(integrated circuits, IC):速度快、体积小、功耗更低、可靠性更高。根据集成度的不同,集成电路可分为如下几种(如图1-2所示):

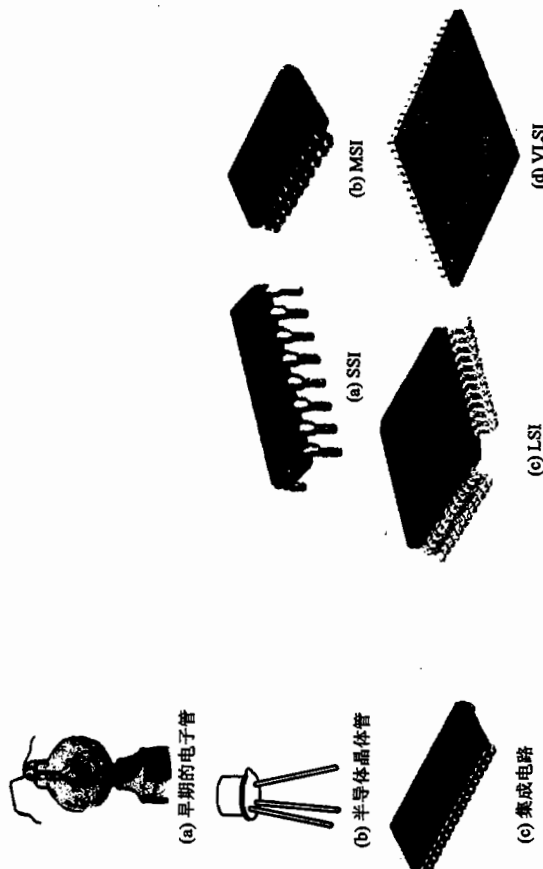


图1-1 电子器件的发展

- SSI(小规模集成电路):晶体管数100个以下。
- MSI(中规模集成电路):晶体管数100~3 000个。
- LSI(大规模集成电路):晶体管数3 000~10⁵个。
- VLSI(超大规模集成电路):晶体管数10⁵~10⁸个。
- ULSI(甚大规模集成电路):晶体管数超过10⁸以上。

图1-2 各种规模的集成电路

2. 计算机的发展史

与电子技术的发展相对应,计算机的发展也经历了相似的几个阶段。

- (1) 电子管计算机(1946年到20世纪50年代后期)
 - 存储器:采用延迟线、磁芯。
 - 无高级语言,甚至没有汇编器。
 - 主要用于科学计算或为军事与国防尖端科技服务。
 - 没有商业化。
- (2) 晶体管计算机(20世纪50年代中期到20世纪60年代后期)
 - 磁芯存储器,存储容量从几千提高到10万字以上。
 - 运算速度从每秒几千次提高到几十万次。
 - 应用于军事与尖端技术、气象、工程设计、数据处理以及其他科学研究等领域。

- (3) 中、小规模集成电路计算机(从20世纪60年代中期到20世纪70年代前期)
 - 功耗、体积、价格等进一步下降。
 - 速度及可靠性相应提高。
 - 应用范围进一步扩大。
- (4) 大规模集成电路计算机(从20世纪70年代初至今)
 - 20世纪60年代后期,半导体存储器取代了磁芯存储器,并不断向大容量、高集成度、高速发展。
 - 20世纪70年代初,出现微型计算机——以微处理器(MPU,单片IC的CPU)为核心的电子计算机。
 - 计算机进入了几乎所有的领域。

1.1.2 摩尔定律

基于上述计算机与电子技术之间的密切关系,可以说计算机的发展就是电子技术发展的一种体现,因此电子技术发展的规律将直接影响计算机的发展。世界上第一个集成电路诞生时,芯片上只包括一个晶体管、两个电阻和一个电阻-电容网络。随着集成电路工艺日趋完善,芯片上所包含的元件数量以每1~2年翻一番的速度不断增长。1965年,Intel公司的创办人之一摩尔发现了一个具有历史意义的现象:“在相等面积(制作成本)上,CPU上的晶体管数量以每18个月倍增的趋势增加,执行性能的提升大体也符合这个趋势。”这就是著名的摩尔定律,它揭示了一个趋势。这个趋势一直延续至今,并且事实证明相当准确。Intel公司微处理器芯片的发展便印证了摩尔定律的正确性。

1971年诞生的4位微处理器4004,是Intel生产的第一微处理器,它大约有2 300个晶体管,以10 μm 工艺制作,内存空间仅640 B,工作频率为108 kHz。

1972年,Intel推出第一个8位微处理器8008,它大约有3 500个晶体管,以10 μm 工艺制作,内存空间16 KB,工作频率为200 kHz。

1974年的8080是第一个真正的微处理器,由6 000个晶体管组成,采用6 μm 工艺,工作频率为2 MHz,内存空间64 KB。

1978年推出的8086是16位微处理器,由29 000个晶体管所组成,采用3 μm 工艺。8086的数据总线与寄存器宽度皆为16位,最大内存空间1 MB,工作频率为4.77 MHz。之后又推出了8088。1980年,IBM公司推出了以8088为微处理器的IBM PC(以及随后的PC XT)。

1980年诞生的80186/80188与8086/88的内部结构相似,但因80186内置的中断控制器与IBM PC所规划的中断控制器不兼容,故大多应用在嵌入式系统中。

1982年,Intel公司推出了x86体系结构,直到今天,x86体系结构仍然是大多数Intel处理器的基础。80286是第一款基于x86体系结构的微处理器,是第一颗具备多任务切换、内存分段保护与虚拟内存功能的16位微处理器。它有134 000个晶体管,采用1.5 μm 工艺,工作频率为6 MHz~25 MHz。1984年,IBM以Intel 80286为CPU架构,推出了PC AT。

1985年诞生的80386采用的是2 μm 工艺,由275 000个晶体管组成,内部寄存器、外部数据与内存总线宽度全部扩展为32位,最大内存空间4 GB,工作频率从16 MHz开始,可外接64 KB~128 KB高速缓冲存储器(Cache)。

1989年, Intel公司以最新的1 μ m工艺开发出更细小的386芯片,即386SX。不同于原先的386(改称为386DX),内部寄存器、外部数据与内存总线皆为32位宽度。386SX内部寄存器为32位。32位。但外部数据总线宽度仅16位,目的是方便与既有的286芯片组搭配出一台低成本的16/32位386计算机。

1989年推出的80486由1 200 000个晶体管组成,开始以1 μm 工艺制作,随后的DX2系列进步到0.8 μm ,最大内存空间4 GB。80486的指令系统与8086/8088/286/386兼容,芯片内部包含8 KB Cache和浮点运算单元(FPU)。

1993 年推出的 80586, 又称奔腾 Pentium, 由 3 100 000 个晶体管组成。Pentium 是最早采用超标量结构的处理器。其内部具有两个整数运算单元、一个浮点运算单元, 16 KB 高速缓存 (指令和数据各占一半), 最大内存空间 4 GB。Pentium 初期推出的速度版本分别是 60 MHz、66 MHz, 以 0.8 μm 工艺制作, 核心电压为 5 V。随后 Intel 采用 0.6 μm 工艺, 推出了 Pentium 75 MHz、90 MHz、100 MHz 与 120 MHz 版本, 晶体管数量增为 330 万个, 但核心电压降低至 3.3 V。

1995 年推出 Pentium Pro (高能奔腾), 具有 5 500 000 个晶体管, 采用 0.35 μm 工艺, 最大内存空间达到 64 GB, 工作频率为 150 MHz~200 MHz。

1997 年推出 Pentium II 由 750 万个晶体管制成, 采用了 0.35 μm 工艺, 最大内存空间达到 64 GB, 工作频率为 200 MHz ~ 300 MHz, 主要架构跟第六代 x86 的 Pentium Pro (P6) 类似。

1999年推出Pentium III, 包含2 800万个晶体管, 采用了0.25 μm 工艺。Pentium III依然沿用了P6的系统架构, 极大地提高了计算机在高级图形、三维动画、视频、音频等方面的性能。

2000年11月,Intel的新一代微处理器——Pentium 4(奔腾4)诞生。Pentium 4由4200万个晶体管组成,采用0.18 μm 工艺,工作频率达1.5 GHz。之后,采用0.13 μm 工艺,工作频率达1.7 GHz。

图 1-3 给出了 Intel 微处理器的发展情况示意图,显然与摩尔定律非常吻合。可以看到,计算机及微处理器的发展史,就是微电子技术的发展史。

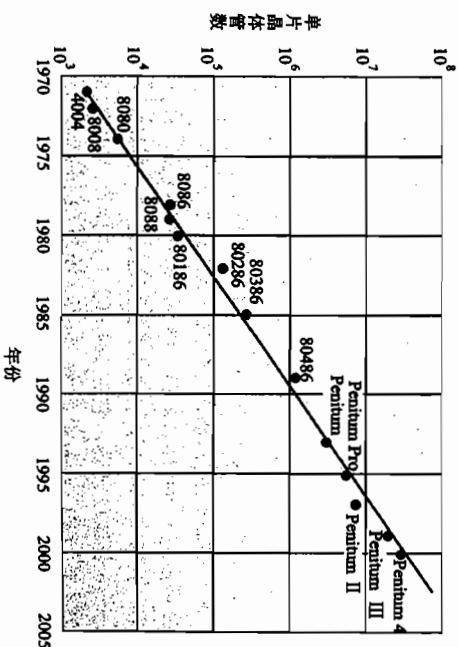


图 1-3 Intel 微处理器的发展概况

1.1.3 计算机发展过程中的重要里程碑

1. 第一台电子计算机

1946 年在美国诞生的“埃尼亚克”(ENIAC)是举世公认的第一台电子计算机,如图 1-4 所示。其设计的目的是为阿贝丁陆军试验场进行弹道计算。主要研制者是宾夕法尼亚大学莫尔学院的莫契利(J. Mauchly)和埃克特(J. P. Eckert)。ENIAC 共使用了 18 000 个电子管,另加 1 500 个继电器以及其他器件,其总体积约 90 m³,重达 30 t,占地 170 m²。这台耗电量为 140 kW 的计算机,运算速度为每秒 5 000 次加法、或者 400 次乘法,至多只能存储 20 个 10 位的十进制数,无程序存储器。



图 1-4 ENIAC

如果将 5 000 次加法看做每秒 5 000 次运算, 这样的计算能力是个什么概念呢? 以自动驾驶为例: 每小时行驶 60 公里时, 要求处理速度达到 10 ~ 100 GIPS (千兆条指令每秒), 即每秒 100 亿 ~ 1 000 亿次运算, 5 000 次运算相当于行车速度为:

$$\frac{5 \times 10^3}{10 \times 10^9} \times 60 \times 10^3 \times 10^2 = 3 \text{ cm/h}$$

显然,这样的速度比蜗牛还缓慢。

ENIAC 有两大缺点,一是没有内存存储器,二是要由人像搭积木一样,将大量运算部件搭配成各种解题布局,每算一题就要重搭一次,又费时,又麻烦。有的题只要计算 1 秒钟,准备工作却要花上几十分钟。对 ENIAC 的改进,便是著名的美籍匈牙利数学家冯·诺依曼提出的存储程序控制。

2. 存储程序概念的提出

1945年6月,以美籍匈牙利数学家冯·诺依曼为首的研制小组与参与研制ENIAC主要人员联名发表了一篇长达101页的报告,即计算机史上著名的“101页报告”,提出了“存储程序控制”的计算机结构(即冯·诺依曼机),奠定了现代计算机的基础。概括来说,冯·诺依曼机的体系结构具有如下特点:

① 计算机(指硬件)由五大基本部件组成:运算器、存储器、控制器、输入和输出系统。

- ② 计算机内部采用二进制来表示指令和数据。
- ③ 将编好的程序和原始数据事先存入存储器中,然后再执行程序,完成预定的任务。
- 3. 第一台存储程序计算机
1949年在英国剑桥大学问世的EDSAC,如图1-5所示,是由3 000个电子管为主要元件的存储结构的计算机。



图 1-5 EDSAC

1951年冯·诺依曼的EDVAC问世,总共只采用了2 300个电子管,但运算速度却比拥有18 000个电子管的ENIAC提高了10倍。

4. 晶体管计算机的诞生

1954年,贝尔实验室研制成功第一台使用晶体管线路的计算机(TRANIC),装有800个晶体管。

1958年,IBM公司制成了第一台全部使用晶体管的计算机RCA501。计算速度从每秒几千次提高到几十万次,主存储器的存储容量,从几千字提高到10万字以上。

1959年,IBM公司又生产出全部晶体管化的数字计算机IBM 7090。

5. 第三代计算机的标志性产品 IBM S/360 系列计算机

1964年,由IBM公司主设计师吉恩·阿姆达尔(G. Amdahl)主持设计的第一个采用集成电路的通用计算机系列IBM S/360系统研制成功,如图1-6所示,该系列有大、中、小型,共6个型号。阿姆达尔提出了一种全新的思路:IBM S/360必须是一种“兼容性”的产品,这意味着大、中、小型IBM S/360系统计算机,都能以相同方式处理相同的指令,运行相同的软件,配置相同的外部设备,而且能够相互连接在一起工作。“兼容性”是一次伟大的观念变革,由此引起的现代计算机的技术进步,至今仍发挥着巨大的作用。IBM S/360系统取得了难以置信的成功,成为第三代计算机的标志性产品。

6. 使用超大规模集成电路的第四代计算机

1965年,美国国防部拨款开发的ILLIAC-IV计算机,是第一台全面以大规模集成电路作为逻辑元件和存储器的阵列计算机,它标志着计算机的发展已到了第四代。1976年问世的被誉为“美国民族智慧星”的克雷公司的Cray系列巨型机(如图1-7所示)以及IBM公司的370系列、43XX、30XX、ES-9000系列和日本富士通公司生产的M系列都是比较有代表性的第四代计算机。



图 1-6 IBM S/360 系统



图 1-7 Cray-2 巨型机

7. 微型计算机的出现

Intel系列CPU的发展,特别是1978年推出的16位CPU——8086,为微型计算机的出现奠定了硬件基础。

1980年,IBM实行开放政策,采用Intel 8088 MPU、委托独立软件公司为其开发各种软件。

1981年8月12日,IBM在纽约宣布IBM个人计算机(PC)问世,如图1-8所示。



图 1-8 IBM PC

表1-1给出了上述以电子技术为特征划分的四代计算机的情况。

表 1-1 四代计算机概况

起迄年份	代别	主要逻辑元件	软件
1946—1957	一	电子管	机器语言、汇编语言
1958—1964	二	晶体管	高级语言、监控程序
1965—1970	三	集成电路	简单操作系统
1970 至今	四	大规模或超大规模集成电路	软件工程的研究及应用、数据库、语言编译系统和网络软件

表1-2给出了几代计算机中央处理器(CPU)集成度和运算速度等性能,反映了CPU的发展概况。

表 1-2 CPU 的集成度和运算性能

代别	年份	字长 (位)	型号	工艺	晶体管数 (万个)	时钟频率 (MHz)	运算速度 (MIPS)
一	1971	4	4004	PMOS	0.2	<1	0.06
二	1974	8	8080	NMOS	0.5	2~4	0.4
三	1978	16	8086	HMOS	2.9	4.77~10	<1
四	1985	32	80386	CMOS	27.5	16~33	6~12

续表							
代别	年份	字长 (位)	型号	工艺	晶体管数 (万个)	时钟频率 (MHz)	运算速度 (MIPS)
五	1993	32	Pentium	CMOS	320	60 ~ 133	100 ~ 200
六	1997	32	Pentium II	CMOS	750	233 ~ 400	~ 300
七	2000	32	Pentium IV	CMOS	4 200	1 300 ~ 3 000	> 2 000
八	2004	64	Pentium IV F	CMOS	16 900	3 200 ~ 3 600	> 5 000

1.2 现代计算机的特点及其类型

计算机按用途可分为专用计算机和通用计算机。一般来说,专用计算机是高效、经济、快速的计算机,但是其功能单一,适用范围很窄;通用计算机功能齐全,适应性很强,但是牺牲了效率、速度和经济性。专用和通用是根据计算机的效率、速度、价格、运行的经济性和适用范围来划分的。

现在,一般意义上讲的计算机都是指通用计算机。

1989年11月IEEE提出一个分类报告,它根据计算机在信息处理系统中的地位与作用,考虑到计算机分类的演变过程和近期可能的发展趋势,把计算机分成六大类:巨型计算机、小型计算机、主机、小型计算机、工作站和个人计算机。目前,国际上比较流行的分类如下:

- 巨型计算机
- 主机(大、中型机)
- 小型计算机
- 个人计算机

1. 巨型计算机(supercomputer)

巨型计算机也称超级计算机,如图1-9所示,是指运算速度快、存储容量大的高性能计算机,它采用了大规模并行处理体系结构,CPU由数以百计、千计的处理器组成,有极强的运算处理能力。对巨型计算机的指标有这样一些规定:首先,计算机的运算速度平均每秒1 000万次以上;其次,存储容量在1 000万位以上。例如,我国研制成功的“银河”计算机,就属于巨型计算机。自从1964年美国Control Data Corporation推出世界首台巨型机CDC 6600以来,由于受科学研究和工程计算的需求牵引,国际巨型计算机技术发展迅猛,其峰值性能已从30多年前的每秒百万次上升到今天的每秒40万亿次(指日本NEC公司的“地球模拟器”系统),平均每10年提高3个数量级。IBM

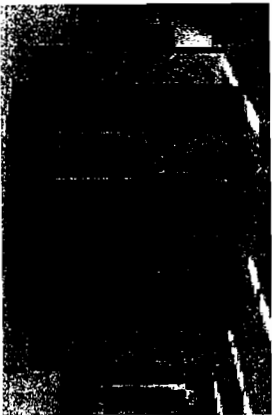


图1-9 银河II十亿次巨型计算机

公司计划在2007年推出峰值性能1 048万亿次的IBM蓝色基因/P巨型机。目前,我国已研制出“银河Ⅲ”百亿次并行巨型计算机。由于巨型计算机结构复杂、价格昂贵,主要应用于尖端科学计算和现代化军事领域。研制这类巨型计算机是现代科学技术,尤其是国防尖端技术发展的需要。核武器、反导弹武器、空间技术、大范围天气预报、石油勘探等都要求计算机具有很高的处理速度、很大的容量,一般的计算机远远不能满足要求。巨型计算机的研制水平是一个国家计算机技术水平的重要标志。

2. 主机(Mainframe computers)

主机包括通常所说的大、中型计算机。大型计算机的运算速度一般在100万次/秒至几千万次/秒,通常用每秒运行多少万次来作为运算速度单位,字长32~64位,主存容量在几百兆字节以上。它比较完善的指令系统、丰富的外部设备和功能齐全的软件系统。其特点是通用,有极强的综合处理能力。1964年4月7日,IBM发布了世界上第一台大型计算机——IBM S/360,如图1-10所示,它具有32位通用寄存器、24位存储器编址及最大7个(并行)I/O通道。2000年,IBM公布了Z系列大型计算机,这是IBM最新的,也是拥有最多改进型号的大型计算机系列。Z系列主机引入了新的64位操作系统,并保留了对24/32位应用程序的兼容。大型主机经历了批处理阶段、分时处理阶段,进入了分散处理与集中管理的阶段。IBM公司一直在大型主机市场处于霸主地位,DEC、富士通、日立、NEC也生产大型主机。不过随着微机与网络的迅速发展,大型计算机正在走下坡路,现在主要朝着超级服务器方向发展。中型计算机的规模介于大型计算机和小型计算机之间。

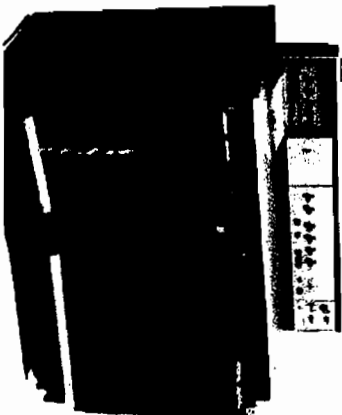


图1-10 IBM S/360大型计算机

3. 小型计算机(minicomputers)

小型计算机的规模小、结构简单、研发周期短,便于及时采用先进工艺技术,软件开发成本低,易于操作维护。它们已广泛应用于工业自动控制、大型分析仪器、测量设备、企业管理、大学和科研机构等,也可以作为大型与巨型计算机系统的辅助计算机。近年来,小型计算机的发展也引人注目。特别是精简指令集计算机(Reduced Instruction Set Computer, RISC)体系结构的广泛应用,大大提高了小型计算机的性能。

4. 微型计算机

微型计算机又简称为μC或MC(micro computer)。以微处理器(CPU)为核心,通过系统总线(bus)将存储器、外围控制电路、输入/输出接口连接起来的系统称为微型计算机。若配有相应的外围设备(如显示器、键盘、打印机等)和系统软件,就组成了微型计算机系统(micro computer system)。

微型计算机具有以下特点:

- (1) 集成度高,体积小,重量轻,价格低廉。

(2) 部件标准化,易于组装及维修。

(3) 高可靠性及适应性。

如 1.1.3 节所述,微型计算机诞生于 20 世纪 70 年代,一方面是由于当时军事、工业自动化技术的发展,需要体积小、功耗低、可靠性好的微型计算机;另一方面,由于大规模集成电路(LSI)和超大规模集成电路(VLSI)的迅速发展,可以在单片硅片上集成几千到几十万个晶体管,为微型计算机的产生打下了坚实的物质基础,引发了新的技术革命。微型计算机一经问世,就势不可阻挡的势头迅猛发展,成为当今计算机发展的一个主流方向。

微型计算机的发展已经历了四代。第一代微型计算机(1971—1973)采用 4 位和低档 8 位微处理器;第二代微型计算机(1974—1977)采用中、高档 8 位微处理器;第三代微型计算机(1978—1984)采用 16 位微处理器;第四代微型计算机(1985—1993)采用 32 位高档微处理器。现在,第五代微型计算机(1993 年至今)采用 64 位高档微处理器。人们一般以字长和典型的微处理器作为微型计算机各发展阶段标志。几乎每两年微处理器的集成度便翻一番,微型计算机 2~4 年更新换代一次。

除常见的台式微型计算机以外,笔记本计算机及掌上电脑也都属于微型计算机。微型计算机的发展趋势呈现以下特点:

- 体积小、通用化
- 维护与操作简单化
- 价格平民化
- 功能多媒体化、网络化

1.3 计算机的性能与发展趋势

1.3.1 计算机的运算速度与存储容量

运算速度是计算机的一项重要性能指标,以前常采用每秒执行的指令条数来表示。但计算机的指令类型很多,而且每一种指令出现的频繁程度也不完全一样,因此这样的指标很难反映计算机真正的运算速度。目前,微型计算机常以微处理器的时钟频率(又称为主频)来衡量运算速度。

计算机的运算速度与许多因素有关,如微处理器的主频、执行何种操作、主存本身的速度(主存速度越快,取指令、取数据就越快)等有关。

早期的计算机运算速度用完成一次加法或乘法运算的时间来衡量。后来曾改用吉普森(Gibson)法,它综合考虑了每条指令的执行时间以及其在全局操作中所占的百分比,其计算公式如下:

$$T_M = \sum_{i=1}^n f_i t_i$$

其中, T_M 为机器运算速度, f_i 为第 i 种指令占全部操作的百分比, t_i 为第 i 种指令的执行时间。目前,计算机的运算速度以单位时间内所执行指令的平均条数来衡量,单位为 MIPS (Million In-

struction Per Second, 每秒百万条指令);也有以 CPI (Cycle Per Instruction), 即执行一条指令所需时钟周期数来表示;或以 FLOPS (Floating Point Operation Per Second), 即每秒浮点运算次数来衡量。

计算机的存储容量是计算机性能的又一重要指标。计算机的存储容量可用字节(B)来表示。一个字节可表示一个英文字母或数字,而汉字则需要用两个字节来表示。规定 8 个二进制位为一个字节。例如,大写英文字母 A 占一个字节,即可用一个 8 位长的二进制数字 01000001 来表示。

字节这一单位太小了。人们常用千字节(KB)、兆字节(MB)甚至吉字节(GB)来表示存储容量,它们之间的换算关系是:

$$\begin{aligned} 1 \text{ KB} &= 1024 \text{ B} \\ 1 \text{ MB} &= 1024 \text{ KB} \\ 1 \text{ GB} &= 1024 \text{ MB} \end{aligned}$$

例如,一个内存条的容量可能是 8 MB,一张 3 英寸高密软盘的容量是 1.44 MB,一张 VCD 光盘的容量是 650 MB。

1.3.2 计算机的发展趋势

目前计算机正以超大规模集成电路为基础,向巨型化、微型化、网络化与智能化的方向发展。

1. 巨型化

巨型化是指计算机的运算速度更高、存储容量更大、功能更强。第四代计算机在实现微型化的同时,也在实现巨型化。虽然从体积上说,当今最大的巨型计算机也未必能和第一台计算机相比,但它的运算速度则达到了第一台计算机的百万倍、千万倍甚至数十亿倍。

2. 微型化

微型计算机已进入仪器、仪表、家用电器等小型仪器设备中,使仪器设备实现“智能化”,而成为工业过程控制的核心。随着微电子技术的进一步发展,更适合移动环境、便携式的笔记本型、掌上型等微型计算机,必将以更优越的性能价格比受到人们的青睐。

3. 网络化

计算机网络是现代通信技术与计算机技术相结合的产物,已经在现代社会生活中发挥着越来越重要的作用,如银行系统、商业系统、交通运输系统等。计算机互联网络是目前世界上最大的信息网。存储设备也在向网络化方向发展,可通过高速光纤通道与存储交换设备连接在一起;存储设备将实现集中管理;存储设备的动态分布和分配使应用软件所需的存储容量变得十分灵活;网络化存储设备可实现网络连到哪里,存储设备也分布到哪里。Sun 公司早就提出了“网络就是计算机”的口号。从世界 IT 发展趋势看,网络计算时代已经到来。

4. 智能化

计算机的智能化就是让计算机来模拟人的感觉、行为、思维过程的机理,使计算机具备逻辑推理、学习等能力。超级计算机性能再好,速度再快,却仍在按人们事先编制好的程序指令办事,仍无法成为能容忍程序错误的计算机。计算机应用的发展在经历了几十年以键盘为工具的复杂人机交互后,正疾速跨入新的以“能听会说”为表征的智能时代。建立在现代科学基础之上的计

算机人工智能是计算机发展的一个重要方向,新一代计算机将可以模拟人的感觉行为和思维过程的机理,具有“看”、“听”、“说”、“想”、“做”的能力,并具有学习、逻辑判断、推理及自然语言能力。

另一个值得关注的计算机未来发展的趋势是将出现新型材料计算机——光子计算机、生物计算机、超导计算机、量子计算机、纳米计算机等。

光子计算机是一种由光信号进行数字运算、逻辑操作、信息存储和处理的新型计算机。光子计算机具有下列优势。

(1) 运算速度快。电子计算机的运算速度取决于每个开关、器件改变状态所需的时间。目前使用的开关速度最快的是硅晶体管,其开关时间很难小于亿分之一秒,即计算机的工作速度很难超过每秒亿次。如果能制造出用激光来传送信号的光计算机,从理论上计算,光晶体管开关时间在皮(10^{-12})秒水平,即这种光开关器件每秒能进行万亿次逻辑运算,与当前运算速度最高的巨型计算机相比还要快数百倍。

(2) 存储容量大。它的存储容量可达 10^{12} 二进制信息位,将是现有电子计算机的几万亿倍。(3) 可以同时处理几路信息。几束光通过光晶体管时能够相互独立,而几路电流输入晶体管时却会混在一起。

(4) 不受磁场影响。电子计算机利用电子传输信息,容易受磁场影响;而光子计算机是利用光子传输信息,不会受磁场影响。

正因为光子计算机的这些优越性能,所以世界上许多国家正在努力发展光子计算机。

生物计算机是利用有机分子作为基本部件的计算机。因为有机分子存在于生物体内,所以这种计算机称为“生物计算机”。生物计算机具有以下显著的优点。

(1) 体积小、功效高。以分子水平的线路为目标的生物化学元件线度可达几百埃(1埃 = 10^{-10} m),一平方毫米的面积上可容纳亿个电路,比目前的电子计算机提高了几百倍。

(2) 可靠性高。因为生物本身具有自我修复的机能,所以即使计算机芯片出了故障也能自我修复。

(3) 耗能少。有机分子构成的生物化学元件是利用化学反应工作的,需要能量少,不存在发热问题。

超导计算机是用超导器件作为元件的计算机。超导计算机的两个基本器件——超导开关和超导存储器具有如下特点。

(1) 开关速度快。目前已达几皮秒,比高速集成电路快几百倍。

(2) 功耗非常小。仅为硅集成电路的几分之一。有人预测,过去需要 10 kW 功率的大、中型计算机,若用超导器件制成的超导计算机,则只需 1 节电池供电。

各具优势、特点的量子计算机与纳米计算机等也处于研究之中。

1.4 计算机的基本结构

1.4.1 冯·诺依曼模型

现代计算机硬件系统的基本结构仍基于冯·诺依曼(Von Neumann)提出的设计思想,即计

算机硬件系统由运算器、控制器、存储器、输入设备和输出设备五大基本部件组成。采用二进制的形式(0、1)表示数据和指令,数据和指令预先存入内存,计算机能自动取出并执行指令,如图 1-11 所示。这就是著名的冯·诺依曼模型。

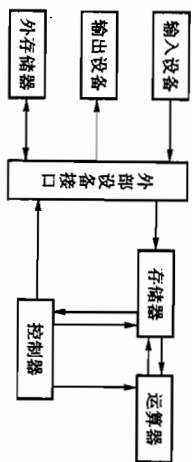


图 1-11 冯·诺依曼模型

1.4.2 计算机系统结构

建筑学、系统结构、架构

“计算机系统结构”一词源于英文 computer architecture,也有译成“计算机体系结构”的。计算机系统结构研究的内容不但涉及计算机硬件,也涉及计算机软件,已成为一门学科。

计算机这个术语最早由 Amdahl 等人在 1964 年提出。他们把系统结构定义为由程序设计者所看到的一个计算机系统的属性,即概念性结构和功能特性。这实际上是计算机系统的内在特性。按照计算机层次结构,不同程序设计者所看到的计算机有不同的属性。使用高级语言的程序员所看到的计算机属性主要是软件子系统和固件子系统的属性,包括程序语言以及操作系统、数据库管理系统、网络软件等用户界面。Amdahl 等人提出的系统结构定义是指机器语言或编译程序设计者所看到的计算机属性,是硬件子系统的概念结构及其功能特性,包括:

- ① 机器内的数据表示,即硬件能直接辨认和处理的那些数据类型。
- ② 寻址方式,包括最小寻址单元和地址运算等。
- ③ 寄存器定义,包括操作数寄存器、变址寄存器、控制寄存器等定义、数量和使用方式。
- ④ 指令系统,包括机器指令的操作类型和格式、指令间的排序和控制机构等。
- ⑤ 中断机构,包括中断的类型和中断响应硬件的功能等。
- ⑥ 机器工作状态的定义和切换,如管态和目态等。
- ⑦ 输入输出结构,包括输入输出的连接方式、处理机/存储器与输入输出设备间数据传送的方式和格式、传送的数据量以及输入输出操作的结束与出错标志等。
- ⑧ 信息保护,包括信息保护方式和硬件对信息保护的支持等。

这些即是程序员为了使其所编写的程序能在机器上正确运行,需要了解和遵循的计算机属性。当然不包括基本的数据流、控制流、逻辑设计和物理实现等。

将不同人员所关注的计算机属性的不同进行这种概念性的延伸,仅局限于一般用户观察到的一般用途的计算机硬件系统,主要包括主机、显示器、键盘和鼠标等。

计算机的设计和制造者除关心计算机的组成原理和实现方法外,甚至还需要对构造计算机各部件的芯片或门电路有深入的了解。因此,他们所看到的是如图 1-12 所示的计算机层次结构。

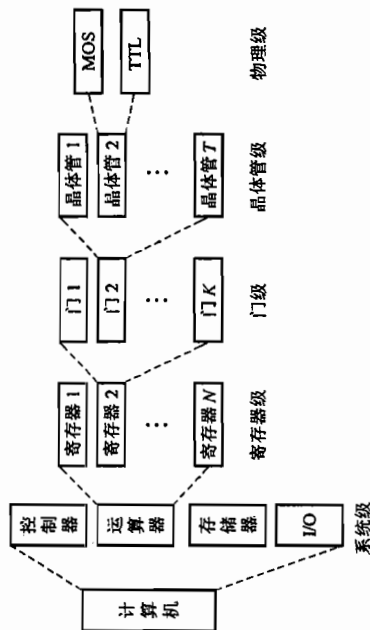


图 1-12 计算机硬件的层次结构

而对于一般工程技术人员等用户来说,所需要了解的是基于冯·诺依曼模型的计算机的主要组成(如图 1-13 所示)及其对计算机整机性能的作用。这也是本书所要讨论的主要内容。

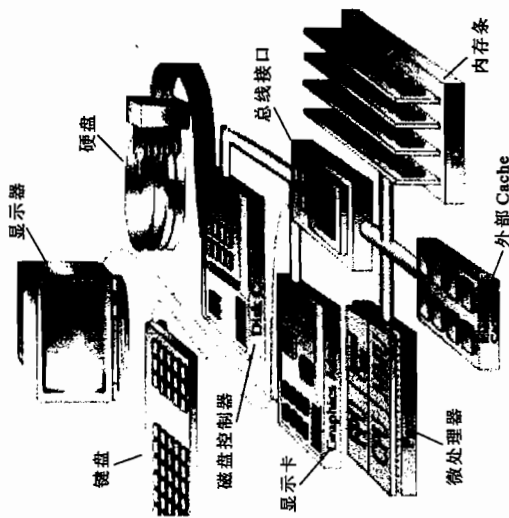


图 1-13 计算机的主要组成

1.4.3 计算机的模块结构

计算机系统由硬件和软件两大部分组成,可以用模块化结构表示,如图 1-14 所示。硬件部分的主要包括中央处理器、内/外存储器和输入/输出设备等。

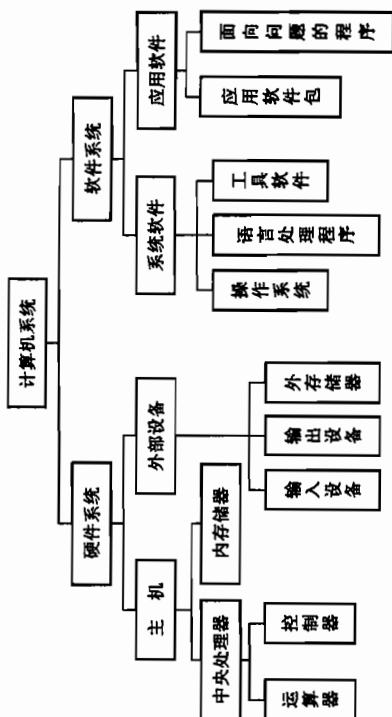


图 1-14 计算机的系统结构

1. 中央处理器

中央处理器(Central Processing Unit, CPU)是计算机系统的核心,包括运算器和控制器两部分。计算机所进行的全部操作都受 CPU 的控制。其中,运算器主要完成各种算术运算和逻辑运算,是对信息加工和处理的部件,由运算器件和用来暂存数据的寄存器、累加器等组成。控制器是对计算机发布命令的“决策机构”,用来协调和指挥整个计算机系统的操作,它本身不具有运算功能,而是通过读取各种指令,并对其进行翻译、分析,而后对各部件做出相应控制;它主要由指令寄存器、译码器、程序计数器、操作控制器等组成。CPU 是计算机的心脏,CPU 品质的高低直接决定了计算机系统的性能。能够处理的数据位数是 CPU 的一个最重要的品质标志。通常所说的 8 位、16 位或 32 位计算机即指 CPU 可同时处理 8 位、16 位或 32 位的二进制数据。

2. 内存存储器

内存存储器(Memory),简称内存,是微型计算机主机的组成部分,用来存放当前正在使用的或随时要使用的程序或数据。CPU 可以直接访问内存。微型计算机以字节为单位线性地组织内存,每个存储单元(一个字节)都有一个唯一的编号。24 位地址总线可以提供的地址为 $2^{24} = 16 \text{ MB}$ 。内存按其工作特点分为只读存储器(Read-Only Memory, ROM)和随机存取存储器(Random Access Memory, RAM)。

3. 外存储器

外部存储器简称外存,也称为辅存,是内存的延伸,其主要作用是长期存放计算机工作所需要的系统文件、应用程序、用户程序、文档和数据等。当 CPU 需要执行某部分程序或使用数据时,这些程序或数据由外存加载到内存中,供 CPU 访问。可见外存的作用是扩大存储系统的容量。目前最常用的外存有硬盘(包括可移动硬盘)、软盘、光盘、磁带和移动存储设备(U 盘)等。通常一台微型计算机至少安装一个硬盘驱动器、一个软盘驱动器和一个光盘驱动器。硬盘驱动器的特点是存储容量大、读写速度快、密封性好、可靠性高、使用方便。

4. 输入输出设备

输出设备主要包括显示设备、打印设备、音频设备和绘图仪等。

- 显示设备:包括显示器和显示适配器(俗称显卡),分辨率为其主要性能参数。
- 打印设备:有击打式(如点阵或针式)打印机和非击打式(如喷墨打印机、激光打印机等)打印机。

- 音频设备:主要包括声卡和音箱。

• 绘图仪:有笔式绘图仪、平板式绘图仪、滚筒式绘图仪、转筒式绘图仪、非笔式绘图仪、喷墨绘图仪、静电绘图仪、热敏绘图仪和电子照相绘图仪等。

输入设备主要包括键盘、鼠标、扫描仪和数码相机等。

- 文字输入设备:键盘、磁卡阅读器、条形码阅读器、纸带阅读器、卡片阅读机等。
- 图形输入设备:光笔、鼠标器、数字化仪、触摸屏等。
- 图像输入设备:扫描仪、数码相机、摄像头等。
- 音频处理设备:声卡。

这里仅对计算机硬件系统做模块化介绍。硬件组成各部分的工作原理、发展中的新技术、性能参数及其对计算机整机性能和功能的影响,将在后续章节中进一步介绍。

习题

- 1/ 计算机的类型有哪些?简述它们的特点。
- 2/ 微型计算机简单分哪几类?各有什么特点?
- 3/ 微型计算机的组成部件有哪些?
4. 简述微处理器的组成单元及各单元的功能和特点。
- 5/ 计算机的运算速度和存储容量怎么衡量?影响它们的主要因素有哪些?
6. 计算机的主要发展趋势有哪些?有什么特点和含义?
7. 简述摩尔定律的含义。
- 8/ 冯·诺依曼计算机的结构和特点是什么?
- 9/ 什么是计算机体系结构?其具体含义是什么?
10. 简述计算机的结构模块及各模块的功能与特点。

计算机系统组成及工作原理

第2章

1946年,美籍匈牙利数学家冯·诺依曼提出了关于计算机组成和工作方式的基本设想。迄今为止,尽管计算机制造技术已经发生了极大的变化,但是就其体系结构而言,仍然是以他的设计思想为基础的,这样的计算机称为冯·诺依曼结构计算机。冯·诺依曼设计思想可以简要地概括为以下3点:

- ① 采用二进制形式表示数据和指令。
- ② 采用存储程序方式。
- ③ 由运算器、存储器、控制器、输入装置和输出装置五大部件组成计算机系统,并规定了这五部分的基本功能。

冯·诺依曼设计思想最重要之处在于明确地提出了“存储程序”的概念,他的全部设计思想实际上是对“存储程序”概念的具体化。

本章以冯·诺依曼设计思想为核心,首先介绍与计算机相关的基础知识,包括二进制计数法与信息表示、逻辑电路基础等,最后讲述计算机的组成及工作原理。

2.1 二进制数与信息表示

数据是对事实、概念或指令的一种特殊表达形式,这种特殊表达形式可以用人工的方式或者用自动化的装置进行通信、翻译转换或者进行加工处理。一般的数字、文字、图像、声音和动画等都是数据,而信息是有用的数据。计算机信息处理的本质就是通过二进制编码形式对数据进行处理。采用二进制编码表示的数字、文字、图像、声音和动画才能由计算机进行处理。因此,了解二进制数的特点、二进制制与其他数制之间的转换关系以及信息编码等概念,对于学习和掌握计算机很有必要。

2.1.1 为什么要使用二进制

我们通常使用十进制数进行计算,而计算机却用二进制数计算,这是由于计算机由电子元件组成,而电子元件有“通”和“断”两种状态、信号有“有”和“无”两种情况,电

流有“正”和“负”两种方向。我们用二进制的“1”和“0”两个数码能很方便地分别表示这两种状态,因此,计算机采用二进制数来表示信息,这种设计最简单,工作也最为稳定。

此外,二进制数的“1”和“0”正好可与逻辑值“真”和“假”相对应,这样就为计算机进行逻辑运算提供了方便。

算术运算和逻辑运算是计算机的基本运算,采用二进制数可以简单方便地进行这两类运算。

2.1.2 信息在计算机内的表示

计算机内信息的表示形式是二进制数字编码。各种类型的信息(如数字、文本、音频、图像、视频等)必须转换成数字量即二进制数字编码的形式,才能由计算机进行处理,如图2-1所示。

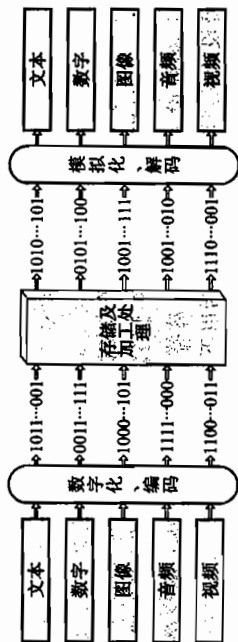


图 2-1 信息在计算机内的表示

计算机中处理的数据分为数值型数据和非数值型数据两大类。

2.1.3 数值型数据

数值型数据指能进行算术运算(加、减、乘、除四则运算)的数据,即我们通常所说的“数”。

数值型数据通常按进位的方法进行计数,称为进位计数制。也就是说数字累计到最大计数(进制数)时,数码长度就增加一位或数码高位数值增1。

一个K进制数N可表示为:

$$N = m_{n-1} \times K^{n-1} + m_{n-2} \times K^{n-2} + \dots + m_0 \times K^0 + m_{-1} \times K^{-1} + m_{-2} \times K^{-2} + \dots$$

其中, K^i 称为K进制第i位的权,简称位权; m_i 称为K进制第i位的系数,共K个。表2-1列出了十进制数、二进制数、八进制数、十六进制数以及十六进制数的表示形式。

表 2-1 常用数制

	二进制	八进制	十进制	十六进制	K进制
基数	2	8	10	16	K
进位	逢2进1	逢8进1	逢10进1	逢16进1	逢K进1
可用数码	0,1	0,1,2,3,4,5,6,7	0,1,2,3,4,5,6,7,8,9	0,1,2,3,4,5,6,7,8,9,A,B,C,D,E,F	0,1,2,3,4,5,6,7,8,9,...,K-1

不同进制的数字有不同的书写规则。通常有如下两种书写方式。

(1) 在数字后面加写相应的英文字母作为标识

B(Binary)——表示二进制数。例如,二进制数的100可写成100B。

O(Octonary)——表示八进制数。例如,八进制数的100可写成100O。

D(Decimal)——表示十进制数。例如,十进制数的100可写成100D。一般约定D可省略,即无后缀的数字为十进制数。

H(Hexadecimal)——表示十六进制数。例如,十六进制数100可写成100H。

(2) 在括号外面加数字下标

(1101)₂——表示二进制数的1101。

(3174)₈——表示八进制数的3174。

(6678)₁₀——表示十进制数的6678。

(2DF6)₁₆——表示十六进制数的2DF6。

1. 数制转换

不同进位的计数制只是描述数值的不同手段,它们之间可以相互转换,转换的原则是保证转换前后所表示的数值相等。

(1) 将K进制数转换为十进制数

其方法为按“权”展开,也就是按照各种进制的权值展开,求出系数与位权的乘积,然后对诸项乘积求和,即可得到转换结果。

例1 将二进制数(1011.101)₂转换为十进制数。

解:将二进制数按权展开如下:

$$(1011.101)_2 = 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 0 \times 2^{-2} + 1 \times 2^{-3} = (11.625)_{10}$$

其他进制数转换为十进制的方法与上例类似。

例2 将十六进制数(FA59)₁₆转换为十进制数。

$$\text{解: } (FA59)_{16} = 15 \times 16^3 + 10 \times 16^2 + 5 \times 16^1 + 9 \times 16^0 = (64089)_{10}$$

(2) 将十进制数转换成K进制

方法是先将整数部分和小数部分分别进行转换,然后再将它们合并起来。

整数部分转换:除以K,取余数。

小数部分转换:乘以K,取整数。

1) 十进制整数转换成K进制整数,采用逐次除以基数K取余数(除K取余)的方法。

① 将给定的十进制数除以K,余数作为K进制数的最低位。

② 把第一次除法运算所得的商再次除以K,余数作为次低位。

③ 重复②的步骤,记下余数,直至最后的商为0,最后的余数即为K进制的最高位。

2) 十进制纯小数转换成K进制小数,采取逐次乘以K,截取乘积的整数部分(乘K取整)方法。

① 将给定的十进制小数乘以K,截取其整数部分作为K进制小数部分的最高位。

② 把第一次乘积的小数部分再乘以K,所得积的整数部分为K进制的小数次高位。

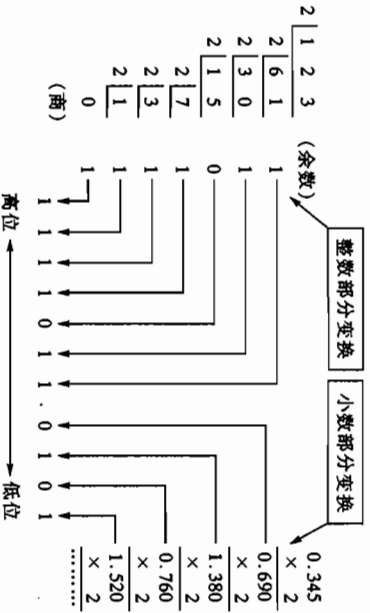
③ 依次进行下去,直至最后乘积为0。若最后乘积不会出现0,则直至达到要求的精度为止。

若要求精确到0.001(千分之一),取10位,因为 $1/2^{10}=0.00097$ 。

若要求精确到0.01(百分之一),取7位,因为 $1/2^7=0.0078$ 。
若要求精确到0.1(十分之一),取4位,因为 $1/2^4=0.0625$ 。

例3 将 $(123.345)_{10}$ 转换为二进制数。

解:



2. 汉字编码及处理技术

西文是拼音文字,基本符号比较少,编码比较容易,因此,在计算机系统中,输入、内部处理、存储和输出都可以使用同一代码。汉字种类繁多,编码比拼音文字困难,因此在不同的场合要使用不同的编码。目前使用的汉字编码主要有三种,包括 GBK、GB 2312 和 BIG-5。

(1) 汉字编码

1) GB 2312 全称为《信息交换用汉字编码字符集 基本集》,俗称国标码,由国家标准总局发布,1981年5月1日实施该编码,新加坡等地也使用此编码。这是一个简化字的编码规范,当然也包括其他的符号、字母、日文假名等,共7445个图形字符,其中汉字占6763个。通常说其中包含有6768个汉字,但其中有5个编码为空白,所以实际上总共有6763个汉字。

GB 2312 规定“对任意一个图形字符都采用两个字节表示,每个字节均采用7位编码表示”,习惯上称第一个字节为“高字节”,第二个字节为“低字节”。GB 2312 中汉字的编码范围:第一个字节为 0XB0~0XF7(对应十进制数为 176~247),第二个字节为 0XA0~0XFE(对应十进制数为 160~254)。

GB 2312 将代码表分为 94 个区,对应第一字节(0XA1~0XFE);每个区 94 个位(0XA1~0XFE),对应第二字节,两个字节的值分别为区号值和位号值加 32(20H),因此也称为区位码。01~09 区为符号、数字区,16~87 区为汉字区(0XB0~0XF7),10~15 区、88~94 区是有待进一步标准化的空白区。

2) BIG-5 又称大五码,主要在中国香港与中国台湾地区使用,是一种繁体字编码。每个汉字由两个字节构成,第一个字节的范围为 0X81~0XFE(即 129~254),共 126 种。第二个字节的范围不连续,分别为 0X40~0X7E(即 64~126)和 0XA1~0XFE(即 161~254),共 157 种。

3) GBK 也是一个汉字编码标准,全称为《汉字内码扩展规范》,由全国信息技术标准化技术委员会于 1995 年 12 月 1 日制定。GB 即“国标”,K 是“扩展”的汉语拼音第一个字母。

GBK 向下与 GB 2312 编码标准兼容,向上支持 ISO 10646-1 国际标准,是前者向后者过渡过程中的一个承上启下的标准。

ISO 10646-1 是国际标准化组织(ISO)公布的一个编码标准,即 Universal Multiple-Octet Coded Character Set(UCS),译为《通用多八位编码字符集》,而在我国台湾地区译为《广用多八位元编码字符集》。我国 1993 年以 GB 13000-1 国家标准的形式予以认可(即 GB 13000-1 等同于 ISO 10646-1)。

ISO 10646-1 是一个包括世界上各种语言的书面形式以及附加符号的编码体系。其中的汉字部分称为“CJK 统一汉字”(C 代表中国, J 代表日本, K 代表朝鲜)。而其中的中国部分,包括了源自 GB 2312、GB 12345、《现代汉语通用字表》等法定标准的汉字和符号以及源自我国台湾地区的 CNS 11643 标准中第 1、2 字面(基本等如同于 BIG-5 编码)、第 14 字面的汉字和符号。

GBK 亦采用双字节表示,总体编码范围为 8140~FEFE,第一字节在 81~FE 之间,第二字节在 40~FE 之间。总计 23 940 个码位,共收入 21 886 个汉字和图形符号,其中汉字(包括部首和构件)21 003 个,图形符号 883 个。

GB 2312、GBK 与 BIG-5 编码范围的比较如表 2-3 所示。

表 2-3 汉字编码范围

名称	第一节	第二节
GB 2312	0XB0~0XF7(176~247)	0XA0~0XFE(160~254)
GBK	0X81~0XFE(129~254)	0X40~0XFE(64~254)
BIG-5	0X81~0XFE(129~254)	0X40~0X7E(64~126) 0XA1~0XFE(161~254)

(2) 汉字输入码

输入码所解决的问题是如何使用西文标准键盘把汉字输入到计算机内。目前,汉字输入码有许多种,其中最常用的是拼音编码和字形编码。

拼音编码是按照汉语拼音规则来输入汉字的,不需要特殊记忆,只要会汉语拼音就可以输入汉字。但拼音输入法也有缺点:一是同音字太多,重码率较高,输入效率较低;二是对用户的发音要求较高;三是难于处理不识的生字。

字形编码是以汉字的形状确定的编码,即按汉字的笔画部件用字母或数字进行编码,如五笔字型、表形码便属于此类编码。字形编码的优点是重码少,不受方言干扰,只要经过一段时间的训练,输入汉字的效率会有大大的提高。现在社会上,大多数打字员都是用字形编码进行汉字输入。但字形编码的缺点是记忆的东西较多,需要经过一段时间学习才能掌握,长时间不用会忘掉。

(3) 汉字字库

字库就是我们使用计算机时显示或打印汉字的图像源。计算机调用字库显示汉字时,不是直接调用汉字的图像,而是调用这个汉字的内码,程序通过这个内码到相应的图像源(字库)当中寻找相应的图像信息,并绘制到屏幕上或者打印到纸上。汉字字库分点阵与矢量两种。

点阵字库是把每一个汉字都分成 16×16 或 24×24 个点,然后用每个点的虚实来表示汉字的轮廓,常作为显示字库使用。这类点阵字库汉字最大的缺点是不能放大,放大后文字边缘会出现锯齿。矢量字库保存的是对每一个汉字的描述信息,比如一个笔画的起始、终止坐标、半径、弧度等。在显示、打印这一类字库时,要经过一系列的数学运算才能输出结果。这一类字库保存的汉字理论上可以被无限地放大,即放大后汉字的笔画轮廓仍能保持圆滑,打印时使用的字库均为此类字库。Windows 使用的字库也为以上两类。在 FONTS 目录下,如果字体文件的扩展名为 FON,表示为点阵字库;扩展名为 TTF,则表示为矢量字库。

2.1.5 数据格式

计算机中常用的数据表示格式有两种,一是定点格式,二是浮点格式。一般来说,定点格式容许的数值范围有限,但要求的处理硬件比较简单。而浮点格式容许的数值范围很大,但要求的处理硬件比较复杂。

1. 定点数的表示方法

定点表示法约定计算机中所有数的小数点位置是固定不变的。由于约定在固定的位置,小

数点就不再使用记号“.”来表示。通常将数表示成纯小数或纯整数。通常把小数点固定在有效数位的最前面或最后面。

小数点的位置固定在最高有效数位之前,符号位之后,记为 $x_s, x_1, x_2, \dots, x_{n-1}, x_n$,这个数是一个纯小数。

小数点的位置固定在最低有效数位之后,记为 $x_s, x_1, x_2, \dots, x_{n-1}, x_n$,这个数是一个纯整数。小数点“.”在机器中的位置是隐含约定的,并不需要真正地占据一个二进制位。

定点数表示法如下图所示。

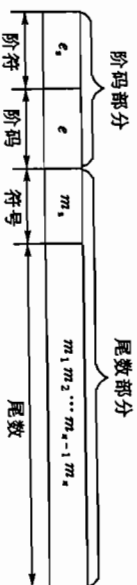


2. 浮点数的表示方法

在浮点表示法中小数点的位置不固定,而是根据需要浮动。任何一个数 n 的浮点表示形式为

$$n = m \times 2^e$$

式中, e 和 m 都是带符号的定点数, e 为阶码部分, m 为尾数部分。尾数为纯小数,阶码为纯整数。在大多数计算机中,浮点数的表示方法如下所示:



阶码的符号位为 e_s ,阶码的大小反映了在数 n 中小数点的实际位置;尾数的符号位为 m_s ,它是整个浮点数的符号位,表示了该浮点数的正负。

与定点表示法相比,在相同字长下,浮点表示法所能表示的数的范围要大得多,使用比较方便。

2.1.6 原码、补码和反码

1. 机器数与真值

在实际的计算过程中,参加运算的数大多是带符号的,既有正数,也有负数。通常在数字面前冠以“+”或“-”符号来表示正负。而在计算机中,数据都是存放在存储单元内,由二进制数码“1”和“0”来表示。对于“+”、“-”号,也只能用这两个数码(状态)来区别,这样,数的符号在计算机中就数码化了。

通常规定,在数的前面增设一位符号位,正数符号用“0”表示,负数符号用“1”表示。因此,在计算机中的数是用符号位和若干数值位来表示的。

例6 以8位字长的存储单元为例,假设在计算机中用8位二进制代码表示一个数+91。

由于91D=1011011B,所以其二进制格式为:

0 1011011

数字位1011011,表示数值的大小“91”;符号位“0”表示“+”。

用8位二进制代码表示一个数“-91”,则格式为:

1 1011011

数字位1011011,表示数值“91”;符号位“1”表示“-”。

在计算机内部,数字和符号都用二进制代码表示,两者合在一起,构成数的机内表示形式,称为机器数。也就是说,将已经数码化了的带符号数称为机器数,而把原来的数称为这个机器数的真值。

当然,机器数的形式是人们规定的。在计算机中为了运算方便,机器数常采用三种表示方法:原码、补码、反码。下面将分别进行讨论。

2. 原码表示法

原码表示法是一种最简单的机器数表示法,带符号数的原码表示形式与机器数形式完全相同,最高位为符号位,其余位为数值位。原码的表示法如下:

① 用最高位表示符号,符号位为“0”表示该数为正;符号位为“1”表示该数为负。

② 其有效数值部分用二进制数的“绝对值”表示。

例7 $X = +1101$, 则 $[X]_{\text{原}} = 01101$;

$X = -1101$, 则 $[X]_{\text{原}} = 11101$ 。

原码具有以下特点。

① “0”的原码有两种表示形式:

$$[+0]_{\text{原}} = 00000000, [-0]_{\text{原}} = 10000000。$$

② n 位带符号数的原码所表示的数的范围为: $+2^{n-1}-1 \sim -2^{n-1}-1$ 。

例如,8位带符号数的原码所表示的数的范围为: $+2^7-1 \sim -2^7-1$,即 $+127 \sim -127$ 。

01111111 ~ 11111111。

原码表示法简单易懂,但不便于进行算术运算。例如,两数相加时,必须首先判断符号位是否相同,若相同,做加法;若不相同,做减法。做减法时还须判断两数绝对值的大小,以便用大数减小数,最后还需给运算结果置以正确的符号。为了解决这一问题,产生了数的补码表示法,它可使负数的加法或减法运算简化为单一的加法运算。

3. 补码表示法

补码表示法的基本思想是使符号位参加运算,从而简化加、减法的运算规则,使减法运算转化成加法运算,从而简化机器的运算器。补码的表示法如下:

① 用最高位表示符号,符号位为“0”表示该数为正;符号位为“1”表示该数为负。

② 数值部分,正数的补码与原码相同;负数的补码,其数值部分各位按位取反,且在最低位上加1。

例8 $X = +1101$, 则 $[X]_{\text{补}} = 01101$;

$X = -1101$, 则 $[X]_{\text{补}} = 10011$ 。

补码具有以下特点。

显然这两个结果都是错误的。为什么会发生这种错误？计算机中的数据是有位数限制的，例如，5 位补码可以表示的十进制数的范围是从 -16 ~ +15。当所得的结果超过计算机所允许的额定位数时，就会出现错误。这种现象被称为溢出。

由于单个符号位只能表示正、负两种情况，当产生溢出时，符号位的含义就会发生混乱。如果将符号位扩充为两位(S₁、S₂)，其所能表示的信息量将随之扩大，既能判别是否溢出，又能指出结果的符号。两个符号位都作为数的一部分参加运算。双符号位(S₁、S₂)的含义如下：

- 00 结果为正数，无溢出
- 01 结果正溢出
- 10 结果负溢出
- 11 结果为负数，无溢出

由此可见，当两位符号位的值不一致时，表明产生溢出。为了尽可能减少实现代价，采用双符号位方案时，操作数和结果在寄存器和内存中仍保持单符号位，仅在运算时再扩充为双符号位。

2.2 逻辑电路

计算机采用二进制的“0”和“1”两个数码进行运算。实验上，计算机中的所有操作都是利用与二进制的“0”和“1”相对应的电路来实现的。换句话说，计算机通常以与“0”和“1”相对应的两种状态作为基准，进行所有的操作。如果只考虑这两种状态，则可运用代数来处理这种状态的变化，这种代数就称为逻辑代数(logical algebra)或布尔代数(Boolean algebra)。逻辑电路就是通过二极管、晶体管等元器件把布尔代数应用到电子电路上，以实现逻辑运算的电路。这种逻辑电路称为逻辑门电路。

2.2.1 有关逻辑电路的基础知识

1. 逻辑电路中的0和1

二进制中的0和1表示的是“数值”的0和1。在图2-2所示的电路中，0表示灯灭，1表示灯亮。因此，二进制中的0和1就可由电路的这两种状态表示。

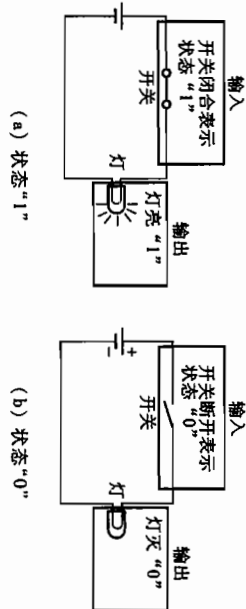


图2-2 电路示意

2. 逻辑代数

逻辑代数又称布尔代数，是英国数学家乔治·布尔(George Boole)为了研究思维规律(逻辑学、数理逻辑)，于1847年和1854年提出的数学模型，是表示二值逻辑的数学方法。逻辑代数是讨论逻辑关系的一门学科，它是分析和设计逻辑电路的数学基础。逻辑代数也是用字母表示变量，但是逻辑代数和普通代数有着根本的区别。逻辑代数中的逻辑变量只有两种可能取值0和1，而且这里的0和1不同于普通代数中的0和1。它只表示两种对立的逻辑状态，并不表示数值的大小。在逻辑运算中，基本的逻辑关系有“与”、“或”、“非”三种。在逻辑代数中，相应的也有三种基本运算，即“与”运算、“或”运算和“非”(求反)运算。逻辑运算的基本定律如表2-4所列。

表2-4 逻辑代数的基本定律和恒等式(基本公式)

表达式	名称	运算规律
$A+0=A$	0-1律	变量与常量的关系
$A \cdot 0=0$		
$A+1=1$		
$A \cdot 1=A$		
$A+A=A$	同一律	逻辑代数的特殊规律, 不同于普通代数
$A \cdot A=A$		
$A+\bar{A}=1$	互补律	
$A \cdot \bar{A}=0$		
$\bar{\bar{A}}=A$	还原律	
$A+B=B+A$		
$A \cdot B=B \cdot A$	交换律	与普通代数规律相同
$(A+B)+C=A+(B+C)$		
$(A \cdot B) \cdot C=A \cdot (B \cdot C)$	结合律	
$A \cdot (B+C)=A \cdot B+A \cdot C$		
$A+BC=(A+B)(A+C)$	分配律	逻辑代数的特殊规律, 不同于普通代数
$\overline{A+B}=\bar{A} \cdot \bar{B}$		

3. 真值表

将某事成功与否的所有条件，逐条分析、列举出来，所得的表格称为真值表。例如，表2-5列出了3个布尔表达式的真值表。真值表也表示了一种逻辑门电路所有可能的输入组合和相应的输出，从而定义了这种逻辑门电路的功能。

表 2-5 真值表

X = A · B · C				X = A + B + C			
输入条件		输出		输入条件		输出	
A	B	C	X	A	B	C	X
0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	1
0	1	0	0	0	1	0	0
0	1	1	0	0	1	1	1
1	0	0	0	1	0	0	0
1	0	1	0	1	0	1	1
1	1	0	0	1	1	0	1
1	1	1	1	1	1	1	1

4. 逻辑元件

逻辑电路是通过二极管和晶体管把布尔代数应用到电子电路上,来实现逻辑运算的。构成逻辑电路的这些元件称为逻辑元件。

(1) 二极管

二极管在正向电压作用下电阻很小,处于导通状态,相当于一只接通的开关;在反向电压作用下,电阻很大,处于截止状态,如同一只断开的开关,如图 2-3 所示。利用二极管的这种开关特性,可以组成各种逻辑电路。

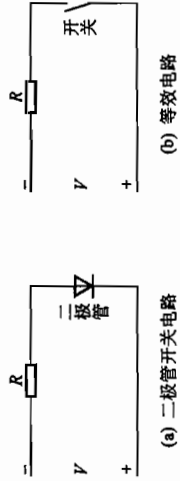


图 2-3 二极管开关电路及其等效电路

(2) 晶体管

晶体管可以工作在三个区域:放大区、截止区和饱和区。对应这三个区域,晶体管具有放大、截止和饱和三种工作状态。

在数字电路中,晶体管作为开关主要工作于截止和饱和两种状态,而放大状态是截止和饱和之间的过渡状态,它主要应用于模拟电路中。

截止状态:当输入电压 $V_1 < V_{BE}$ 时,发射结反偏, $I_B = I_C = I_E \approx 0$, $V_{CE} \approx V_{CC}$,集电结也反偏。C-E 间相当于开关断开,这种状态称晶体管的截止状态。

导通状态:当输入电压 $V_1 > V_{BE}$,发射结正偏, I_B 、 I_C

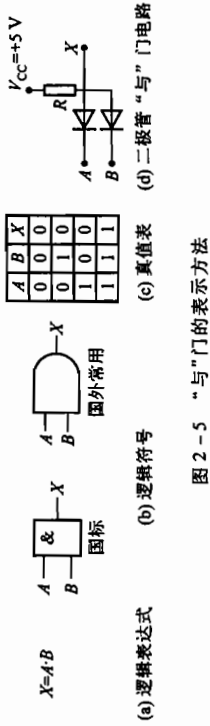
增大,输出电压 $V_{CE} = V_{CC} - I_C \times R_C$ 不断下降,降至 0.7 V 以下时,集电结也正偏,晶体管饱和,C-E 间相当于开关接通,称为晶体管的导通状态,如图 2-4 所示。

2.2.2 逻辑门电路

所谓“门”就是一种开关,在一定条件下它允许信号通过;若条件不满足,信号就通不过。因此,门电路的输入信号与输出信号之间存在一定的逻辑对应关系,所以门电路又称为逻辑门电路。

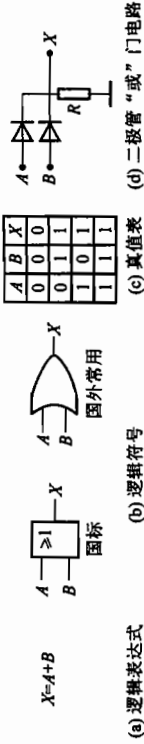
1. “与”门

“与”门和“与”逻辑关系对应。当所有条件都满足时,结果才成立;反之,有一个或一个以上条件不满足,结果就不成立,这种逻辑关系称为“与”逻辑关系。利用二极管管的开关特性可构成“与”门电路,图 2-5 是“与”门的表示方法和由二极管构成的“与”门电路。



2. “或”门

“或”门和“或”逻辑关系对应。当有一个或者一个以上条件得到满足时,结果就成立;仅当所有条件都不满足时,结果才不成立,这种逻辑关系称为“或”逻辑关系。图 2-6 是“或”门的表示方法和由二极管构成的“或”门电路。



3. “非”门

“非”门和“非”逻辑关系对应。“非”门接受一个输入值,生成一个与输入值相反的输出现值。图 2-7 展示了“非”门的表示方法和由晶体管构成的“非”门电路。

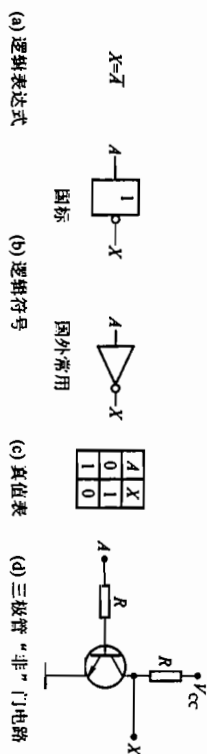


图 2-7 “非”门的表示方法

4. “与非”门

“与非”逻辑运算是先进行“与”运算再进行“非”运算的两级逻辑运算。“与非”逻辑运算的各种表示方法如图 2-8 所示。

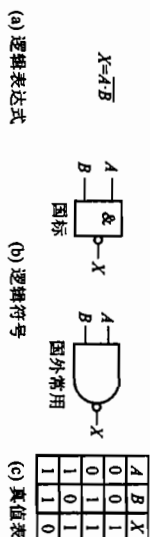


图 2-8 “与非”逻辑的表示方法

5. “或非”门

“或非”逻辑运算是先进行“或”运算再进行“非”运算的两级逻辑运算。“或非”逻辑运算的各种表示方法如图 2-9 所示。

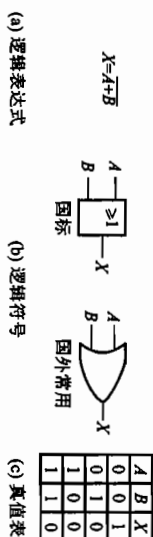


图 2-9 “或非”逻辑的表示方法

6. “异或”门

“异或”逻辑运算是两个输入值不同时,输出 1;否则,输出 0。“异或”逻辑运算的各种表示如图 2-10 所示。

2.2.3 组合逻辑电路

按照逻辑功能的不同,数字电路分成两大类,一类是组合逻辑电路,另一类是时序逻辑电路。

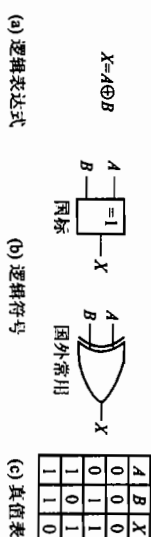


图 2-10 “异或”逻辑的表示方法

组合逻辑电路是具有有一组输入和一组输出的非记忆性逻辑电路,它的基本特点是任何时刻的输出信号状态仅取决于该时刻各个输入信号状态的组合,而与电路在输入信号作用前的状态无关,如图 2-11 所示。



图 2-11 组合逻辑电路

组合逻辑电路是由门电路组成的,但不包含存储信号的记忆单元,输出与输入间无反馈通路,信号是单向传输,且存在传输延迟。

由二极管与晶体管可构成“与非”门和“或非”门等组合逻辑电路,如图 2-12 所示。

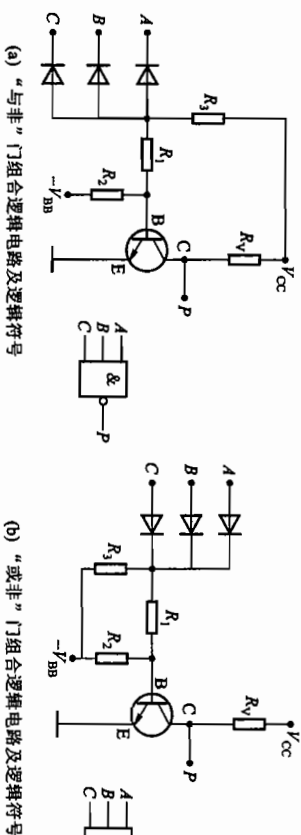


图 2-12 “与非”和“或非”门组合逻辑电路

组合逻辑电路的种类很多,主要有全加器、译码器、编码器、多路选择器等。下面分别加以介绍。

1. 二进制加法器

在数字系统中,算术运算都是利用加法进行的,因此加法器是数字系统中最基本的运算单元。由于二进制运算可以用逻辑运算来表示,因此可以用逻辑设计的方法来设计运算电路。在

数字系统中,加法分为全加和半加,所以加法器也分为全加器和半加器。

加数和被加数相加时,不考虑低位向本位的进位,这种加法叫做半加。

加数和被加数相加时,若考虑低位向本位的进位,则这种加法叫做全加。

(1) 半加器

在二进制加法运算中最简单的运算是1位加法,它只有下面4种组合:

$$0+0=0$$

$$0+1=1$$

$$1+0=1$$

$$1+1=0 \quad \dots\dots \text{产生进位 } 1$$

半加器不考虑低位向本位的进位,因此它有两个输入端和两个输出端。设加数(输入端)为A、B;和为S;向高位的进位为C。根据上述条件,可以画出图2-13(a)所示的真值表;推导出其逻辑表达式,如图2-13(b)所示;最后用逻辑电路(用“异或”门和“与”门构成)实现半加器,如图2-13(c)所示。

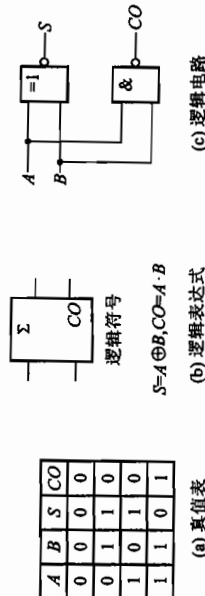


图 2-13 半加器

(2) 全加器

若两个二进制数相加时,在考虑本位的两个加数的同时,还考虑来自低位的进位数,则称为全加。实现全加运算功能的电路称为全加器。由于全加器考虑低位向高位的进位,所以它有3个输入端和两个输出端。图2-14所示是全加器的真值表、逻辑表达式及逻辑电路。

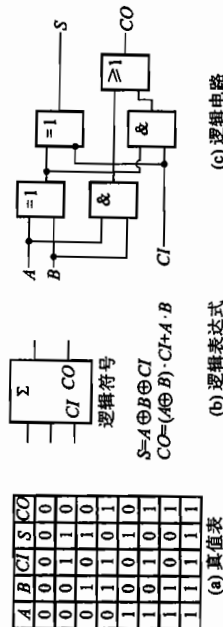


图 2-14 全加器

(3) 多位二进制加法电路

用全加器可以实现多位二进制加法运算。实现4位二进制加法运算的逻辑电路图如图2-15所示。

图中低位进位输出作为高位进位输入,依此类推,这种进位方式称为异步进位。

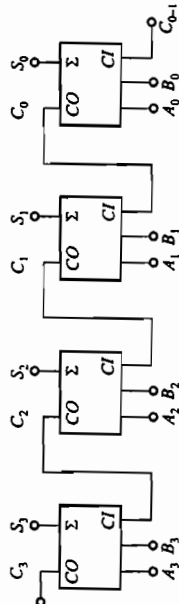


图 2-15 采用异步进位的4位二进制加法器逻辑图

异步进位方式中,进位信号是由后级向前级一级一级传输的。由于门电路具有平均传输延迟时间 t_{pd} ,经过n级传输,输出信号要经过 $n \times t_{pd}$ 时间才能稳定,即总平均传输延迟时间等于 $n \times t_{pd}$ 。所以,异步进位方式仅适用于位数不多,工作速度要求不高的场合。

2. 编码器和译码器

将特定的信号用二进制代码表示的过程叫做编码。将某一组二进制代码的特定含义译出的过程叫译码。

(1) 编码器

因为n位二进制数有 2^n 种状态,所以它可代表 2^n 组信息。在编码过程中一般采用编码矩阵和编码表。图2-16所示是一个3位二进制编码器。

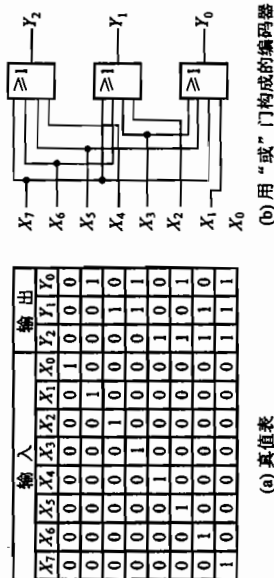


图 2-16 编码器

输入: $X_0 \sim X_7$ 为8个需要编码的信号。

输出: Y_2, Y_1, Y_0 为3位二进制代码。

由于该编码器有8个输入端,3个输出端,故称为8线-3线编码器。

(2) 译码器

译码是编码的逆过程,目的是将编码所表示的信息恢复为编码前的信息。实现译码功能的电路称为译码器。

译码器有多个输入端和多个输出端,每个输出端只对应于一种输出组合,即一种输入组合

(表示一个代码)用一个输出端来表示。假如输入端个数为 n ,则输出端个数最多有 2^n 个。

2.2.4 时序逻辑电路

时序逻辑电路是一种在任何时刻的输出不仅取决于该时刻电路的输入,而且与电路过去的输入有关的逻辑电路。组合逻辑电路在任一时刻的输出信号仅仅与当时的输入信号有关;而时序逻辑电路在任一时刻的输出信号不仅与当时的输入信号有关,而且与电路原来的状态有关。

从结构上看,组合逻辑电路仅由若干逻辑门组成,没有存储电路,因而无记忆功能;而时序逻辑电路除包含组合逻辑电路外,还含有存储电路,因而有记忆功能,如图2-17所示。

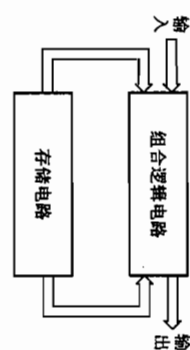


图 2-17 时序逻辑电路

在时序逻辑电路中,存储电路主要是触发器,是必不可少的,而组合逻辑电路是可选的。

时序电路具有以下特点:

- ① 除了组合逻辑电路以外,还包含有存储电路,具有记忆过去输入信号的能力。
- ② 存储电路的状态反馈到输入端,与输入信号共同决定其组合逻辑电路部分的输出。

2.2.5 典型时序电路——触发器

触发器是一种基本的时序电路,是能够存储一位二值信号的基本单元电路。触发器具备以下3个基本特点:

- ① 具有两个能自行保持的稳定状态,用来表示逻辑状态的0和1或二进制数的0和1。
- ② 根据不同的输入信号可以置成1或0状态。
- ③ 在输入信号消失以后,能将获得的新状态保存下来。

触发器的分类方式有很多种,按电路结构可分为:基本RS触发器、同步触发器、主从触发器、边沿触发器(包括维持阻塞触发器)等。不同电路结构的触发器有不同的动作特点。按逻辑功能可分为:RS触发器、D触发器、JK触发器、T和T'触发器等几种类型。

1. 基本RS触发器

基本RS触发器的电路如图2-18(a)所示。它由两个“与非”门按正反馈方式闭合而成。图2-18(b)是基本RS触发器逻辑符号。它有两个输入端R和S,两个输出端Q和 \bar{Q} ,符号R和S分别表示复位和置位输入。复位端R加低电平,即 $R=0$,“与非”门 G_2 输出“1”;置位端S加低电平,即 $S=0$,“与非”门 G_1 输出“1”。因此,复位和置位输入信号中只能有一个为低电平,低电平加在复位端,触发器复位, $Q=0$;低电平加在置位端,触发器置位, $Q=1$ 。所以,基本RS触发器称为置位复位触发器。在逻辑符号中用“0”表示低电平是有效信号。

通过在R、S端加不同的信号,同时考虑Q和 \bar{Q} 的状态,可得其真值表,如图2-18(c)所示。当 $R=S=0$ 时,两个输出端Q和 \bar{Q} 都为1,破坏了触发器输出互补的逻辑关系,特别当 $R=S=0$ 信号同时消失时,由于 G_1 、 G_2 门的传输延迟时间的不确定,触发器的新状态也是不确定的,故在真值表中描述为不定状态。这种情况应当避免。

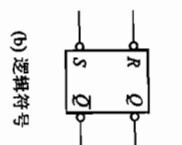
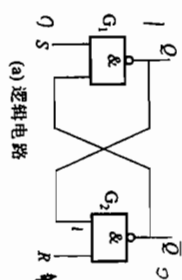


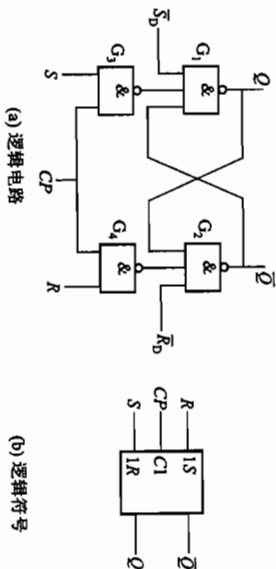
图 2-18 基本RS触发器

输入	输出
S	Q_{n+1}
R	功能
0	× 不定
0	0 置0
1	0 1 置1
1	1 Q_n 不变

(c) 特征表

2. 同步RS触发器

同步RS触发器是由一时钟脉冲信号CP控制的RS触发器。当要求触发器状态不是单纯地受R、S端信号控制,还要求按一定时间节拍把R、S端的状态反映到输出端时,就必须再增加一个控制端。只有控制端出现脉冲信号时,触发器才动作;至于触发器输出为什么状态,仍然由R、S端的高低电平来决定。采用这种触发方式的触发器,称为同步RS触发器,如图2-19所示,它由“与非”门构成,图(b)为其逻辑符号。



(b) 逻辑符号

图 2-19 同步RS触发器

S	R	Q_n	Q_{n+1}
0	0	0	0
0	0	1	1
0	1	0	0
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	不定
1	1	1	不定

(c) 特征表

在图2-19(a)中, G_1 、 G_2 门构成基本RS触发器, G_3 、 G_4 门组成控制电路,CP是控制脉冲。所谓同步就是触发器状态的变化与时钟脉冲同步。当 $CP=0$ 时, G_3 、 G_4 门被封锁,R、S信号不能进入, G_3 、 G_4 门输出均为高电平,则触发器输出保持原来状态;当 $CP=1$ 时,R、S信号才能经过 G_3 、 G_4 门影响到输出。S为直接置1端,R为直接置0端,它们的电平可以不受CP信号的控制而直接影响到触发器的输出。利用基本RS触发器的真值表,可得同步RS触发器的功能表,如图2-19(c)所示。因为CP脉冲的加入,要考虑CP脉冲作用前后Q端的状态,所以将CP脉冲作用前Q端的状态用 Q_n 表示,称为触发器的原状态;CP脉冲作用后Q端的状态用 Q_{n+1} 表示,称为触发器的次状态。将这种考虑了CP脉冲作用前后Q端状态的表格称为特征表或功能表。

3. JK触发器

由同步RS触发器的特性表(图2-19(c))可知,当 $R=S=1$ 时,触发器输出状态不定,须避免这种情况出现,这给使用带来不便。为此引入JK触发器,从电路设计上避免出现这种

情况。

考虑到RS触发器的输出 Q 和 \bar{Q} 互补的特点,将输出 Q 和 \bar{Q} 反馈到输入端,通过两个“与门”使加到 R 和 S 端的信号不能同时为1,从而满足RS触发器要求的约束条件。为区别于原来的RS触发器,将对应于原图中的 R 用 K 表示, S 用 J 表示,如图2-20所示。这种改接后的电路称为主从JK触发器。其逻辑符号如图2-20(b)所示。

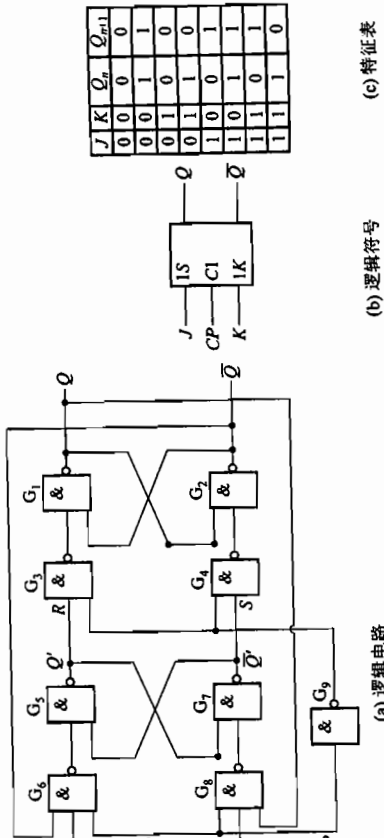


图 2-20 主从 JK 触发器

4. D 触发器

JK 触发器功能较完善,应用广泛,但需两个输入控制信号(J 和 K)。如果在 JK 触发器的 K 端前加上一个“非”门,再接到 J 端,如图 2-20(a)所示,可使输入端只有一个,在某些场合用这种电路进行逻辑设计可使电路得到简化。将这种触发器的输入端符号改用 D 表示,称为 D 触发器。由 JK 触发器的特征表可得 D 触发器的特征表如图 2-21(c)所示。

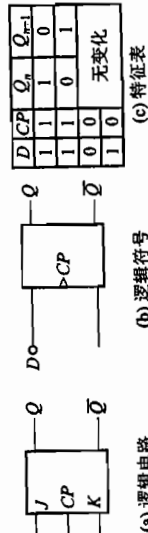


图 2-21 D 触发器

5. T' 触发器

T' 触发器又称受控翻转型触发器。这种触发器的特点很明显: $T=0$ 时,触发器由 CP 脉冲触发后,状态保持不变; $T=1$ 时,每进来一个 CP 脉冲,触发器状态就改变一次。由 JK 触发器转换而来的 T' 触发器如图 2-22 所示。

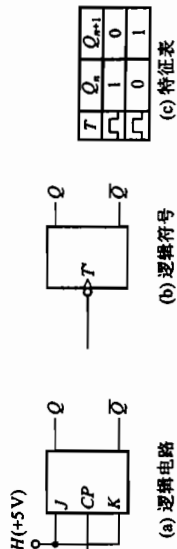


图 2-22 T' 触发器

2.2.6 计数器

计数器是常用的时序逻辑电路之一,如图 2-23 所示,它主要由具有记忆功能的触发器构成。计数器不仅仅用来记录脉冲的个数,还大量用于分频、程序控制及逻辑控制等,在计算机及各种数字仪表中,都得到了广泛的应用。

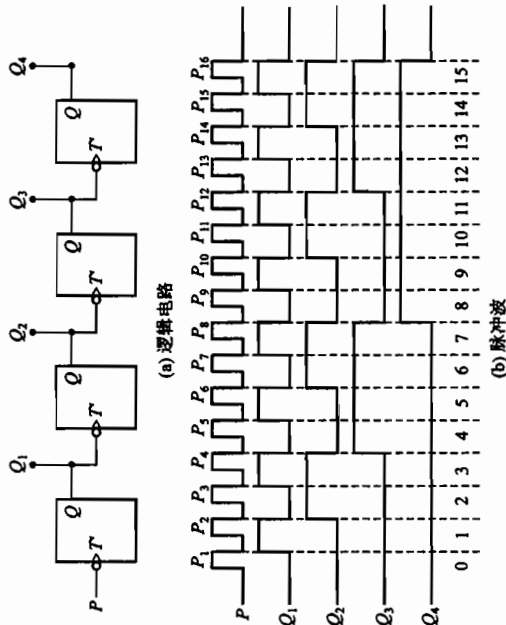


图 2-23 计数器

计数器对输入的时钟脉冲进行计数,每输入一个 P 脉冲,计数器状态变化一次。根据计数器的计数循环长度 M ,可将某种计数器称为模 M 计数器(M 进制计数器)。通常,计数器的状态按二进制制数的递增或递减规律来编码,相应的计数器称为加法计数器或减法计数器。

一个计数型触发器就是一位二进制计数器。 N 个计数型触发器可以构成同步或异步 N 位二进制加法或减法计数器。当然,计数器状态编码并非必须按二进制制的规律编码,可以给 M 进制计数器任意地编排 M 个二进制码。

2.2.7 寄存器

寄存器是一种接收、存储和输出二进制数码或信息的逻辑部件。按寄存器的功能不同,可分为数码寄存器和移位寄存器。

(1) 数码寄存器

数码寄存器——存储二进制数码的时序电路组件,具有接收和寄存二进制数码的功能。前面介绍的各种触发器,就是一种可以存储一位二进制数码的寄存器,用 n 个触发器就可以存储 n 位二进制数。

图 2-24 所示是由 D 触发器组成的 4 位寄存器的逻辑电路图。 $D_0 \sim D_3$ 是并行数据输入端, $Q_0 \sim Q_3$ 是并行数据输出端。

CP 为时钟脉冲, $Q_0 \sim Q_3$ 是并行数据输出端。该电路的数码接收过程:将需要存储的 4 位二进制数码送到数据输入端 $D_0 \sim D_3$,在 CP 端送一个时钟脉冲,脉冲上升沿作用后,4 位数码并行地出现在 4 个触发器的 Q 端。

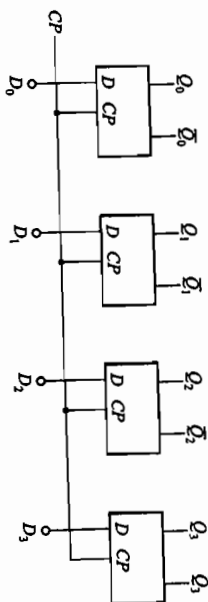


图 2-24 D 触发器组成的 4 位寄存器

(2) 移位寄存器

移位寄存器不但可以寄存数码,而且在移位脉冲作用下,寄存器中的数码可根据需要向左或向右移动 1 位。移位寄存器也是数字系统和计算机中应用很广泛的基本逻辑部件。

图 2-25 所示是由 D 触发器组成的 4 位右移寄存器。设移位寄存器的初始状态为 0000,串行输入数码 $D_1 = 1011$,从高位到低位依次输入。在 4 个移位脉冲作用下,输入的 4 位串行数码 1011 全部存入寄存器中。

移位寄存器中的数码可从 Q_3, Q_2, Q_1 和 Q_0 并行输出,也可从 Q_3 串行输出。串行输出时,要继续输入 4 个移位脉冲,才能将寄存器中存放的 4 位数码 1011 依次输出。

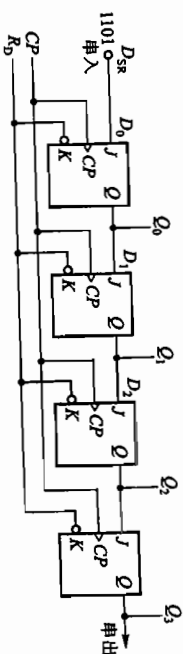


图 2-25 4 位右移寄存器

2.3 计算机的硬件组成

计算机就是由上述逻辑电路构成的,为完成特定任务进行二进制运算的、复杂的集合体,也就是说,计算机是由众多的、具有各种各样功能的逻辑电路汇集而成。前面提到的加法器、计数器、寄存器都是构成计算机的基本逻辑电路。

计算机一开始是作为一个计算工具出现的。如能脱离人的干预而自动地完成计算,它必须具备以下几个基本部件:①要有能进行运算的部件——运算器;②要有能记忆原始数据、运算程序及运算结果的部件——存储器;③要有能发出各种控制信息,以使计算机各部件协调工作的部件——控制器。④要有能将原始数据及运算程序输入计算机,并将结果及其他信息输出的部件——输入、输出设备。典型的计算机硬件系统就是由这五大部件构成,这种模型最早是由数学家冯·诺依曼提出的,故称为冯·诺依曼体系结构。冯·诺依曼体系结构的计算机是以存储程序方式工作的,即在控制器的控制下,计算机的各个部分根据预先编制的程序自动连续地进行工作。冯·诺依曼体系结构的计算机如图 2-26 所示。

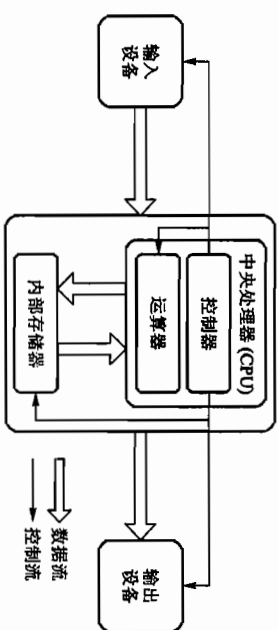


图 2-26 冯·诺依曼体系结构的计算机

运算器的功能是实现算术和逻辑运算,这是通过其内部的算术逻辑部件 (Arithmetic and Logic Unit, ALU) 来实现的。控制器发出各种控制信号,使整个运行过程自动进行。控制器和运算器合称为中央处理器 (Central Processing Unit, CPU)。存储器由记忆单元组成,用于存放数据、中间结果及一系列指令。输入/输出设备称为外部设备 (简称外设),用于输入原始数据、控制命令及输出运行结果等。

在计算机中,有两股信息在流动。一股是数据流,各种原始数据由输入设备输入至存储器中,在运算过程中,数据从存储器读入运算器进行运算,运算的中间结果存入存储器中或由外设输出。另一股信息流为控制信息流,人们给计算机的各种命令 (即程序) 也以数据的形式由存储器送入控制器。由控制器译码后发出控制信号,控制计算机各部件的工作。

2.3.1 CPU 的组成

CPU 是使整台计算机能够运行的最核心、最重要的组件。当计算机系统开始运行时,CPU

从内存中读取操作指令与数据,通过运算器运算出结果后存回内存,然而通过主机板上的I/O设备输出运算结果,从而达到数据处理的目的。

CPU 为控制整台计算机运作的中心枢纽,其内部包括控制器、运算器、寄存器组等。运算器是计算机对数据进行加工处理的中心,它主要由算术逻辑部件、累加器、标志寄存器和寄存器组成。控制器是计算机的控制中心,它决定了计算机运行过程的自动化,它不仅要保证程序的正确执行,而且要能够处理异常事件。控制器一般包括指令寄存器、译码器、操作控制器等。图 2-27 是一个典型的 CPU 组成示意图。

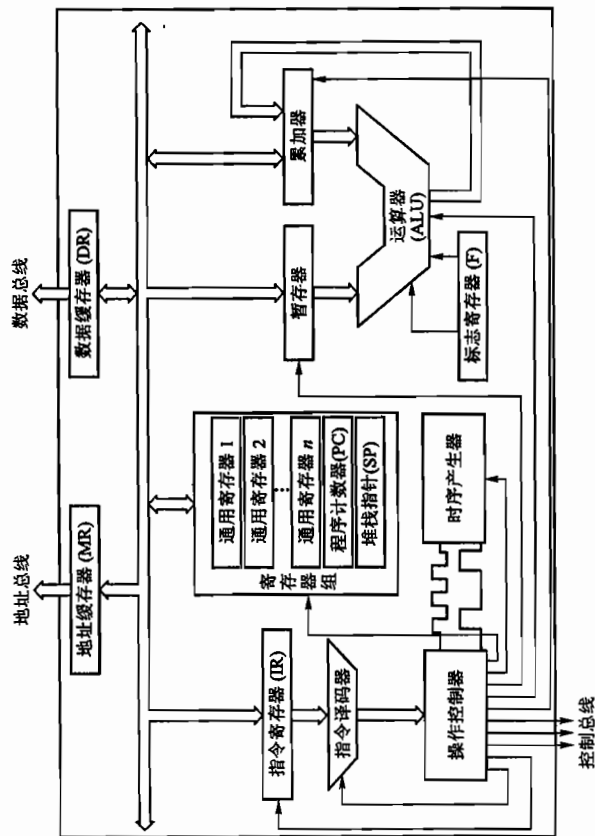


图 2-27 CPU 组成示意图

运算器(ALU)主要完成对二进制信息的定点算术运算、逻辑运算和各种移位操作。

通用寄存器组用来保存参加运算的操作数和运算的中间结果。

标志寄存器用以指示运算结果状态(如加法运算进位时置 1 进位标志等)和控制工作条件(如设置中断屏蔽标志等)。标志寄存器在不同的机器中有不同的规定,标志寄存器中的状态位通常作为转移指令的判断条件。

累加器是一个具有特种功能的寄存器,它虽叫累加器,但并不进行加法运算,主要用来临时存储运算器运算过程的结果和其他数据,并能把存在其中的数据左移或右移。

若要使计算机能够按程序要求的顺序自动执行程序,就必须把存在存储器中的指令按顺序依次取出来并加以执行,因此,必须有一个电路能追踪、记忆指令所在的地址,这个电路就是程序计数器(Program Counter, PC)。计算机加电时,计算机的复位电路自动给程序计数器赋以程序中

的第一条指令所在地址,计算机根据这一地址取出第一条指令。程序计数器的内容自动加该指令的长度(字节数),以指向下一条指令的地址;计算机处理完一条指令后,根据程序计数器记忆的地址,取下一条指令,程序计数器再自动加该指令的长度。当遇到跳转指令、调用子程序指令或遇到中断时,程序计数器内容自动转到所需的地方。

指令寄存器用来保存当前正在执行的那条指令。当要执行一条指令时,先从存储器取出该指令,并送至指令寄存器(Instruction Register, IR),该寄存器的输出就是指令译码器(Instruction Decoder, ID)的输入,经指令译码器译码(识别这条指令的功能)后,即可向操作控制器发出具体操作的特定信号。

从上述可知,CPU 内的每个功能部件都完成特定的功能。然而,信息怎样才能在各部件之间传送呢?也就是说,数据的流动是由什么部件控制的呢?通常把许多数字部件之间传送信息的通路称为“数据通路”。信息从什么地方开始,中间经过哪个寄存器或多路开关,最后传到哪个寄存器,都要加以控制。在各寄存器之间建立数据通路的任务,是由称为“操作控制器”的部件来完成的。操作控制器的功能就是根据指令操作码和时序信号,产生各种操作控制信号,以便正确地建立数据通路,从而完成取指令和执行指令的控制。

计算机时刻在高速地工作,每一个动作的时间是非常严格的,不能有任何差错。时序产生器的作用,就是对各种操作实施时间上的控制。

2.3.2 指令系统

指令系统指的是 CPU 所能够处理的全部指令的集合,是 CPU 的根本属性。指令系统决定了 CPU 能够运行什么样的程序。不同的 CPU,不同的指令集,构成了不同的指令系统。指令系统是计算机硬件和软件之间的桥梁,是汇编程序设计的基础。指令系统包含了许多执行各种类型操作的指令,每条指令完成一种特定的操作。计算机系统完成的每项工作,往往需要执行一系列相应的指令。

所有采用高级语言编写的程序,都需要翻译(编译或解释)成为机器语言后才能运行,这些机器语言中所包含的就是一条条的指令。

1. 指令的格式

一条指令一般包括两个字段:操作码和地址码。操作码其实就是指令序列号,用来告诉 CPU 需要执行的是哪一条指令。地址码则复杂一些,主要包括源操作数地址、目的操作数地址和下一条指令的地址。在某些指令中,地址码可以部分或全部省略,比如一条空指令就只有操作码而没有地址码。

(1) 指令的操作码

指令的操作码字段表明了所执行的操作。每一条指令都有一个唯一确定的操作码。指令的操作码经过译码产生执行本指令操作所需的微操作控制信号,控制计算机的各部件完成规定的操作。如果指令系统共有 m 条指令,指令中操作码字段的位数为 N 位,则有如下关系式: $m \leq 2^N$ 或 $N \geq \log_2 m$ 。

(2) 指令的地址码结构

对于一般的指令来说,除操作码之外,指令中还应包含以下信息:

- 参加运算的操作数存放的地址。
- 执行某种操作后的结果存放的地址。

指令的地址码字段的结构又分为三地址指令、二地址指令和一地址指令等几种。

2. 指令的分类

一台计算机的指令系统可以有上百条指令,这些指令按其功能可以分成以下几种类型。

① 数据传送类指令:CPU 在进行算术和逻辑操作时,总需要操作数,所以数据的传送是一种最基本、最主要的操作。数据传送类指令用于控制数据信息在计算机各组成部件之间的传送操作。具体可实现寄存器与寄存器之间、寄存器与存储器之间以及寄存器与 I/O 端口之间的数据传送操作。

由于数据在传送过程中不发生改变,没有产生新的结果,因此,这类指令执行时,不会影响标志寄存器。

② 算术逻辑运算类指令:多数计算机具有加、减、乘、除四种基本的算术运算指令和“与”、“或”、“非”、“异或”四种基本的逻辑运算指令。带有浮点运算部件的计算机还可以具有浮点运算指令。

③ 程序控制类指令:用于控制程序的执行方向,使程序具有测试、分析与判断的能力。包括转移控制指令、子程序调用和返回指令、中断控制指令等。

④ 输入输出类指令:输入输出(I/O)类指令用来实现主机与外部设备之间的信息交换。

3. 寻址方式

寻址方式是指令系统设计的重要内容,它对于丰富程序设计手段、方便程序编制、提高程序的质量、减少访问主存的次数以及压缩程序占用的内存空间等,都起着重要作用。每台计算机的指令系统都有自己的一套寻址方式。

① 立即寻址:指令所需操作数就在指令码中,操作码后面一个(或多个)字节就是操作数,即指令中的形式地址字段提供的就是操作数本身。

② 寄存器寻址:指令所需操作数在 CPU 内部寄存器中,指令中形式地址字段提供的是存放操作数的寄存器编码。由于寄存器的编码短,采用寄存器操作数,不仅可以减少指令的长度,而且执行指令时,全部操作在 CPU 中完成,不需使用总线操作周期来存取操作数,这就使得指令执行速度较快。这种寻址方式主要用来存取位于寄存器中的数据。

③ 直接寻址:指令所需操作数在某存储单元中,指令中形式地址字段给出的就是该存储单元的地址。

④ 寄存器间接寻址:指令所需操作数在存储单元之中,采用这种寻址方式时,应事先把操作数的偏移地址存入一个寄存器(BX、BP、SI 或 DI)中。寄存器间接寻址是把形式地址字段指定寄存器的内容作为操作数的有效地址,由该地址所指定的存储单元的内容作为操作数。如果指令中指定的寄存器是 BX、SI 或 DI,则操作数在数据字段中。

在寄存器间接寻址中,寄存器的内容如同一个地址指针。如果寄存器的内容在程序运行期间进行修改,那么使用这种寻址方式的同一指令可以对不同存储单元进行操作。因此,这种寻址方式适用于数组变量或表格处理。

2.3.3 CPU 的工作过程

CPU 的工作过程实际上就是指令的执行过程,图 2-28 是一条加法指令“ADD A, [3000H]”的执行过程。

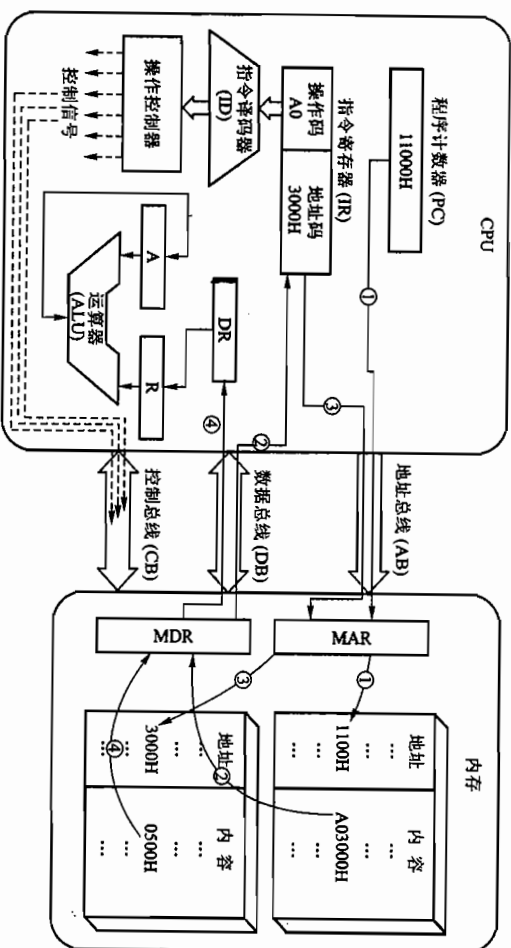


图 2-28 CPU 的工作过程

第一步,取指令。这一步的任务是将指令(包含操作码和地址码)取出来并送到控制器的指令寄存器(IR)。取指令的数据传输路径如图 2-28 中①、②所示。

1) PC→MAR:将程序计数器(PC)中存放的指令地址送入内存地址寄存器(MAR)。控制器发出的相应控制命令为:PC(1100)→AB, AB→MAR。

2) M→MDR:将指定的存储单元的内容 A03000H(加法指令)送至内存数据寄存器(MDR)。控制器发出的相应控制命令为:IR。

3) MDR→IR:将数据寄存器(MDR)中的指令送到指令寄存器(IR)。控制器发出的相应控制命令为:MDR(A03000H)→DB, DB→IR。

4) PC+1→PC:程序计数器(PC)自动加 1,为取下一条指令做好准备。控制器发出的相应控制命令为:PC+1。

第二步,取加数。指令中的操作码经指令译码器译码后识别出是加法运算,接着从内存中取出加数。取操作数的数据传输路径如图 2-28 中③、④所示。

1) 3000H→MAR:指令中的地址码 3000H 经地址形成部件产生操作数地址送到地址寄存器。控制器发出的相应命令。

2) M→MDR:将指定的存储单元的内容 0500H(加数)送至内存数据寄存器(MDR)。控制

器发出的相应控制命令为: R。

3) $MDR \rightarrow R$: 数据寄存器的内容送到运算器的暂存寄存器(R)。控制器发出的相应控制命令为: $(MDR) \rightarrow DB, DB \rightarrow R$ 。

第三步, 求和数。把累加器 A 中的被加数和暂存寄存器 R 中的加数送入加法器, 最后把和数送入累加器 A, $A(A) + (R) \rightarrow A$ 。

习 题

一、选择题

- 与十进制 56 等值的二进制数是()。
A. 111000 B. 111001 C. 101111 D. 110110
- 二进制数 111010011 转换成十六进制数是()。
A. $(323)_{16}$ B. $(1D3)_{16}$ C. $(133)_{16}$ D. $(3D1)_{16}$
- 16 位计算机中整数的有效范围是()。
A. $-32\ 768 \sim 32\ 767$ B. $-32\ 767 \sim 32\ 767$ C. $0 \sim 65\ 535$ D. $-32\ 768 \sim 32\ 767$ 或 $0 \sim 65\ 535$
- 带 + 或 - 号的数, 如 $N1 = +1001010$, 称为()。
A. 无符号数 B. 真值 C. 浮点数 D. 机器数
- 计算机内部的所有信息都采用()。
A. 二进制形式表示 B. 十六进制形式表示 C. 十进制形式表示 D. 多种进制数表示
- CPU 包括()。
A. ALU 和累加器 B. ALU 和控制器 C. 运算器和控制器 D. ALU 和主存储器
- 用以指定待执行指令所在地址的是()。
A. 指令寄存器 B. 数据计数器 C. 程序计数器 D. 累加器
- 在()表示中, 二进制数 11111111 表示十进制数 -1。
A. 原码 B. 反码 C. 补码 D. 移码(增码)
- 原码 -0 的反码是()。
A. +0 B. -127 C. 0 D. +127
- 若使用双符号位, 则发生正溢的特征是双符号位为()。
A. 00 B. 01 C. 10 D. 11
- 一个逻辑门电路, 若有任一输入为 1 时, 其输出为 0, 则此逻辑门电路位为()。
A. 异或门 B. 与门 C. 或非门 D. 与非门

非门

12. 下列描述中正确的是()。

- A. 控制器能理解、解释并执行所有的指令及存储结果
B. 一台计算机包括输入、输出、控制、存储及算术逻辑运算 5 个单元
C. 所有的数据运算都在 CPU 的控制器中完成
D. 以上答案都正确

D. 或门

13. 以下不属于“存储程序”计算机体系结构(即冯·诺依曼体系结构)应具备的特征是()。

- A. 将编好的程序先存入存储器中, 然后再启动计算机
B. 计算机内部采用二进制数
C. 计算机硬件由 5 大基本部件组成
D. 所有计算机都采用同样的指令系统

14. 运算器由很多部件组成, 其核心部分是()。

- A. 数据总线 B. 多路开关
C. 算术逻辑单元 D. 累加器

15. 某种微型计算机内存取字节编址, 每执行一条指令, 程序计数器()。

- A. 自动加 1 B. 保持不变
C. 自动增 2 D. 自动增加本指令的字节数

二、填空题

- 计算机中的指令由()和()两部分组成。
- 加法器有两个输入数据, 一个来自(), 另一个来自()。
- CPU 的两个组成部分是()和()。加法器有两个输入数据, 一个来自(), 另一个来自()。
- 设字长为 $N+1$ 位(含 1 位符号位), 补码的表示范围为()。
- 微型计算机中使用最普遍的英文字符编码是()。
- 十进制数 36.875 转换成二进制数是()。
- 通常, 约定二进制数的最高位为()位, “0”表示(), “1”表示()。
- 在计算机中常用的两种表示小数的方法, 一种是()表示法, 另一种是()表示法。
- ALU 用来执行当前指令所规定的()和()。
- 根据产生微操作控制信号的方式不同, 控制器可分为()和()。

三、简答题

- 计算机的硬件是由哪些部件组成的? 它们各有哪些功能?
- 简述 ALU 的基本功能。
- 简述中央处理器的构成。
- 控制器可细分为哪些部分? 各部分的功能如何?
- 在浮点数中, 阶码的正负和尾数的正负各代表什么含义? 对实际数值的正负与大小有何影响?
- 控制电路有哪几种硬件实现方法? 各有何特点?
- 指令字中有哪些字段? 各有何作用?
- 指令执行的基本过程分哪几个步骤? 请简单说明各步骤完成的任务。

“个人计算机(Personal Computer,PC)”。

从第一代PC机IBM-PC/XT到今天,随着CPU处理能力的不断提高,PC机的性能有了飞速的提高,CPU的型号一直是决定PC机更新换代的主要标志。伴随着CPU从8088到Pentium 4的性能的不断提高,PC机也经历了7代的发展历程。表3-1列出了各代PC及所用的CPU。

表 3-1 各代PC及所用CPU

PC机	CPU	出现年代	晶体管数目
第一代	8086和8088	1978—1981	29 000
第二代	80286	1984	134 000
第三代	80386D和80386SX	1987—1988	275 000
第四代	80486SX、80486DX、80486DX2和80486DX4	1990—1992	1 200 000
第五代	Pentium	1993—1995	3 100 000
	Cyrix 6x86	1996	—
	AMD K5	1996	—
第五代改进型	IDT WinChip C6	1997	3 500 000
	Pentium MMX	1997	4 500 000
	IBM/Cyrix 6x86MX	1997	6 000 000
第六代	IDT WinChip2 3D	1998	6 000 000
	Pentium Pro	1995	5 500 000
	AMD K6	1997	8 800 000
第六代改进型	Pentium II	1997	7 500 000
	AMD K6-2	1998	9 300 000
	Pentium III Kalmi	1999	930 万
第七代	Pentium III CuMine	2000	2 800 万
	Pentium III Tualatin	2001	4 400 万
	Pentium 4 Willamette	2000	42 000 万
第七代改进型	Pentium 4 Northwood	2001	55 000 万
	Pentium 4 Prescott	2004	125 000 万
	Pentium 4 Smithfield	2005	230 000 万

1982年,Intel公司推出了著名的x86体系结构。直到今天,x86体系结构仍然是大多数Intel处理器的基础。现在所使用的CPU基本上都是x86系列及兼容CPU,x86系列即采用x86指令集的CPU。虽然随着CPU技术的不断发展,Intel陆续研制出更新型的i80386、i80486、Pentium,直到今天的Pentium 4系列,但为了保证微型计算机能继续运行以往开发的各类应用程序,以保护和继承丰富的软件资源,Intel公司所生产的所有CPU仍然继续使用x86指令集。另外,除Intel公司之外,AMD和Cyrix等厂家也相继生产使用x86指令集的CPU。由于这些CPU能运行所有为Intel CPU所开发的软件,所以微型计算机业内人士就将这些CPU列为Intel CPU的兼

第3章

微型计算机的中央处理器

中央处理器(CPU)是组成微型计算机的核心部件,承担着计算机的运算和控制任务。微型计算机的运算速度和性能在很大程度上是由CPU决定的。本章将在第2章的基础上进一步介绍CPU的组成结构、性能指标、发展趋势和CPU的关键技术。

3.1 CPU概述

CPU是Central Processing Unit(中央处理器)的缩写。在微型计算机中,CPU被集成在一块被称为微处理器(MPU)的大规模集成电路芯片上,因此,通常把用在微型计算机中的CPU称为微处理器。CPU由运算器和控制器组成。如果把计算机比作一个人,那么CPU就是心脏,其重要作用由此可见一斑。

微处理器被誉为20世纪最伟大的发明之一。1971年1月,Intel公司的工程师Ferdinando Faggin、Ted Hoff和Stan Mazor研制成功世界上第一块4位微处理器芯片Intel 4004,标志着第一代微处理器问世。微处理器和微型计算机时代从此开始。这一创举开始了人类将智能内嵌于无生命设备的历程。Intel 4004包含2 300个晶体管,尺寸规格为3 mm×4 mm,计算性能远远超过当年的ENIAC。第一块真正实用的微处理器是诞生于1974年的8080,采用8080的微型计算机很快流行开来,为Intel公司日后的发展打下了基础。几年之后,比8080更新一代的产品8086处理器面市。8086采用16位通用目标寄存器、16位数据总线以及20位地址总线,支持1 MB系统内存,内含29 000个晶体管。

虽然Intel公司在1971年就研制成功了Intel 4004 CPU,但第一代微型计算机却是在1981年出现的。1980年7月,IBM微型计算机技术总设计师埃斯特利奇(Don Estridge)领导“跳棋计划”的13人小组,秘密来到佛罗里达州波克顿镇的IBM研发中心,开始开发后来被称为IBM PC的产品。一年后的8月12日,IBM公司在纽约宣布第一台IBM PC诞生,开创了计算机历史的新篇章。第一台IBM PC采用了Intel 8088(18086简化版)微处理器,操作系统是Microsoft提供的MS-DOS。IBM将其命名为

CPU 管芯面积不能太大的情况下,一级高速缓存的容量不可能做得太大。服务器 CPU 的一级高速缓存的容量通常在 32 KB ~ 256 KB 之间。

二级高速缓存是 CPU 的第二层高速缓存,分内部和外部两种芯片。内部的芯片二级高速缓存运行速度与主频相同,而外部的二级高速缓存则只有主频的一半。二级高速缓存的容量也会影响 CPU 的性能,原则是越大越好。现在 PC 机 CPU 的二级高速缓存通常为 256 KB ~ 2 MB。

3.2.4 CPU 的扩展指令集

CPU 依靠指令来处理数据和控制系统。每款 CPU 在设计时就规定了一系列与其硬件电路相配合的指令系统。指令集的强弱也是 CPU 的重要指标。指令集是提高微处理器效率的最有效工具之一。从现阶段的主流体系结构来看,指令集可分为复杂指令集和精简指令集。Intel 的 MMX (Multi Media Extended)、SSE、SSE2 (Streaming-Single instruction multiple data-Extensions 2)、SSE3 和 AMD 的 3DNow! 等都是 CPU 的扩展指令集,分别增强了 CPU 的多媒体、图形图像和 Internet 等的处理能力。通常把 CPU 的扩展指令集称为“CPU 的指令集”。SSE3 指令集也是目前规模最小的指令集,此前 MMX 包含有 57 条指令,SSE 包含有 50 条指令,SSE2 包含有 144 条指令,SSE3 包含有 13 条指令。目前,SSE3 也是最先进的指令集,Intel Prescott 处理器已经支持 SSE3 指令集,AMD 会在未来双核处理器中加入对 SSE3 指令集的支持,全美达 (Transmeta) 的处理器也将支持这一指令集。

3.2.5 工作电压

工作电压指的是 CPU 正常工作所需的电压。早期 CPU (80286、80386 及 80486 时代) 的工作电压一般为 5 V,当时的制造工艺相对落后,以至于 CPU 的发热量太大,使得寿命缩短。随着 CPU 的制造工艺与主频的提高,近年来各种 CPU 的工作电压有逐步下降的趋势,以解决散热的問題。目前,一般台式计算机 CPU 工作电压已低于 2 V,而笔记本计算机专用的 CPU 工作电压就更低了,甚至达到 1.2 V。这样功耗就大大降低,而散热也大为减少,当然,其生产成本也就大幅度提高。

3.2.6 地址总线与数据总线的宽度

1. 地址总线宽度

地址总线宽度决定了 CPU 可以访问的物理地址空间,简单地讲就是 CPU 到底能够使用多大容量的内存。对于 32 位地址线的宽度,最多可以直接访问 4 096 MB (4GB) 的物理地址空间。

2. 数据总线宽度

数据总线负责整个系统的数据传输,而数据总线宽度则决定了 CPU 与二级高速缓存、内存以及输入/输出设备之间一次数据传输的信息量。

3.2.7 制造工艺

线宽是指芯片上的最基本功能单元——门电路的宽度。实际上,门电路之间连线的宽度同门电路的宽度相同,所以线宽可以描述制造工艺。缩小线宽意味着晶体管可以做得更小、更密

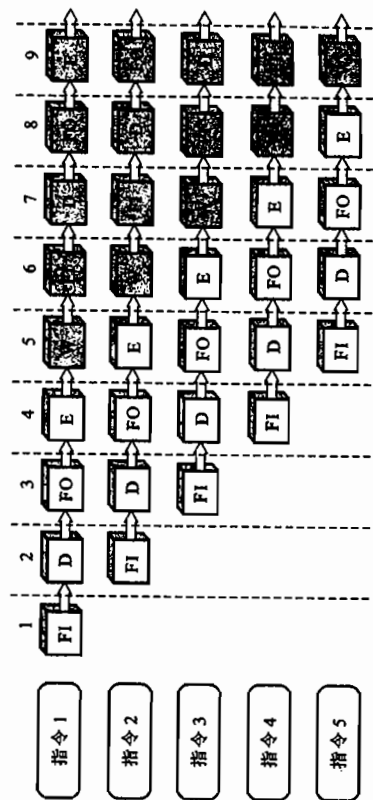
集,可以降低芯片功耗,系统更稳定;CPU 得以运行在更高的频率下,而且在相同的芯片复杂度下可使用更小的晶圆,于是成本降低了。随着线宽的不断降低,以往芯片内部使用的铝连线的导电性能将不敷使用,现在的处理器普遍采用导电特性更好的铜连线。Pentium CPU 的制造工艺是 0.35 μm , Pentium II 是 0.25 μm , Pentium III 是 130 nm, Pentium 4 所采用的制造工艺是 90 nm,而最新的 Pentium 4 已达到了 65 nm 的制造工艺。

3.3 提高 CPU 性能先进技术

3.3.1 流水线与超标量结构

在早期的计算机系统中,无论是指令的执行还是数据的处理,都是严格按照串行的方式来进行的:指令或数据流入后,无论需要执行或处理的时间是长还是短,除了正在运行的部件以外,其余的都处于闲置状态。这样的处理方式非常简单,但效率也惊人的低下,因为这些闲置的部件完全可以在等待的时间里完成其他同类任务。

计算机执行任务的过程类似于工厂的流水线。计算机中的流水线由许多可以完成特定功能的部件组成,这些部件的数量多少称为流水线的级数。流水线中的功能部件按照串行或并行的方式连接。计算机执行任务时,会把任务细分为若干个小任务,交由这些部件执行。部件之间设有缓存寄存器,用于保存上个部件已经执行的任务。在相同的时钟的控制下,任务一级一级地往下流动,最后完成整个任务,如图 3-1 所示。



FI—取指令, D—译码, FO—取操作数, E—执行, W—写

图 3-1 流水线

Intel 在 80486 CPU 中开始使用流水线技术。在 CPU 中由 5 ~ 6 个不同功能的电路单元组成一条指令处理流水线,然后将一条 x86 指令分成 5 ~ 6 步,再由这些电路单元分别执行,这样就能实现在一个 CPU 时钟周期内执行一条指令,因此提高了 CPU 的运算速度。经典的 Pentium 处理器的每条整数流水线都分为 4 级流水,即指令预取、译码、执行、写回结果。

超流水线是通过细化流水线提高主频,使得在一个时钟周期内完成一条甚至多条指令,其实只是以时间换取空间。例如,Pentium 4 的流水线就多达 20 级。流水线的级数越多,完成一条指令的速度就越快,从而能适应主频更高的 CPU。

超标量 (superscalar) 技术是指在 CPU 中有一条以上的流水线,并且每个时钟周期内可以完成一条以上的指令。采用超标量技术的 CPU 集成了多个算术逻辑部件 (ALU)、多个浮点处理单元 (FPU)、多个译码器,以并行处理的方式来提高性能。Pentium 是 Intel 处理器家族中最早采用超标量结构的处理器,带有两条独立的处理管线,每周执行两条指令。

如果是一个超标量系统,假设有 3 个指令控制单元和两个运算单元,那么就可以在完成了第一条指令的取址工作后直接开始第二条指令的取址,这时第一条指令在进行译码,然后开始第三条指令取址,第二条指令开始译码,第一条指令开始取操作数……,这样就形成一个 5 级流水线。显然,5 级流水线的平均理论处理速度是不用流水线的 4 倍。

流水线系统最大限度地利用了 CPU 资源,使每个部件在每个时钟周期都工作,大大提高了效率。但是,流水线有两个非常大的问题:相关和转移。

在一个流水线系统中,如果第二条指令需要用到第一条指令的结果,那么这种情况叫做相关。上面的 5 级流水线为例:当第二条指令需要取操作数时,第一条指令的运算还没有完成,如果这时第二条指令就去取操作数,就会得到错误的结果。所以,这时整条流水线不得不停顿下来,等待第一条指令的完成。这是一个很讨厌的问题,特别是对于比较长的流水线,比如 20 级,这种停顿通常要损失十几个时钟周期。

目前,解决这个问题的方法是乱序执行。乱序执行的原理是在两条相关指令中插入不相关的指令,使整条流水线顺畅。比如上面的例子中,开始执行第一条指令后直接开始执行第三条指令(假设第三条指令不相关),然后才开始执行第二条指令,这样当第二条指令需要取操作数时,第一条指令刚好完成,而且第三条指令也快要完成了,整条流水线不会停顿。当然,流水线的阻塞现象还是不能完全避免的,尤其是当相关指令非常多的时候。

另一个问题是条件转移。在上面的例子中,如果第一条指令是一个条件转移指令,那么系统就不清楚下面应该执行哪一条指令,这时必须等第一条指令的判断结果出来才能执行第二条指令。条件转移所造成的流水线停顿甚至比相关还要严重得多,所以,现在采用分支预测技术来处理转移问题。虽然程序中充满着分支,而且选择任何一条分支都是有可能的,但大多数情况下总是选择某一分支。比如,一个循环的末尾是一个分支,除了最后一次循环结束时需要跳出循环外,其他的时候总是选择继续循环这条分支。

根据这一原理,分支预测技术可以在没有得到结果之前预测下一条指令是什么,并执行它。现在的分支预测技术能够达到 90% 以上的正确率,但是,一旦预测错误,CPU 仍然不得不清理整条流水线并回到分支点。这将损失大量的时钟周期。所以,进一步提高分支预测的准确率也是正在研究的一个课题。

流水线越长,相关和转移这两大问题就越严重。所以,流水线并不是越长越好,超标量也不是越多越好,找到一个速度与效率的平衡点才是最重要的。

3.3.2 高速缓存技术

虽然 CPU 主频的提升会带动系统性能的改善,但系统性能的提高不仅仅取决于 CPU,还与

系统架构、指令结构、信息在各个部件之间的传输速度及存储部件的存取速度等因素有关,特别是与 CPU 和内存之间的存取速度有关。若 CPU 工作速度较高,但内存存取速度较低,则造成 CPU 等待,降低处理速度,浪费 CPU 的能力。例如,500 MHz 的 Pentium III CPU,一条指令的执行时间为 2 ns,与其相配的内存 (SDRAM) 存取时间为 10 ns,比前者慢了许多,因此 CPU 的性能不能充分发挥。减少 CPU 与内存之间的速度差异最有效的方法就是采用高速缓存 (cache),它是当代处理器普遍采用的一种技术。高速缓存是一个速度等于或接近于上级存储的小容量存储器,用于保存先前被存取过的上级存储器的数据以及邻近小范围的数据。

1. 高速缓存的实现原理

高速缓存的实现原理是基于 CPU 运行时的两个基本局限性:时间局限性和空间局限性。

所谓时间局限性,是指 CPU 在某一时刻使用到的数据或指令,在下一时刻也会被重复利用。比如 3D 游戏中,同一场景会在不同时间被渲染多次,如果在第一次渲染中高速缓存存储了相关指令、数据,那么在下次需要重复渲染时,就能够直接从高速缓存中读取相关内容。

而空间局限性,指的是 CPU 在读取某一地址的数据时,也有可能用到该地址附近的数据。也就是说,CPU 需要的数据在很多时候是连续的。例如在听歌或看电影时,数据流总是连续的 (正常播放状态下)。这样的应用对于 CPU 来说是很利,数据预读单元也能够发挥最大的作用。

高速缓存正是利用了上述两个局限性,得以实现并工作。设计师们能够充分利用这两个局限,在容量较小的高速缓存中存入 CPU 在将来某时刻需要的内容。需要指出的是,很多程序在执行指令或数据时,所呈现出来的局限性是不同的。有可能执行指令的时候呈现出时间局限性,而数据呈现出空间局限性,因此设计师们把一级高速缓存分成了指令缓存 (instruction cache) 和数据缓存 (data cache)。

高速缓存正是依据局限性原理,不断地将与当前指令集相关的一个不太大的后继指令集从内存读入高速缓存,然后再向 CPU 高速传送,从而达到速度匹配。CPU 对存储器进行数据请求时,通常先访问高速缓存。由于局限性原理不能保证所请求的数据百分之百地在高速缓存中,这里便存在一个命中率的问题。命中率是指 CPU 在任一时刻从高速缓存中可靠获取数据的几率。命中率越高,正确获取数据的可能性就越大。一般来说,高速缓存的存储容量比主存的容量小得多,但不能太小,太小会使命中率太低,也没有必要过大,过大不仅会增加成本,而且当容量超过一定值后,命中率随容量的增加将不会有明显的增长。只要高速缓存的空间与主存空间在一定范围内保持适当比例的映射关系,高速缓存的命中率还是相当高的。如 256 KB 高速缓存映射 64 MB 内存时,命中率达到 90% 以上。至于没有命中的数据,CPU 只好直接从内存获取,同时把它存入高速缓存,以备下次访问。

2. 高速缓存的基本结构与地址映射方式

高速缓存通常由相联存储器实现,如图 3-2 所示。相联存储器的每一个存储块都具有额外的存储信息,此信息称为标签 (Tag)。当访问相联存储器时,将地址和标签同时进行比较,从而对标签相同的存储块进行访问。

高速缓存在计算机存储系统中没有分配固定的地址,这样程序员在写程序时就不用考虑指令是运行在内存中还是高速缓存中,高速缓存对于计算机上层来说是完全透明的。

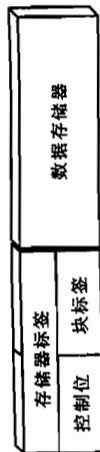


图 3-2 高速缓存的相联存储器结构

CPU 在读取数据时,会首先向内存和高速缓存都发送一个查找指令。如果所需要的数据在高速缓存中(命中),则直接从高速缓存读取数据,以节约时间和资源。CPU 对高速缓存的搜索叫做标签搜索。

在高速缓存中数据或指令是以存储块(block)为单位存储的,一个块又包含了很多字节,现在主流的设计是一块包含 64 B。每一块拥有一个标签。因此,假设 CPU 需要一个标为 Tag 1 的中的数据,它会对高速缓存中的块进行搜索,一旦找到相同标签的块,就对其中的数据进行搜索。

前面已经说过,高速缓存中的数据存储是以块为单位的,而块存储在称为框架(block frame)的数据容器中,框架则直接与内存相对应,如图 3-3 所示。很明显,高速缓存中可能包含数个框架。那么这些高速缓存框架是怎么与内存相对应相联系的呢?有三种方法。

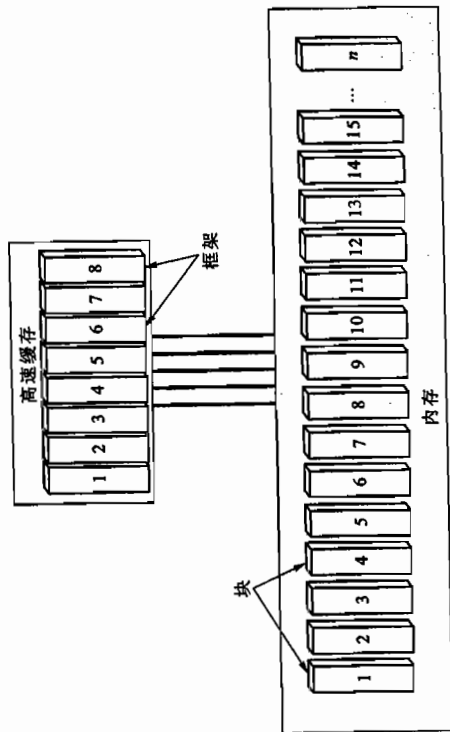


图 3-3 高速缓存的组成结构

第一种方法叫完全相联法,如图 3-4 所示。在这种方法中,内存中的数据可以存储在任何一个高速缓存框架中,同一数据也可以存储在不同的框架中。这样数据的存储相当灵活,CPU 在查找时也很方便;只需在框架中对比找出需要的标签行,即实现命中,显著提高了命中率。然而这样做的缺陷也很明显:对于容量较大的高速缓存来说,CPU 需要在成百的框架中查找需要的标签行,延迟将大大增加。因此这种设计方式只适用于容量较小的高速缓存。

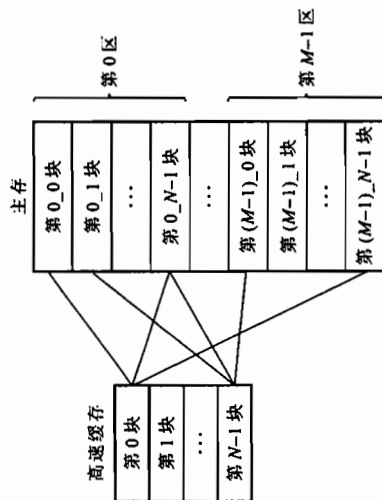


图 3-4 完全相联法

由于完全相联法的这种局限性,设计师们很快提出了另一种旨在降低延迟的组织方式:直接映射法,如图 3-5 所示。和完全相联法不同,在直接映射法中,内存会“记住”存储数据的高速缓存框架地址,以后再次存储时就只能使用该框架。这样做的好处是使 CPU 只需要进行一次标签搜索,在以后的读取操作中就可以直接找到所需标签行所在的框架,从而达到降低延迟的目的。

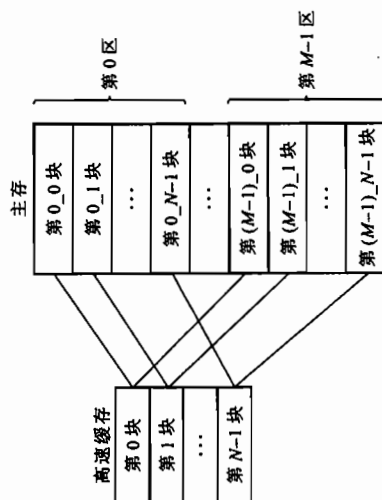


图 3-5 直接映射法

至于内存会将数据存入高速缓存的哪个框架中,可用块地址与整个框架数的同余算法实现。举个简单的例子,有个 1 KB 的缓存,块大小为 64 B,则总共有 16 个缓存块,也就是有 16 个框架。那么在内存中首地址为 12480 的内存块应该保存在缓存的哪个框架中呢? $12480/64 = 195$, $195 \bmod 16 = 3$,则它应该放入第 4 个框架中。这样一来,内存中的数据能很快地读入到缓存中的某个块中,CPU 也能很快地在这个块中找到所要的数据,这样就省下了对比各个框架的时间,

自然延迟就小了。但是,如果第4个框架中装入了内存块195的数据,则其他同余同样是3的35,51,67等这些块就不能装入了。当CPU需要35,51,67这些块的时候,就会发生冲突,导致出现缓存块丢失的情况,大大降低了命中率。

直接映射法和完全相联法都只解决了高速缓存运行问题的一个方面。于是人们又设计了一种综合前两者优点的方法——组相联法,如图3-6所示。先将高速缓存分成不同的组,每个组中放入不同的框架。内存数据的存储对于组来说是固定的(直接映射法),这样就有效地控制了延迟。而每个框架中的块又按照完全相联的方法,灵活存储,这样又提高了命中率。显然,一组中放入的框架越少,其命中率和延迟就能够控制得越好。

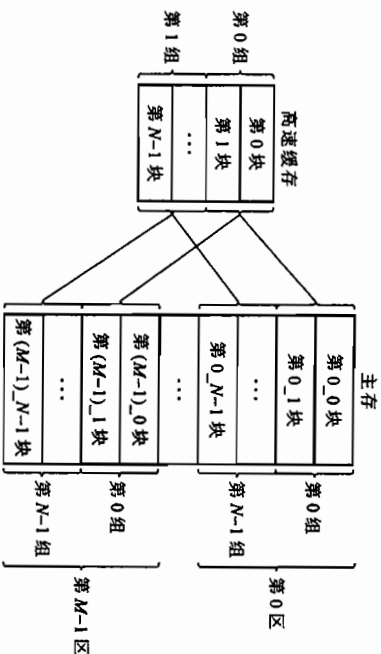


图 3-6 组相联法

组相联法很好地解决了命中率延迟之间的矛盾,因此在现代高速缓存中得到了广泛的应用。

3. 高速缓存与 DRAM 的存取策略

在 CPU 与主存之间增加了高速缓存之后,便产生了数据在 CPU 和高速缓存及主存之间如何存取的问题。读、写各有两种方式。

(1) 贯穿读出式(look through)

该方式将高速缓存隔在 CPU 与主存之间,如图 3-7 所示,CPU 对主存的所有数据请求都首先送到高速缓存,在高速缓存中进行查找。如果命中,则切断 CPU 对主存的请求,并将数据送出;如果没有命中,则将数据请求传给主存。该方法的优点是降低了 CPU 对主存的请求次数,缺点是延迟了 CPU 对主存的访问。

(2) 旁路读出式(look aside)

在这种方式中,CPU 发出数据请求时,并不是单通道地穿过高速缓存,而是向高速缓存和主存同时发出请求,如图 3-8 所示。由于高速缓存速度更快,如果命中,则高速缓存将数据回送给 CPU 的同时,还来得及中断 CPU 对主存的请求;如果没有命中,则高速缓存不做任何动作,由 CPU 直接访问主存。它的优点是没有任何时间延迟,缺点是每次 CPU 对主存访问时都同时访问高速

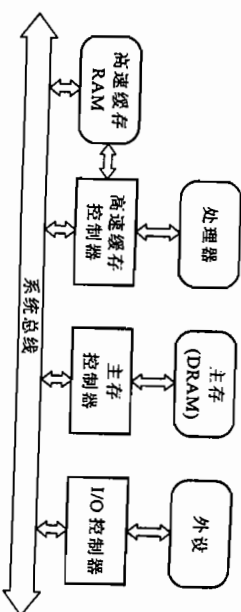


图 3-7 贯穿读出式结构

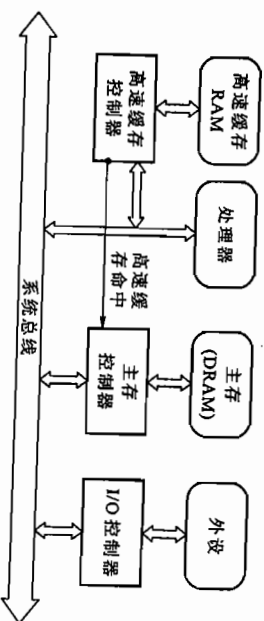


图 3-8 旁路读出式

缓存,这样就占用了一部分总线时间。

(3) 写回法

当 CPU 更新高速缓存时,并不同时更新内存中的相应数据。这种方法减少了访问内存的次数,缩短了访问时间。但在保持与内存内容的一致性方面存在隐患,并且必须为每个缓存块设置一个修改位,以反映此块是否被 CPU 修改过。

(4) 全写法

和写回法相反,当高速缓存数据更新时,也同时更新内存中的数据。高速缓存不用设置修改位或相应的判断器。这种方法的好处是,当高速缓存命中时,由于缓存和内存是同时写入的,所以可以很好地保持缓存和内存内容的一致性。但缺点也很明显。由于每次写入操作都要更新所有的存储体,如果一次有大量的数据要更新,就要占用大量的内存带宽。而内存带宽本来就不宽裕,如果写操作占用太多带宽的话,主要的读操作就会受到比较大的影响。

(5) 写一次法

这是一种基于上面两种方法的写策略,它的特点是,除了第一次更新高速缓存的时候要同时更新内存外,其他时候都和写回法一样,只修改高速缓存。这样就在内存一致性和延迟中找到了一个较好的平衡。

4. 高速缓存的替换策略

所谓替换策略,是指高速缓存中数据的更新方法。高速缓存的容量比较小,要保证其中的数据正是 CPU 所需要的,就需要不停地对高速缓存进行更新,换掉不需要的旧数据,加入新内容。

目前,常见的高速缓存替换策略有三种,下面分别简单介绍。

(1) 先进先出

先进先出(First In First Out, FIFO),即替换最早进入高速缓存的数据。这种算法在早期的高速缓存里使用较多,那时候的高速缓存的容量还很小,数据在高速缓存中停留的时间都不会太久,经常是 CPU 一用完就不得不被替换下来,以保证 CPU 所需要的其他数据能在高速缓存中找到。但这样命中率也会降低,因为这种算法所依据的条件是数据在高速缓存中的时间,而不是其在高速缓存中的使用情况。

(2) 最不经常用

最不经常用(Least Frequency Used, LFU),即替换被 CPU 访问次数最少的行。LFU 算法是将每个行设置计数器,起始为 0,每被 CPU 访问一次,就加 1,当需要替换时,就将计数最小的那一行替换掉,同时将其计数的值置 0。这种算法利用了时间局限性原理,但是每次替换完都把其他行置 0,使得把这种局限性的时间限定在了两次替换之间的时间间隔内。由于替换太频繁,使这一时间间隔太短,并不能完全反映出 CPU 近期的访问情况。

(3) 近期最少使用

近期最少使用(Least Recently Used, LRU),即替换在一段时间里,被 CPU 访问次数最少的行,它是 LFU 的改进版本。其原理是在每行中设置一个计数器,哪一行被 CPU 访问,则该行置 0,其他行增 1,在一段时间内,如此循环,待到要替换时,把计数值最大的行替换掉。这种算法相当于延长了替换时间,从而能够更加真实地反应行的使用情况。有一点要说明的是,有时候行被替换出,并不代表它一定用不到了,而是高速缓存容量不够了。这种算法是目前最优秀的,大部分的高速缓存的替换策略都采用这种算法。

5. 高速缓存的分级体系设计

微处理器的性能可以通过如下公式进行估算:

$$\text{性能} = k \left[\frac{f}{CPI} - (1 - H)N \right]$$

式中: k 为比例常数; f 为工作频率, CPI 为执行每条指令需要的周期数, H 为高速缓存的命中率, N 为存储周期数。

从公式中可以看出,为了提高处理器的性能,应提高工作频率,减少执行每条指令需要的周期数,提高高速缓存的命中率。同时分发多条指令和采用乱序控制,可以减少 CPI 值;采用转移预测和增加高速缓存容量,可以提高 H 值。为了减少存储周期数 N ,可采用高速的总线接口和不分块的高速缓存方案。以前提高处理器的性能,主要靠提高工作频率和提高指令级的并行度,今后则主要靠提高高速缓存的命中率,设计出无阻塞高速缓存分级结构,如图 3-9 所示。

高速缓存分级结构的主要优势在于,对于一个典型的一级缓存系统,80% 的内存访问请求都发生在 CPU 内部,只有 20% 的内存访问请求是与外部内存打交道。而这 20% 的外部内存访问请求中的 80% 又与二级缓存打交道。因此,只有 4% 的内存访问请求定向到 DRAM 中。

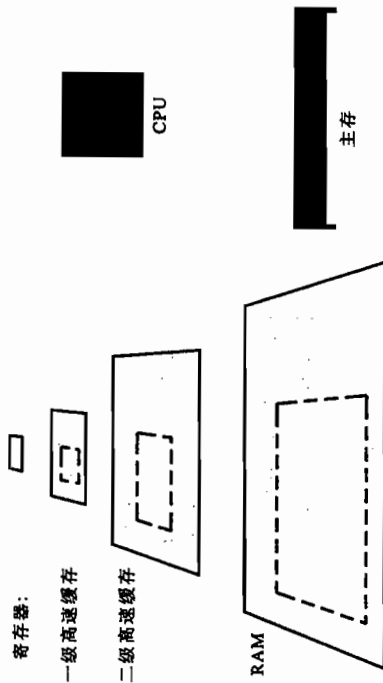


图 3-9 高速缓存的分级体系设计

高速缓存分级结构的不足在于高速缓存组数目受限,需要占用线路板空间和一些支持逻辑电路,会使成本增加。但是综合比较结果还是采用分级高速缓存。

在一级高速缓存结构方面,一般采用回写式静态随机存储器(SRAM)。目前,一级高速缓存容量有加大的趋势。

二级高速缓存的设计分芯片内置和外置两种。

由于一级高速缓存容量的限制,为了进一步提高内存存取速度,在 Intel 的 486 和 Pentium 时期,在 CPU 外部(通常在主板上)再设置一高速缓冲存储器,即二级高速缓存。CPU 在读取数据时,先在一级高速缓存中查找,再从二级高速缓存寻找,然后是内存。从 Intel Pentium II 及 AMD K7 开始,二级高速缓存都内置到 CPU 里。现在 CPU 都采用内置二级高速缓存。

3.3.3 NetBurst 架构 Pentium 4 中的高速缓存实现技术

NetBurst 架构是 Intel 推出 Pentium 4 时提出的新的 CPU 内核结构,其高速缓存设计十分先进,在一定程度上代表了未来的发展趋势。

NetBurst 架构 Pentium 4 的一级高速缓存和其他 CPU 有些不一样。它没有了以前的指令缓存(instruction cache),取而代之的是回溯缓存(trace cache);并且其容量度量单位也不再是 KB,而是 μ -Ops。这是怎么回事呢?

先来看看 CPU 运行程序时的基本操作流程:首先 CPU 会根据程序计数器(PC)提供的地址取得指令,然后对指令进行解码,得出操作地址和操作码。再根据操作地址获得操作数,并由 ALU 结合操作码计算出操作数,最后保存在寄存器中。其中最复杂的就是解码操作。现代 CPU 为了提高指令执行效率,借鉴了 RISC 的设计理念,将不同的 x86 指令解码为长短统一、内容简单的微指令,并以超标量的方式执行。

以前的 CPU 设计中,一般等到 CPU 需要执行某指令时,才对其进行解码。引入了分支预测

后,就可以预先读取CPU可能用到的指令,并解码,以备CPU需要时直接调用,这就是回溯缓存的作用。它可以分步预测得到的x86指令事先解码为微指令,并存储在一个被称为回溯片断(trace segment)的逻辑组中。这些解码后的指令非常灵活,它能够在进入回溯片断前决定是否直接进入流水线来执行。当回溯片断中的微操作指令执行后,回溯缓存就会马上变成执行模式(execute mode),在这种模式下,微指令就会依据预测时的顺序由CPU提取,并执行。此时,回溯缓存存储的指令就是CPU可以直接执行的微操作,并且由于是存放在一级高速缓存中,使得CPU在需要时可以马上取得已经译码的微指令,再加上超低的延迟,可以使CPU的执行频率迅速提高,从而使CPU的主频达到前所未有的高度。还有一个重要的好处就是可以减少CPU本身的x86解码单元,从而简化CPU结构。总的来说,引入回溯缓存的最终目的就是减少x86指令解码器,缓解长流水线出现预测失误所带来的性能损失,并代替原始的一级高速缓存中的指令缓存。

由于Intel一直没有公布回溯缓存的具体大小,只是透露其可以装入12条微指令,因此其单位是12μ-Ops。

这种大胆的高速缓存设计最大的弊端在于,一旦分支预测出现错误,后果将极其严重。因此Intel希望借助于较大的二级高速缓存来降低预测出错引起的性能损失,这就要求二级高速缓存拥有极高的命中率。因此Intel在NetBurst架构中采用了8路组相联(其命中率可以达到完全相联的水平)、容量为2MB的二级高速缓存,并提供256B的带宽,以确保其有低的延迟。但实际效果并不理想。由于流水线级别的增加给分支预测带来了不可估计的难度,即使有2MB的二级高速缓存,也难以避免Pentium 4的“高频低能”现象。

对于AMD的K7/K8来说,其高速缓存设计同Intel也有相似之处。AMD历来重视一级高速缓存,均为64KB的指令缓存与数据缓存,加上2路组相联的设计,使AMD的CPU拥有高命中率和低延迟的一级高速缓存。然而在指令缓存中,AMD的设计思路又与其他CPU有所不同。首先,其指令缓存中并不存储传统的x86指令,而是存储着分支预测出来的x86指令的信息。再根据这些信息将指令解码成1~2条“宏操作”来执行。由于AMD CPU采用了3条流水线的超标量结构,因此其可以保证同时执行6条这样的宏操作。显然,相比之下AMD CPU的指令执行效率要优于Intel,因此就可以解释为什么AMD 2.4GHz优于Intel 3.2GHz。然而,从这样的一级高速缓存设计来看,似乎和Intel有异曲同工之妙。

对于二级高速缓存,AMD CPU历来只有256KB或512KB。这是由于AMD的高速缓存设计思想所致。在一般的高速缓存中,下级高速缓存总是要保留一个上级高速缓存的映像,即一级高速缓存中的数据在二级高速缓存中也能找到,二级高速缓存中的数据在三级高速缓存中也能找到。然而,AMD并没有这样做,它的二级高速缓存中保存的数据都是一级高速缓存中替换下来的,以保证CPU在下次使用时能够在二级高速缓存中找到。因此,AMD的高速缓存结构呈现出“一级高速缓存+二级高速缓存的‘怪异’现象”,它的二级高速缓存对于整体性能并不起决定性作用。Sempion 2600+(128KB二级高速缓存)同2800+(256KB二级高速缓存)性能相同,就是这个原因。

3.3.4 扩展指令集

从Pentium MMX开始,为了提升处理器各方面的性能,Intel和AMD在x86指令集的基础

上,各自开发新的指令集。指令集中包含了对多媒体、3D处理等方面的支持,这些指令集能够提高处理器在多媒体和3D的方面处理能力,但是需要必要的软件支持。

1. MMX 技术

MMX技术是单指令多数据流(SIMD)技术在Pentium和Pentium II的具体实现,它可使单条指令处理多条信息。将这种并行操作技术与超标量体系结构相结合,可以极大地提高Pentium CPU的性能。MMX技术的另一基础是向下兼容,即MMX技术与已有的操作系统和软件完全兼容,它不引进新的状态寄存器、控制寄存器以及新的条件码等。

MMX技术核心体现在如下3点:

- 4种新的数据类型
- 8个64位宽的MMX寄存器
- 57条新指令

(1) MMX 指令集

共有57条MMX指令添加到Pentium原有的指令系统中,这些MMX指令可分成如下7组:

- 算术指令
- 比较指令
- 转换指令
- 逻辑指令
- 移位指令
- 数据传送指令
- 清除MMX状态(EMMS)指令

指令的助记符表示法是在一般的指令操作符(如ADD、CMP、XOR等)之前加字P,表示打包数据,之后可加表示算法和数据类型的字母,如PADDD和PADDU等。

- US——无符号数饱和算法(unsigned saturation)
- S——有符号数饱和算法(signed saturation)
- B——紧缩字节
- W——紧缩字
- D——紧缩双字
- Q——四字

MMX指令的先进性体现在以下3个方面。

(2) 单指令多数据流(SIMD)结构

这在前面的章节中已经介绍过,它是MMX技术提高计算性能的基础。以前的微处理器在处理8位或16位数据时,仅能用寄存器或寄存器的低8位或16位。对于具有32位或64位带宽的微处理器来说,处理器资源处于未充分利用状态。MMX指令则充分利用了微处理器64位带宽的处理能力,一次并行处理多个不相关的小型数据元,例如,对于单指令多数据加法,一条指令可同时进行8组数的运算。

(3) 扩充的饱和运算模式(saturation mode)

定点运算经常会发生运算结果的溢出,处理器遇到这种情况时,通常是作为异常来处理,设

$$\begin{array}{r}
 X=(1.2, 3.5, 8.9, 0.5) \\
 Y=(1.2, 3.5, 8.9, 5.7) \\
 \hline
 + \quad + \quad + \quad + \\
 \hline
 Z=(2.4, 6.10, 16.18, 5.12)
 \end{array}$$



置溢出标志并产生中断。而在图形和图像的数据处理中,常用8位无符号数表示一个像素点的颜色值或者像素点的R、G、B颜色分量值,数据溢出会使像素点的颜色值从最大值变为最小值,或从最小值变为最大值。例如,8位的像素点数据设定0表示黑色,255表示为白色。当运算结果为256时,溢出的运算结果为0。结果在一片白色区域中突然冒出一个黑点。因此,溢出必须得到处理,但异常处理却常常要打断程序的正常运行,使处理速度变慢。采用扩充的饱和和运算方式,可省却溢出处理,提高处理速度。

如MMX的加、减法指令中,PADD, PADDD, PSUBS, PSUBD具有饱和和运算功能,即当运算结果超出最大值时按最大值处理,当运算结果低于最小值时按最小值处理,并且不设置溢出标志。这样既符合图形、图像处理的实际情况,又减少了溢出判断处理,加快了运算速度。

(4) 积和运算方式

MMX的乘法指令中,PMADDWD指令是一条关键指令,它具有乘法累加操作功能。图3-10说明了它的操作功能,将两个紧缩字类型的数中相应元素(16位)相乘,生成4个32位的积,再将左侧的两个积相加,得出一个结果,右侧的两个积相加,得出另一个结果,这样生成一个紧缩双字类型的数据。

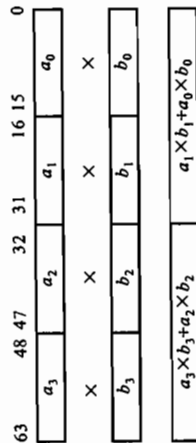


图3-10 PMADDWD指令功能

积和运算方式特别适合于向量计算与矩阵计算。在傅里叶变换、离散余弦变换等频率变换中和有限脉冲响应滤波、无限脉冲响应滤波中,使用PMADDWD指令会大大减少编程的指令数量。

MMX的其他指令也颇具特色,这里就不一一详述了。总之,MMX指令集是一种SIMD并行处理指令。MMX技术是建立在指令级并行的RISC超标量结构之上的,是为了扩充CPU的多媒体和通信功能而设计的。

2. SSE指令集

SSE是Streaming SIMD Extension(流式SIMD扩展指令集)的缩写,而其中SIMD含意为Single Instruction Multiple Data stream(单指令多数据流),所以SSE指令集也叫单指令多数据流扩展。该指令集最先运用于Intel的Pentium III系列处理器。SSE指令集是为提高处理器浮点性能

而开发的扩展指令集,它共有70条指令,其中包含提高3D图形运算效率的50条SIMD浮点运算指令、12条MMX整数运算增强指令和8条优化内存中的连续数据块传输指令。理论上这些指令对当时流行的图像处理、浮点运算、3D运算、多媒体处理等众多应用起到全面提升的作用。SSE指令与AMD公司的3DNow!指令彼此互不兼容,但SSE包含了3DNow!中的绝大部分功能,只是实现的方法不同而已。SSE也向下兼容MMX指令,它可以通过SIMD和单时钟周期并行处理多个浮点数据,从而有效地提高浮点运算速度。

3. 3DNow!指令集

3DNow!指令集由AMD公司开发,该指令集在SSE指令集之前推出,被广泛运用于AMD的K6、K6-2和K7系列处理器上,具有21条扩展指令。在整体上3DNow!与SSE非常相似,它们都具有8个新的寄存器,但是3DNow!的数据带宽是64位的,而SSE是128位。所以3DNow!只能存储两个浮点数,而SSE是4个。但是SSE的侧重点有所不同,3DNow!指令集主要针对3D建模、坐标变换和效果渲染等3D处理,在相应的软件配合下,可以大幅度提高处理器的3D处理性能。AMD公司后来又又在Athlon系列处理器上开发了增强的Enhanced 3DNow!指令集,新的增强指令达52条。目前最为流行的Athlon 64系列处理器仍支持3DNow!指令。

4. SSE2和SSE3指令集

SSE2和SSE3指令集主要用于Pentium 4系列处理器。

SSE2指令集包含了144条指令,由两个部分组成:SSE部分和MMX部分。SSE部分主要负责处理浮点数,而MMX部分则专门计算整数。SSE2的寄存器容量是MMX寄存器的两倍。在指令处理速度保持不变的情况下,通过SSE2优化后的程序和软件运行速度也能够提高一倍。由于SSE2指令集与MMX指令集兼容,因此被MMX优化过的程序很容易被SSE2再进行更高层次的优化,达到更好的运行效果。SSE2对于处理器的性能的提升是十分明显的。虽然在同频率的情况下,Pentium 4性能不如Athlon XP,但由于Athlon XP不支持SSE2,所以经过SSE2优化后的程序在Pentium 4中的运行速度要明显高于Athlon XP。而AMD方面也注意到了这一情况,在随后的K8系列处理器中,都加入SSE2指令集。

SSE3指令集是Intel公司在SSE2指令集的基础上发展起来的。相比于SSE2,SSE3在SSE2的基础上又增加了13个额外的SIMD指令。SSE3中13个新指令的主要目的是改进线程同步和特定应用程序领域,例如多媒体和游戏。这些新增指令强化了处理器在浮点转换至整数、复杂算法、视频编码、SIMD浮点寄存器操作以及线程同步等5个方面的表现,最终达到提升多媒体和游戏性能的目的。Intel是从Prescott核心的Pentium 4开始支持SSE3指令集的。

3.3.5 64位技术

64位技术是相对于32位而言的,这个位数指的是CPU的通用寄存器(General-Purpose Registers, GPR)的数据宽度为64位。64位指令集就是运行64位数据的指令,也就是说处理器一次可以处理64位数据。64位处理器并非现在才有的,在高端的精简指令集计算机(RISC) (Reduced Instruction Set Computer, RISC)中很早就有64位处理器了,比如SUN公司的UltraSparc III、IBM公司的POWER5以及HP公司的Alpha等。

64位计算主要有两大优点:可以进行更大范围的整数运算,可以支持更大的内存。但不能

简单地认为 64 位处理器的性能是 32 位处理器性能的两倍。实际上在 32 位应用下, 32 位处理器的性能甚至会更强。

要实现真正意义上的 64 位计算, 光有 64 位的处理器是不行的, 还必须得有 64 位的操作系统以及 64 位的应用软件才行, 三者缺一不可, 缺少其中任何一个要素都是无法实现 64 位计算的。

目前, 主流 CPU 使用的 64 位技术主要有 AMD 公司的 AMD64 技术、Intel 公司的 EM64T 技术和 IA-64 技术。其中 IA-64 是 Intel 独立开发, 不兼容现在的传统的 32 位计算机, 仅用于 Itanium (安腾) 及其后续产品 Itanium 2。

1. AMD 的 64 位技术

AMD 的 Athlon 64 系列处理器的 64 位技术是在 x86 指令集的基础上加入了 x86-64 的 64 位扩展 x86 指令集, 这就使得 Athlon 64 系列处理器可兼容原来的 32 位的 x86 指令, 同时支持 x86-64 的扩展 64 位计算, 并且具有 64 位的寻址能力, 使得它成为真正的 64 位 x86 构架处理器。在采用 x86-64 架构的 Athlon 64 处理器中, 新增了几组寄存器, 能够提供更高的执行效率。寄存器是处理器用来创建和存储 CPU 运算结果的地方, 标准的 x86 构架中包括 8 组通用寄存器, 而在 AMD 的 x86-64 架构中又增加了 8 组, 将通用寄存器的数目提高到了 16 组。在此基础上, x86-64 指令集另外增加了 8 组 128 位的 XMM 寄存器, 也叫做 SSE 寄存器。它能够给单指令多数数据流技术 (SIMD) 运算提供更多的存储空间, 这些 128 位的寄存器能够提供在矢量和标量计算模式下进行 128 位双精度处理, 这也为 3D 数据处理、矢量分析和虚拟技术提供了良好的硬件基础。由于提供了更多的寄存器, 按照 x86-64 标准生产的处理器可以更高效地处理数据, 在一个时钟周期内能够传输更多的信息。

2. Intel 的 64 位技术

EM64T 的全称是 Extended Memory 64 Technology, 即扩展 64 位内存技术。EM64T 是 Intel IA-32 架构的扩展, 即 IA-32e (Intel Architecture-32 extension)。IA-32 处理器通过附加 EM64T 技术, 便可在兼容 IA-32 软件的情况下, 允许软件利用更多的内存地址空间, 并且允许软件进行 32 位线性地址写入。EM64T 特别强调的是与 32 位和 64 位内存的兼容性。Intel 为新核心增加了 8 个 64 位通用寄存器 (R8~R15), 并且把原有通用寄存器全部扩展为 64 位, 可以提高整数运算能力。增加 8 个 128 位 SSE 寄存器 (XMM8~XMM15) 是为了增强多媒体性能, 包括对 SSE、SSE2 和 SSE3 的支持。

Intel 为支持 EM64T 技术的处理器设计了两大模式: 传统 IA-32 模式 (legacy IA-32 mode) 和 IA-32e 模式 (IA-32e mode)。在支持 EM64T 技术的处理器内有一个称为扩展功能激活寄存器 (extended feature enable register, IA32_EFER) 的部件, 其中的第 10 位控制着 EM64T 是否激活, 决定是 IA-32e 模式有效 (IA-32e mode active) 还是长模式有效 (long mode active, LMA)。当 LMA=0 时, 处理器便作为一颗标准的 32 位 (IA32) 处理器运行在传统 IA-32 模式下; 当 LMA=1 时, EM64T 便被激活, 处理器运行在 IA-32e 模式下。

3.3.6 超线程技术

当运行程序或对程序进行输入时, 程序便把线程发送到 CPU, 告诉 CPU 怎样响应输入。当

双击鼠标打开 Word 文档, CPU 就会发送一组指令来执行相关操作。对于 CPU 来说, 一个线程就是必须执行的指令集合。

CPU 能够连续地处理每条指令, 直到处理完一个线程为止。在线程执行以后, CPU 就会重新读入下一个要执行的指令位置。不同线程可以彼此相互中断, 即强迫 CPU 把当前处理的计算结果存储在堆栈里。这样做的缺陷是 CPU 每次只能处理一个线程。

从原理上来说, 超线程技术属于 Intel 版本的多线程技术。这种技术可以让单 CPU 拥有处理多线程的能力, 而物理上只使用一个处理器。超线程技术为每个物理处理器设置了两个入口——架构状态 (Architecture State, AS) 接口, 从而使操作系统等软件将其识别为两个逻辑处理器。这两个逻辑处理器像传统处理器一样, 都有独立的 IA-32 架构, 它们可以分别进入暂停、中断状态, 或直接执行特殊线程, 并且每个逻辑处理器都拥有高级可编程中断控制器 (Advanced Programmable Interrupt Controller, APIC)。虽然支持超线程的 Pentium 4 能同时执行两个线程, 但不同于双处理器平台或双内核处理器, 超线程中的两个逻辑处理器并没有独立的执行单元、整数单元、寄存器、高速缓存等资源。它们在运行过程中仍需要共用执行单元、高速缓存和系统总线接口。在执行多线程时两个逻辑处理器均是交替工作, 当两个线程同时需要某一个资源时, 其中一个逻辑处理器要暂停并要出让资源, 等待那些资源闲置时才能继续工作。因此, 超线程技术所带来的性能提升远不能等同于两个相同时钟频率的处理器带来的性能提升。可以说 Intel 的超线程技术仅仅是对单个处理器运算资源的优化利用。超线程技术的原理如图 3-11 所示。

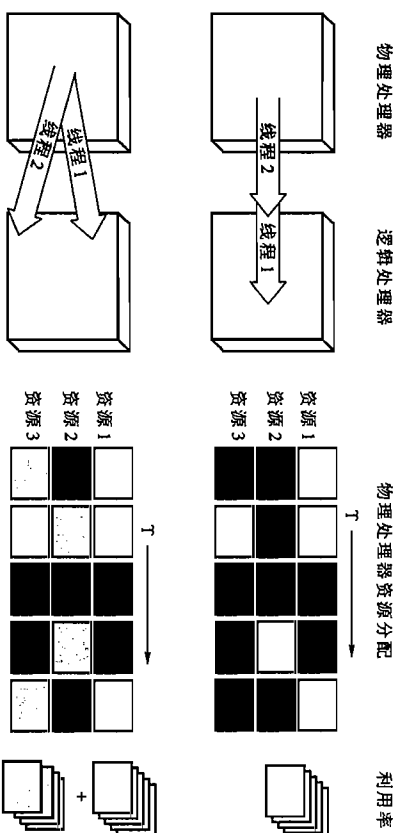


图 3-11 超线程技术

3.3.7 多核心技术

无论是台式计算机还是笔记本电脑, 影响其性能的最关键因素还是 CPU 的性能。基于摩尔定律, 在短短十几年之间, CPU 的工作频率及性能都有了大幅度的提升, 为我们建立高效的娱乐、办公环境奠定了基础。

但是, 在最近几年, CPU 工作频率的提升遇到了瓶颈。无论是 Intel 还是 AMD, 昔日的架构

已经难以继续走下去。多核心的架构缓解了单核 CPU 通过提升频率提高其性能所遭遇的瓶颈,为提升 CPU 的性能开辟了另一条出路。可以说,单纯追求处理器频率的时代已经过去,未来处理器的发展将在“频率、多核心、微架构”中做出平衡。

可以把单核心及双核心的效能功耗表现做一个对比。如果双核心电压及频率均降低至 85%,总功耗约等于 $2 \times 0.85 \times 0.85 \times 0.85$,与单核心的功耗基本相同,但其每个时钟周期执行指令数却是单核心的 1.7 倍 (2×0.85)。由此可知多核心设计比单纯以单核心提升频率,拥有更高的效能功耗比(performance per watt)。虽然多核心处理器成本较高,但随工艺的进步使成本降低,多核心将会和微架构改良一样,成为未来处理器效能提升的核心技术。

在未来,多核心不再是单纯地将相同的核心加进同一芯片上,而是分开通用核心及专门处理核心,称为异构多核心体系结构(Heterogeneous Multi-Core Architecture)。一些复杂的处理如压缩、语音辨识、物理运算等将交由专门处理核心处理,而且各核心内建立互连结构(Interconnect fabric)提升运算效率。目前,Intel 正进一步研发多核心处理器,并与软件业界紧密合作,希望配合软件的支持。

多核心 CPU 的出现将大幅度提升处理器的运算能力,也是未来 CPU 发展的趋势。多核心可以赋予轻薄笔记本计算机更强大的处理能力,让笔记本计算机真正可以在体积、重量与性能之间取得平衡。

与单核处理器相比,多核处理器在运算能力和效率方面具有较大的优势,最终必将成为一种广泛普及的计算模式。Intel 和 AMD 都在致力于多核心 CPU 的研发,在 2005 年下半年 Intel 与 AMD 推出了双核心的处理器。

所谓双核心处理器,简单地说就是在同一块 CPU 基板上集成两个处理器核心,并通过并行总线将两个处理器核心连接起来,如图 3-12 所示。双核心并不是一个新的概念,而只是单芯片多处理器(Chip Multi Processors, CMP)技术最基础、最简单、最容易的一种实现。其实,在 RISC 处理器领域,双核心甚至多核心都早已实现。CMP 最早是由美国斯坦福大学提出的,其思想是在一块芯片内实现对称多处理(Symmetrical Multi-Processing, SMP)架构,且并行执行不同的进程。早在上个世纪末,惠普和 IBM 就已经提出了双核处理器的可行性设计。IBM 在 2001 年就推出了基于双核心的 POWER4 处理器,随后 Sun 和惠普公司先后推出了基于双核心架构的 UltraSPARC 以及 PA-RISC 芯片,但这些都双核心处理器架构都应用于高端的 RISC 领域。直到前不久 Intel 和 AMD 相继推出自己的双核心处理器,双核心处理器才真正走入了主流的 x86 领域。

双核心处理器技术的引入是提高处理器性能的有效方法。因为处理器实际性能是处理器在每一个时钟周期内所能执行指令的总数,因此增加一个核心后,处理器每个时钟周期内可执行的单元数将增加一倍。

双核心处理器的一个主要特点是:它运行起来像是一个双处理器架构,但实际上只是一个单处理器架构。因此,软件必须进行专门的设计才能够充分利用多个核心的优势。目前,软件中针

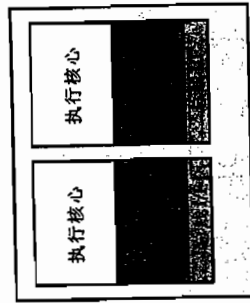


图 3-12 双核心处理器架构



图 3-13 CPU 的封装

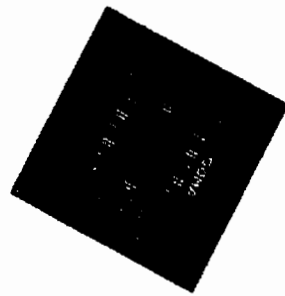


图 3-14 CPU 有引脚一面

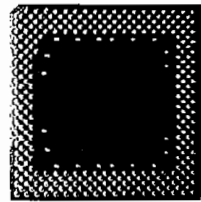
对超线程技术而设计的应用,在双核心处理器中也能够充分体现出来。

3.4 CPU 的封装与接口类型

3.4.1 CPU 的封装

从外表上看,CPU 通常是矩形或正方形的块状物,通过密密麻麻的众多管脚与主板相连。这就是 CPU 的外衣——CPU 的封装。而内部,CPU 的核心是一片大小通常不到 1/4 英寸的薄薄的硅晶片(die),如图 3-13 所示。这块小小的硅片上,密布着数以百万计的晶体管,它们好像大脑的神经元,相互配合协调,完成各种复杂的运算和操作。在通过严格的测试以后,已经刻蚀出各种电路结构的硅片就可以送到封装厂进行切割,划分成单个处理器的芯片并置入到封装中。封装可不仅仅是件漂亮的外衣。由于有封装的保护,处理器核心与空气隔离可以避免污染物的侵害。

除此以外,良好的封装设计还有助于芯片散热。同时,它是连接处理器和主板的桥梁。封装技术也在不断发展,经历了 8086/8088 时代的双列直插(Dual In-line Package, DIP)封装,386/486 时代的塑料方形扁平封装(Plastic Quad Flat Package, PQFP)。目前最常见的是针栅阵列(Pin-Grid Array, PGA)封装。图 3-14 所示是采用 PGA 封装的 Pentium CPU 有针脚的一面。这种封装通常是正方形的,在中央区周围均匀地分布着 3~4 排甚至更多排的引脚,引脚能插入主板 CPU 插座上对应的插孔。随着 CPU 总线宽度的增加,功能的增强,CPU 的引脚数目也不断增多,同时对散热、电气特性有了更高的要求,从而演化出了交错针栅阵列(Staggered Pin-Grid Array, SPGA)、塑料针栅阵列(Plastic Pin-Grid Array, PPGA)及反转芯片针栅阵列(Flip Chip Pin-Grid Array, FC-PGA)封装。



FC-PGA:在这种封装中,芯片被反转,使片模或处理器的芯片部分暴露在处理器的上部。通过将片模暴露出来,使散热解决方案可直接用到片模上,从而更有效地冷却芯片。为了通过隔绝电源信号和接地信号来提高封装的性能。在处理器底部的电容放置区域(处理器中心)安有离散电容和电阻。芯片底部的针脚是锯齿形排列的。此外,针脚的排列方式使得处理器只能以一

种方式插入插座。FC-PGA封装用于 Pentium III 和 Celeron 处理器,它们都使用 370 个引脚。

FC-PGA2:FC-PGA2 封装与 FC-PGA 封装类似,但 FC-PGA2 封装中安装有集成式散热器(HS)。集成式散热器是在生产时直接安装到处理器芯片上的。由于集成式散热器与片模有很好的热接触,并且提供了更大的表面积,以更好地散发热量,所以它显著地改善了散热效果。FC-PGA2 封装用于 370 个引脚的 Pentium III 和 Celeron 处理器以及 478 引脚的 Pentium 4 处理器。

3.4.2 CPU 的接口

CPU 需要通过某个接口与主板连接才能进行工作。经过这么多年的发展,CPU 的接口有引脚式、卡式、触点式等。目前,CPU 都采用引脚式接口,而主板上也有相应类型的插座。CPU 接口类型不同,其插孔数、体积、形状都有所不同,所以不能互相接插。

不同接口 CPU 的引脚数各不相同。CPU 接口类型的命名,习惯用引脚数来表示,比如,目前 Pentium 4 系列处理器所采用的 Socket 478 接口,其引脚数就为 478;而 Athlon XP 系列处理器所采用的 Socket 939 接口,其引脚数为 939。

理论上,CPU 性能的好坏和引脚数的多少是没有关系的,而且 CPU 上的引脚也并不是每一个都起作用,也就是说 CPU 上有些引脚其实是没有任何作用的。这是因为 CPU 厂商在设计 CPU 时,考虑到今后一段时间内的功能扩展和性能提高,通常会预留一些暂时不起作用的引脚。不过,随着 CPU 技术的发展,需要越来越多的引脚,以实现更丰富的功能以及更高的性能。例如,集成双通道内存控制器所需要的引脚数量,要比只集成单通道内存控制器所需要的引脚数多得多。因此,总的来说,CPU 引脚数有越来越多的趋势,基本上可以认为引脚越多的 CPU 其架构也越先进。

1. Socket 423

Socket 423 接口是最初的 Pentium 4 处理器的标准接口。Socket 423 的外形和前几种 Socket 接口类似,如图 3-15 所示,对应的 CPU 引脚数为 423。随着 DDR 内存的流行,Intel 开发了支持 SDRAM 及 DDR 内存的 1845 芯片组,CPU 接口也改成了 Socket 478,Socket 423 接口也就停止使用了。

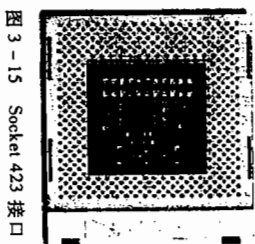


图 3-15 Socket 423 接口

2. Socket 478

最初的 Socket 478 接口是 Intel 用来取代 Socket 423 的用于 Pentium 4 系列处理器的接口类型,引脚为 478 个。Socket 478 的 Pentium 4 处理器面积很小,其引脚排列极为紧密。Intel 公司的采用 Northwood 核心的 Pentium 4 和基于 Pentium 4 的 Celeron 系列 CPU 都采用此接口。目前,这种 CPU 已经逐步退出市场。

但是,Intel 于 2006 年初推出了一种全新的 Socket 478 接口,如图 3-16 所示,这种接口是目前 Intel 公司采用 Core 架构的处理器 Core Duo 和 Core Solo 的专用接口,与早期的台式机用 Pentium 4 系列的 Socket 478 接口相比,虽然引脚同为 478 个,但是其引脚定义以及电压等重要参数完全不同,所以二者之间并不兼容。随着 Intel 公司的处理器全面向 Core 架构转移,今后采用新 Socket 478 接口的处理器将会越来越多,例如,即将推出的 Core 架构的 Celeron M 也会采用这种接口。

3. Socket 775

Socket 775(见图 3-17)又称为 Socket T,是应用于 LGA775 封装的 CPU 所对应的接口。目前,采用此种接口的有 LGA775 封装的单核心的 Pentium 4、Pentium 4 EE、Celeron D 以及双核心的 Pentium D 和 Pentium EE 等 CPU。与采用 Socket 478 接口的 CPU 不同,Socket 775 接口的 CPU 的底部没有传统的引脚,而代之以 775 个触点,通过与之对应的 Socket 775 插槽内的 775 根触针接触来传输信号。Socket 775 接口不仅能够有效提升处理器的信号强度和频率,同时也可以提高处理器生产的良品率,降低生产成本。随着 Socket 478 的逐渐淡出,Socket 775 已经成为 Intel 台式机 CPU 的标准接口。

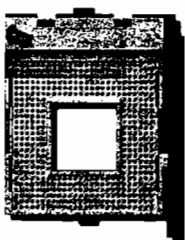


图 3-16 Socket 478 接口

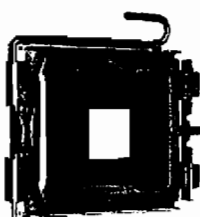


图 3-17 Socket 775 接口

4. Socket A

Socket A 接口,也叫 Socket 462,是 AMD 公司 Athlon XP 和 Duron 处理器采用的接口,如图 3-18 所示。Socket A 接口具有 462 个插孔,可以支持 133 MHz 的外频。

5. Socket 754

Socket 754 是 2003 年 9 月 AMD 64 位台式机 CPU 最初发布时的接口,具有 754 根引脚,只支持单通道 DDR 内存,如图 3-19 所示。目前,采用此接口的 CPU 有用于台式计算机的 Athlon 64 低端型号和 Sempron 高端型号以及笔记本电脑的 Mobile Sempron、Mobile Athlon 64 以及 Turion 64。随着 AMD 从 2006 年开始全面转向支持 DDR2 内存,Socket 754 接口将逐渐被 Socket AM2 接口所取代,从而使 AMD 的台式机处理器接口走向统一。与此同时,用于笔记本电脑的 Socket 754 接口也将逐渐被具有 638 根引脚、支持双通道 DDR2 内存的 Socket S1 接口所取代。Socket 754 接口将在 2007 年底完成自己的历史使命,但其寿命反而要比一度号称要取而代之的 Socket 939 接口长得多。

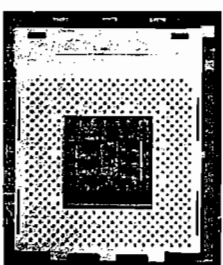


图 3-18 Socket A 接口

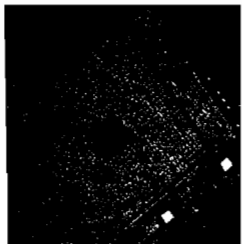


图 3-19 Socket 754 接口

6. Socket 939

Socket 939 是 AMD 公司 2004 年 6 月才推出的 64 位 CPU 接口标准,具有 939 个引脚,支持双通道 DDR 内存。目前,采用此接口的有面向入门级服务器/工作站市场的 Opteron 1xx 系列以及面向 PC 市场的 Athlon 64、Athlon 64 FX 和 Athlon 64 X2。除此之外部分专供 OEM 厂商的 Sempron 也采用了 Socket 939 接口。Socket 939 接口的处理器和与过去的 Socket 940 插槽是不能混插的,但是 Socket 939 仍然使用了相同的 CPU 风扇系统模式。随着 AMD 从 2006 年开始全面转向支持 DDR2 内存,Socket 939 接口为 Socket AM2 接口所取代,在 2007 年初完成自己的历史使命,从推出到被淘汰其寿命还不到 3 年。

7. Socket AM2

Socket AM2 是 AMD 于 2006 年 5 月底发布的 64 位台式计算机 CPU 的接口标准,具有 940 个引脚,支持双通道 DDR2 内存,如图 3-20 所示。虽然同样具有 940 个引脚,但 Socket AM2 与 Socket 940 在引脚定义以及排列方面都不相同,并不能互相兼容。目前采用 Socket AM2 接口的有低端的 Sempron、中端的 Athlon 64、高端的 Athlon 64 X2 以及顶级的 Athlon 64 FX 等全系列 AMD 台式计算机 CPU,支持 200 MHz 外频和 1 000 MHz 的 HyperTransport 总线频率,支持双通道 DDR2 内存,其中 Athlon 64 X2 以及 Athlon 64 FX 最高支持 DDR2-800, Sempron 和 Athlon 64 最高支持 DDR2-667。按照 AMD 的规划,Socket AM2 接口将逐渐取代 Socket 754 接口和 Socket 939 接口,从而实现台式计算机 CPU 接口的统一。

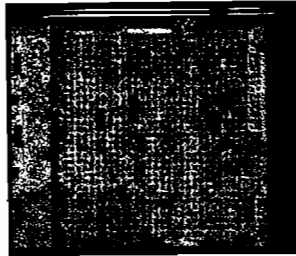


图 3-20 Socket AM2

3.5 CPU 的内核

核心又称为内核,是 CPU 最重要的组成部分。CPU 中心那块隆起的芯片就是核心,是用单晶硅以一定的生产工艺制造出来的,CPU 所有的计算、接收/存储命令、处理数据都由核心执行。各种 CPU 核心都具有固定的逻辑结构,一级缓存、二级缓存、执行单元、指令级单元和总线接口等逻辑单元都有科学的布局。

为了便于设计、生产、销售及及管理,CPU 制造商通常会对各种 CPU 核心编制相应的代号,这也就是所谓的 CPU 核心类型。

不同的 CPU(不同系列或同一系列的不同型号)都会有不同的核心类型(例如,Pentium 4 的 Northwood、Willamette 以及 K6-2 的 CXT 和 K6-2 的 ST-50 等),甚至同一种核心都会有不同版本(例如,Northwood 核心就分为 B0 和 C1 等版本)。核心版本的变更是为了修正上一版本存在的一些缺陷,并提升一定的性能,而这些变化普通消费者是很少去注意的。每一种核心类型都有其相应的制造工艺(例如 0.25 μm 、0.18 μm 、0.13 μm 以及 0.09 μm 等)、核心面积(这是决定 CPU 成本的关键因素,成本与核心面积基本上成正比)、核心电压与电流的大小、晶体管数量、各级缓存的大小、主频范围、流水线架构和支持的指令集(这是决定 CPU 实际性能和工作效率的

关键因素)、功耗和发热量的大小、封装方式(例如,S.E.P.PGA、FC-PGA、FC-PGA2 等)、接口类型(例如,Socket 370、Socket A、Socket 478、Socket T、Slot 1、Socket 940 等)、前端总线频率(FSB)等。因此,核心类型在某种程度上决定了 CPU 的性能。

一般说来,新的核心类型往往比老的核心类型具有更好的性能,例如,Northwood 核心的 Pentium 4 1.8 GHz 的性能就要好于 Willamette 核心的 Pentium 4 1.8 GHz。但这也不是绝对的,这种情况一般发生在新核心类型刚推出时,由于技术不完善或新的架构和制造工艺不成熟等原因,可能会导致新的核心类型的性能反而不如老的核心类型的性能。例如,早期 Willamette 核心、Socket 423 接口的 Pentium 4 的实际性能,就不如 Tualatin 核心、Socket 370 接口的 Pentium III 和 Celeron,现在的低频 Prescott 核心的 Pentium 4 的实际性能不如同频的 Northwood 核心的 Pentium 4 等。但随着技术的进步以及 CPU 制造商对新核心的不断改进和完善,新核心的中后期产品的性能必然会超越老核心的产品。

CPU 核心的发展方向是更低的电压、更低的功耗、更先进的制造工艺、集成更多的晶体管、更小的核心面积(这会降低 CPU 的生产成本从而最终会降低 CPU 的销售价格)、更先进的流水线架构和更多的指令集、更高的前端总线频率、集成更多的功能(例如集成内存控制器等)以及双核心和多核心(也就是 1 个 CPU 内部有两个或多个核心)等。CPU 核心的进步对普通消费者而言,最有意义的就是能以更低的价格买到性能更强的 CPU。

3.6 典型 CPU 介绍

从第一台 PC 所使用的 Intel 8088 到今天,CPU 经历 7 代的发展历程,图 3-21 列出各个时代典型的 CPU 型号。从图中可看出,主流 CPU 一直是以 Intel、AMD、Cyrilx 公司的产品为主。由于 Cyrilx 公司被我国台湾的威盛公司并购后,推出的 CPU 很少,所以目前形成了 CPU 生产厂商 Intel 与 AMD 两雄争霸的局面。本节将对主流的 Intel 与 AMD 的 CPU 做一简要介绍。

3.6.1 Intel 系列 CPU

1971 年,当时还处在发展阶段的 Intel 公司推出了世界上第一款微处理器 4004。从此以后,Intel 便与微处理器结下了不解之缘。1978 年,Intel 公司再次领导潮流,首次生产出 16 位的微处理器,并命名为 i8086,同时还生产出与之相配合的数学协处理器 i8087。这两种处理器使用相互兼容的指令集,但在 i8087 指令集中增加了一些专门用于对数、指数和三角函数等数学计算的指令。由于这些指令集应用于 i8086 和 i8087,所以人们也把这些指令集统称为 x86 指令集。虽然以后 Intel 又陆续生产出 80286、80386 等更先进和更快的新型 CPU,但都仍然兼容原来的 x86 指令,而且 Intel 在后继 CPU 的命名上沿用了原先的 x86 序列(80286、80386、80486),直到后来因商标注册问题(美国的法律禁用阿拉伯数字注册),Intel 将 80486 后的新一代 CPU 用拉丁文 Pentium 注册商标。Pentium 在拉丁文里面就是“五”的意思了。Intel 公司还替它起了一个相当好听的中文名称“奔腾”。Pentium 的后继产品分别以 Pentium II、Pentium III、Pentium 4 命名。

Intel 公司从 Pentium II 推出一种专门为低端计算机设计的微处理器,这种微处理器使用新的品牌名称 Celeron,中文名称为“赛扬”。它与同档次 Pentium 的主要区别是减少了 CPU 内集成的

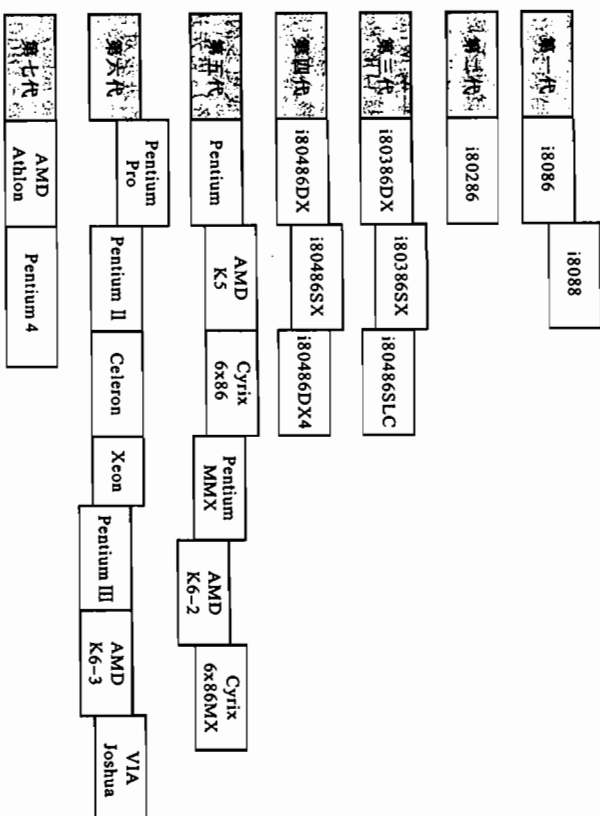


图 3-21 各个时代典型的 CPU 型号

二级缓存的容量(减少了3/4),由于赛扬有很高的性价比,所以受到了低端用户的欢迎。

2000 年,Intel 发布了 Pentium 4 处理器。基于 Pentium 4 处理器的个人计算机,可以让用户创建专业品质的影片,通过因特网传递电视品质的影像,并进行实时语音、影像通信,实时 3D 渲染,快速进行 MP3 编码/解码运算,在连接因特网时运行多个多媒体软件。Pentium 4 还引入了 NetBurst 新结构,NetBurst 是 Intel 应用时间最长的一代构架。以下是 NetBurst 结构带来的好处:

- ① 更快的系统总线。
- ② 高级传输缓存。
- ③ 高级动态执行,包含执行追踪缓存和高级分支预测。
- ④ 超长管道处理技术。
- ⑤ 快速执行引擎。
- ⑥ 高级浮点运算以及多媒体指令集(SSE2)等。

当程序指令与数据开始处理时,会进入系统总线队列。Pentium III 处理器系统总线外频设定在 133 MHz,每时钟周期传输 64 位数据,提供 $8 \text{ B} \times 133 \text{ MHz} = 1.064 \text{ MBps}$ 的数据带宽;而 Pentium 4 处理器的系统总线虽然仅为 100 MHz,同样是 64 位数据带宽,但由于其利用了与 ACPI 相同的“四倍速”(即 FSB400)技术,因此数据传输速率高达 3 200 MBps。因此,Pentium 4 处理器向系统的其他部分传输数据的速率比目前所有的 x86 处理器都快,也打破了 Pentium III 处理器受

系统总线瓶颈的限制。其后 Intel 又不断改进系统总线技术,推出了 FSB533、FSB800 的新规格,将数据传输速率进一步提升。

Pentium 4 采用 SSE2 指令集,这套指令集增加了 144 条全新的指令。对于 128 位的压缩数据,SSE 指令集,仅能以 4 个单精度浮点数的形式来处理,而 SSE2 指令集能采用多种数据结构来处理:

- 4 个单精度浮点数(SSE 和 SSE2)
- 2 个双精度浮点数(SSE2)
- 16 个字节数(SSE2)
- 8 个字数(word)(SSE2)
- 4 个双字数(SSE2)

1. Willamette 核心的 Pentium 4

这是早期的 Pentium 4(如图 3-22 所示)和基于 Pentium 4 的 Celeron 采用的核心,最初采用 Socket 423 接口,后来改用 Socket 478 接口(Celeron 只有 1.7 GHz 和 1.8 GHz 两种,均采用 Socket 478 接口),采用 0.18 μm 制造工艺,前端总线频率为 400 MHz,主频范围从 1.3 GHz 到 2.0 GHz(Socket 423)和 1.6 GHz 到 2.0 GHz(Socket 478),二级缓存分别为 256 KB (Pentium 4)和 128 KB(Celeron),核心电压 1.75 V 左右,封装方式采用 Socket 423 的 PGA INT2、PGA INT3、OO1 423、PGA FC - PGA2 和 Socket 478 的 PGA FC - PGA2 以及 Celeron 采用的 PGA 等。Willamette 核心制造工艺落后,发热量大,性能较低,已经被淘汰,为 Northwood 核心所取代。

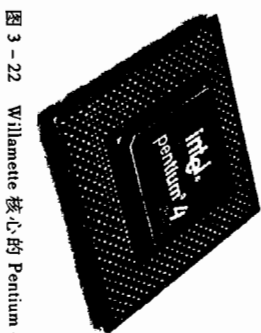


图 3-22 Willamette 核心的 Pentium 4

2. Northwood 核心的 Pentium 4

这是目前主流的 Pentium 4(如图 3-23 所示)和 Celeron 所采用的核心,其与 Willamette 核心最大的改进是采用了 0.13 μm 制造工艺,并且都采用 Socket 478 接口,核心电压 1.5 V 左右,二级缓存分别为 128 KB(Celeron)和 512 KB(Pentium 4),前端总线频率分别为 400/533/800 MHz(Celeron 都只有 400 MHz),主频范围分别为 2.0 GHz 到 2.8 GHz(Celeron)、1.6 GHz 到 2.6 GHz(400 MHz FSB Pentium 4)、2.26 GHz 到 3.06 GHz(533 MHz FSB Pentium 4)和 2.4 GHz 到 3.4 GHz(800 MHz FSB Pentium 4),并且 3.06 GHz Pentium 4 和所有的 800 MHz Pentium 4 都支持超线程技术(hyper-threading technology),封装方式采用 PGA FC - PGA2 和 PGA。按照 Intel 的规划,Northwood 核心很快又会为 Prescott 核心所取代。

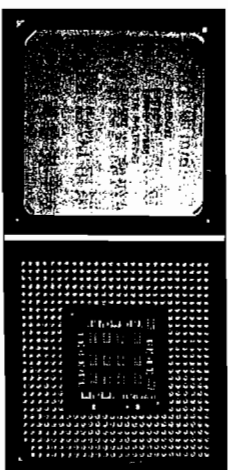


图 3-23 Northwood 核心的 Pentium 4

3. Prescott 核心的 Pentium 4

这是 Intel 最新的 CPU 核心,目前仅用于 Pentium 4,如图 3-24 所示。与 Northwood 最大的

区别是, Prescott 核心采用了 0.09 μm 制造工艺和更多的流水线结构, 初期采用 Socket 478 接口, 以后会全部改用 LGA 775 接口, 核心电压 1.25 ~ 1.525 V, 前端总线频率为 533 MHz (不支持超线程技术) 和 800 MHz (支持超线程技术), 主频分别为 533 MHz FSB 的 2.4 GHz 和 2.8 GHz 以及 800 MHz FSB 的 2.8 GHz、3.0 GHz、3.2 GHz 和 3.4 GHz; 与 Northwood 相比, 其一级高速缓存从 8 KB 增加到了 16 KB, 而二级高速缓存则从 512 KB 增加到了 1 MB, 封装方式采用 PPGA。按照 Intel 的规划, Prescott 核心会很快取代 Northwood 核心, 并且很快就会推出 Prescott 核心的 533 MHz FSB 的 Celeron。

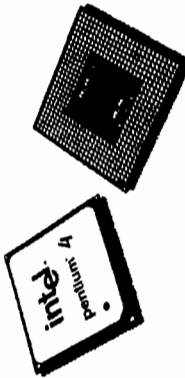


图 3-24 Prescott 核心的 Pentium 4

4. Smithfield 核心的 Pentium D

Smithfield 核心的 CPU 是在一块单晶片硅上结合了两 Prescott 核心。CPU 频率发展的摩尔定律遭遇挫折后, 中高端 CPU 正在向双核或多核发展。进入 2006 年后, 双核 CPU (双核处理器即是基于单个半导体硅晶片的一个处理器上拥有两个功能一样的处理器核心, 也就是将两个物理处理器核心整合到一个核心中) 已成为服务器、工作站, 甚至普通台式计算机和笔记本电脑的标准配置。

Pentium D 是 Intel 公司的双核心处理器系列之一, 如图 3-25 所示。它把两颗 Pentium 4 Prescott 核心放在同一块芯片上, 两个核心共享前端总线, 每个核心都拥有独立的 1 MB 二级高速缓存, 两个核心一共拥有 2 MB。Pentium D 系列中的“D”就代表双核心, 包含 Pentium D 8xx 和 9xx 两大系列, 主要面向入门级服务器及主流的工作站用户。

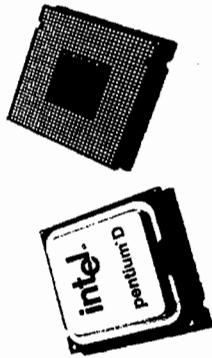


图 3-25 Pentium D

(1) Pentium D 8xx

首批 8xx 系列的 90 nm Pentium D 的核心名称为 Smithfield, 包括 820 (2.8 GHz)、830 (3.0 GHz)、840 (3.2 GHz) 三个型号, 采用 800 MHz 系统总线, 面向主流市场。每个核心拥有独立的总线接口、独立的 16 KB 的一级高速缓存, 独立的支持 ECC 的 1 MB 的二级高速缓存, 支持 SSE2 和 SSE3 多媒体扩展指令集, 采用 LGA775 封装。另外, Pentium D 830 和 Pentium D 840 还支持增强的 Intel Speed Step 技术, 这项技术从以往在移动 CPU 的应用上来看, 能有效地降低噪音和发热。

该处理器不支持超线程技术, 支持 Pentium D 的主板是 i945/955 系列及其更高版本。

(2) Pentium D 9xx

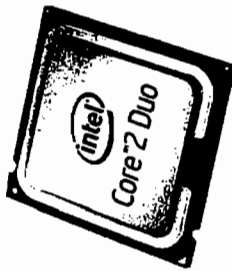
在 Intel 率先采用 65 nm 生产工艺后, 又推出了 Prescott 核心的 Pentium D 900 系列处理器。新一代的 Pentium D 900 系列包括 920 (2.8 GHz)、930 (3.0 GHz)、940 (3.2 GHz)、950 (3.4 GHz) 这 4 种规格处理器, 除此之外, Intel 还将推出一款主频高达 3.6 GHz 的 Pentium D 960 处理器, 它们都采用了更先进的 65 nm 制造工艺, 二级高速缓存容量由 1 MB \times 2 提升为 2 MB \times 2 (即 4 MB), 支持 Intel 虚拟化技术 (可以在一台电脑上虚拟实现多个操作系统, 然后让每个操作系统

各司其职), 采用了 65 nm 工艺。将二级高速缓存扩大了一倍。就内部流水线状况来讲, Presler 和 Smithfield 并无太多差别, 但 Presler 的二级高速缓存是 2 048 KB, 达到了 Prescott 两倍, Presler 的晶体管数达到了 3.76 亿只, 而 Smithfield 只有 2.3 亿只, 而且先进的 65 nm 工艺使 Presler 的发热和功耗都比同频率的 Smithfield 低。

5. Intel 全新的微架构 Core

在 2006 年度春季的 Intel 开发者论坛 (Intel Developer Forum, IDF) 上, Intel 宣布了下一代处理器将统一采用全新的微架构, 并将其正式命名为 Core (酷睿)。

Core 微架构拥有双核心、64 位指令集、4 发射的超标量体系结构和乱序执行机制等技术, 使用 65 nm 制造工艺生产, 支持 36 位的物理寻址和 48 位的虚拟内存寻址, 支持包括 SSE4 在内的 Intel 所有扩展指令集。Core 微架构的每个核心拥有 32 KB 的一级指令缓存、32 KB 的二级缓存。Core 双核处理器如图 3-26 所示。



基于 Core 微架构的笔记本电脑处理器的产品代号为“Merom”, 台式计算机处理器的产品代号为“Conroe”, 而服务器处理器的产品代号为“Woodcrest”。拥有如此强悍性能的 Core 微架构处理器在功耗方面将比先前的产品有大幅下降, 每种产品都拥有自己的最高热设计功耗 (TDP), Merom 最高为 35 W, Conroe 最高为 65 W, Woodcrest 最高为 80 W。

基于 Core 微架构的 Core 双核处理器改写了计算规则, 并且可以提供突破性的性能以及超低的功耗, 是 Intel 卓越台式机和笔记本电脑平台 (分别为 Intel 欢跃技术和 Intel 迅驰移动计算技术) 的核心组件。Core 微架构的特点主要体现在下面的 5 项新技术。

(1) Intel 宽区动态执行技术

Core 微架构具有 4 组解码器, 而之前的 P6 微架构和 AMD K8 微架构都只有 3 组解码器。解码器的增加如同公路车道的增加, 单位时间内处理的指令更多。Core 微架构还首次加入了“宏融合”技术。在传统微架构中, 每个 x86 指令的解码是完全独立的, 但 Intel 发现某些特定关系的指令是可以合并的, 可以在一个时钟周期内执行。这样, Core 微架构在特殊情况下, 每个时钟周期内 4 个解码器能执行 5 组指令。据 Intel 解释, 在大部分程序中, 约每 10 ~ 15 条指令就会有 1 组可以被合并。

新增的 1 组解码器, 结合“宏融合”技术, 在不增加频率的情况下提高了处理器执行指令的速度。

(2) Intel 智能功率能力

Pentium M 的节电性能是不是曾经很诱人? 在节电方面, Core 微架构采用了 Intel 移动微架构的一些技术, 例如 Speed-Step 频率控制, 但 Core 微架构走得更远, 增加了更多的新技术。除了更先进的 65 nm 生产工艺、应变硅、低-K 电介质材料等技术外, 最重要的是 Core 微架构加入了逻辑控制机制, 可独立开、关处理器的各运算单元, 只有在需要时才会被开启, 不需要时自动关闭, 这样可以节省不少电能, 进一步降低处理器的功耗。

(3) Intel 高级智能高速缓存

过去, Pentium D 只是将两颗核心封装在一起, 二级高速缓存是独立的, 两颗核心不能直接沟通。如果一颗核心的二级高速缓存需要读取另一颗核心的二级高速缓存, 则需要通过前端总线及北桥来完成, 这样势必造成在某些应用中的延迟。例如, 对于高效的视频编码程序, 可以让两颗核心分别担任解码和编码工作, 数据直接在二级高速缓存中快速交换, 而独立的二级高速缓存会影响数据交换速度。

Core 微架构对此进行了改进: 两颗核心能共用二级高速缓存, 当一颗核心的工作负担加重时, 可以获得更多的二级缓存空间, 而且一颗核心能直接读取另一颗核心的二级缓存数据。

(4) Intel 智能内存访问

AMD K8 微架构整合了内存控制器, 因而在读/写性能方面更优异。而 Core 微架构仍然将内存控制器交给了北桥。据 Intel 解释, 这样做是因为处理器的发展周期和内存的发展周期不同步, Intel 可以升级北桥来使用最新的内存 (例如 DDR2), 而不用重新设计处理器。不过, Core 微架构为了与内存快速交换数据, 在处理器中加入了称为“内存消歧”的技术。

在传统微架构中, 内存读/写是按指令顺序排队执行的。Core 微架构的“内存消歧”技术加入了智能分析机制, 能预知某些读/写操作是完全独立的, 能使其提早执行或同时执行。这样就缩短了处理器对内存的读/写等候时间, 从而提升效率。

(5) Intel 高级数字媒体增强技术

从 Pentium MMX 处理器开始, Intel 就不断给处理器增加新的多媒体指令, 例如, SSE、SSE2 及 SSE3。Core 微架构针对 SSE 指令执行进行了改良, 称为“Intel 高级数字媒体增强技术”。Intel 传统微架构在执行 128 位 SSE、SSE2 及 SSE3 指令时, 需要把指令分拆为两个 64 位指令 (需要两个时钟周期), 而 Core 微架构只需要一个时钟周期就能完成, 执行效率提高一倍。目前, SSE 指令集已被普遍地用于各种流行软件中, 包括视频、音频、数学运算等软件。

3.6.2 AMD 系列 CPU

作为排在 Intel 公司之后世界第二大 CPU 制造商, AMD 一直是 Intel 最有力的竞争对手。AMD 打破了 Intel 在微处理器领域的垄断局面。K5 是 AMD 第一款独立生产的 x86 级 CPU, 但其寿命较短, 不久便被 K6、K6-2 和 K6-III 所取代。目前, 主流的 AMD CPU 是 K7 与 K8 系列。

1. AMD K7 系列 CPU

K7 是 K6 处理器之后, AMD 推出的第七代处理器, 很多方面比 Intel 的 Pentium 4 先进。1999 年 8 月推出了第一款 K7 系列 CPU, AMD 将其命名为 Athlon (速龙), 如图 3-27 所示。与当时 Intel 的 Pentium III 处理器竞争, 它采用了 0.25 μm 的制造工艺, 起始主频为 500 MHz, 无论是主频还是性能都超过了当时的 Pentium III 处理器。K7 系列的台式计算机处理器包括 Athlon XP、Athlon 和 Duron 三个系列。其中 Athlon 1.4 GHz、1.3 GHz、1.2 GHz、1.1 GHz 和 1.0 GHz 采用了 200 MHz 前端总线, 所以只能同 PCI600 内存搭配使用; 前端总线频率为 266 MHz 的 Athlon 处理器的主频有 1.4 GHz、1.33 GHz、1.2 GHz、1.13 GHz 和 1.0 GHz, 它们可以与 PC2100 内存搭配使用。这些处理器



图 3-27 雷鸟 Thunderbird

都采用了 Thunderbird (雷鸟) 核心, 0.18 μm 工艺生产, 配备 128 KB 一级高速缓存, 256 KB 全速二级缓存, 核心电压也仍然为 1.75 V。

现在最新的 Athlon 处理器有属于 K8 的 Athlon 64 系列, 专为 AMD 64 位平台而设计, 并兼容现有的 x86 平台。

雷鸟的最大特点是其具有 128 KB 一级高速缓存、256 KB 二级高速缓存, 相对于当时 Pentium III 和 Pentium 4 处理器具有一定的优势, 因此其性能同 Intel 处理器不相上下, 甚至略胜一筹。

Athlon XP 处理器 (如图 3-28 所示) 采用了增强型 0.18 μm 铜布线工艺的 Palomino 核心, 延续了 Athlon TB 的 Socket A 接口, 不过采用了全新的 OPGA (Organic Pin Grid Array) 有机封装、硬件预取功能 (Hardware Prefetch)、SSE 指令集及“Quantispeed Architecture”加速架构, 并加入了温控二极管, 高速缓存仍采用 384 KB 的配置 (128 KB 一级高速缓存 + 256 KB 二级高速缓存), 工作电压为 1.75 V。

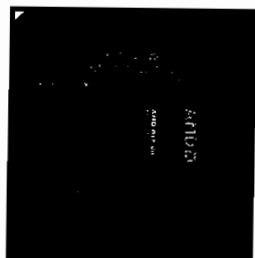


图 3-28 Athlon XP CPU

AMD 第一次推出 K7 以及 Athlon 的时候, 采用的是真实频率标称, 例如 1 GHz 的 Athlon 的运行频率就是 1 000 MHz。从 Athlon XP 处理器起, AMD 不再采用这样的方法, 开始采用“真实性能标称”的叫法, “真实性能标称”就是 CPU PR 值。如 XPI600 + 处理器并它的实际工作频率是 1.4 GHz 左右。

支持 SSE 是 Athlon XP 的另一项改进之处, 最早的 Athlon 处理器采用了 Enhanced 3DNow! 指令集, 除了兼容 MMX 多媒体指令集之外, 还加入了 19 条全新的指令。AMD 在 Athlon XP 上又加入了 59 条全新的指令, 指令集称为 3DNow! Professional, 提供了与 Intel 的 SSE 指令集完全兼容的功能。另外, Athlon XP 中还加入了温控二极管, 这样可以提供比主板上的温度探头更加精确的温度测量, 从而保证了处理器更加安全。

从 Athlon XP 2200 + 起, Athlon XP 系列开始采用 Thoroughbred 核心, 这是 AMD 第一款采用 0.13 μm 制程的台式计算机处理器, 由 3 720 万个晶体管组成, 它的实际频率为 1.8 GHz, 相对于 Palomino 核心并没有很大的提高。随后 AMD 发布的 Athlon 2400 + /2600 + 在其基础上进行了优化, 晶体管数目提升到了 3 760 万个, 核心中增加了一层金属层 (由 8 层变成了 9 层), 以减少晶体管间的阻抗和容抗。当然, 还有其他的一些改变, 都是为了降低处理器内部的电磁干扰, 提高处理器运行期间的稳定性。这两款处理器的频率分别达到了 2 GHz 和 2.13 GHz。

2. Duron 处理器

自从 20 世纪 90 年代末, Intel 率先推出了 Pentium 处理器的精简版 Celeron 并大获成功之后, 为了与 Intel 抗衡, AMD 在不久也推出了当时高端雷鸟处理器的精简版本——Duron 处理器, 如图 3-29 所示。

Duron 的中文名为“毒龙”。根据 AMD 的说法, Duron 源于拉丁字根“durare”, 意思是“持久、耐久”, 而“on”则表示“单元”的意思, 所以 Duron 就是“持久、耐久的单元”的意思。Duron 处理器主要面向低端市场, 最初的 Duron 采用 Spitfire 核心。随着 Palomino 核心的



图 3-29 Duron 处理器

Athlon XP处理器的发布,AMD推出了采用Morgan核心的新一代Duron处理器,同样采用了增强型的0.18 μm 工艺,功耗及发热量相比上一代的Spirifire核心有了明显的改进。Duron处理器配置了128 KB一级缓存和64 KB二级缓存,因此其性能与Athlon XP相比还是有相当的差距的。另外,Morgan还加入了效能极高的SSE指令集,虽然它仍就只有64 KB二级缓存及200 MHz系统总线,但与同频的Celeron相比还是有优势的,而同Pentium 4 Celeron处理器相比的话,则没有什么优势。毕竟它的最高主频只有1.3 GHz和1.4 GHz,而Pentium 4 Celeron处理器的主频已经达到了1.8 GHz。

3. Thoroughbred 核心的 Sempron 系列处理器

随着2004年6月下旬Intel Prescott核心的Celeron D处理器的发布,AMD新一代对抗Intel Celeron的品牌也在2004年7月底面世,它就是全新的Sempron系列处理器,如图3-30所示。

Sempron的名字其实是来源于拉丁语Semper,就是“永远”的意思,后缀“ron”则表示它是一款价格便宜的处理器。其中文译名为“闪龙”,意思是“如闪电般迅捷,如恒星般闪耀”。AMD Sempron系列处理器作为主或中低端市场的产品,其主要对手是Intel 90 nm工艺的Celeron D处理器。

Sempron处理器在型号命名方面,沿用了AMD独有的PR值标称命名方法,AMD推出了Sempron 2200 +、2300 +、2400 +、2500 +、2600 +和2800 +六款产品,工作频率从1.5 GHz到2.0 GHz。这六款Sempron处理器仍然是采用AMD沿用多年的Socket 462接口,并采用与Athlon XP完全一样的Thoroughbred核心,前端总线为333 MHz,一级缓存和二级缓存分别为128 KB和256 KB。与以往Thoroughbred核心的Athlon XP相比,Sempron的PR值明显高出一截。如果说以前Athlon XP的PR值是相对于Intel Pentium系列处理器的话,Sempron较高的PR值可以认为是针对Intel的Celeron系列的。

2005年,AMD对其台式机处理器产品线进行了调整,已经服役多年的Socket 462接口正式退出市场,在低端取而代之的是Socket 754接口的Sempron处理器。Socket 754接口的Sempron处理器在架构上比Athlon XP先进得多,采用了Athlon 64的架构,使用了Hyper Transport传输技术以及内置内存控制器,使得CPU与内存之间的延时得到了大幅的降低,并且支持SSE2指令集。

4. AMD K8 系列 CPU

2004年,AMD推出的新产品中采用了全新的x86微处理器架构,提供64位计算的支持。此新架构是在原来x86-32系统架构的基础上发展过来的,按照延续下来的习惯叫法,我们称之为x86-64微处理器架构。但是,作为独立研发出来的架构,AMD将之称为AMD 64架构。而基于AMD 64架构的一系列处理器,AMD将之统称为K8处理器。

(1) Athlon 64 处理器技术特点

1) 64 位计算能力

这应该是Athlon 64(如图3-31所示)与传统32位处理器最大的区别。在Athlon 64和Athlon 64/FX处理器中,AMD加入了一个被称为x86-64的指令集,正是这个指令集赋予了Athlon 64系列处理器的64位的计算能力。同时,AMD清醒地认识到,64位应用目前还远未成为主流,所以,他们并没有将Athlon 64/FX设计成一款仅支持64位计算的处理器,而是在支持64位代码的基础上提供了对32位和16位代码的良好兼容。使得Athlon 64/FX在执行32位代码时不必重新进行编译,从而避免了效率低下所导致的性能损失。总体来说,Athlon 64/FX既提供了64位计算能力,又完全兼容目前的32位甚至16位应用,是一款全能型处理器。



图3-31 Athlon 64 处理器

2) 超过4 GB的内存寻址能力

除了64位计算能力,x86-64指令集还使Athlon 64系列处理器拥有了64位的地址空间和64位的数据空间,换句话说就是x86-64指令集使Athlon 64/FX可以支持4 GB以上的内存(64位处理器理论上支持 2^{64} 寻址空间,可支持上亿GB的内存),而传统的32位处理器最高仅支持4 GB内存。更大容量的内存支持能力使Athlon 64/FX在诸如大型数据库、CAD/CAM建模以及越来越逼真的3D游戏等高端应用中有着显著的优势。

3) 处理器内部集成内存控制器

这也是Athlon 64系列的独门绝技之一。与传统的处理器-内存控制器(北桥)-内存的方案相比,内存控制器内置可以极大地降低数据收发延迟,缩短读/写请求的响应时间,处理器的性能将因此获得可观的提升。除此之外,处理器内部集成内存控制器还可以简化北桥芯片的结构。目前,一些芯片厂商已经成功设计出了单芯片的Athlon 64主板产品,其中,nVIDIA的nForce3/4系列无疑是杰出的代表。

4) HyperTransport 总线

在Athlon 64处理器架构中,HyperTransport总线负责的是处理器与北桥(或单芯片的nForce3/4)芯片间的数据传输,它的带宽直接影响系统的整体性能。目前,Athlon 64的HyperTransport总线共有两种频率:800 MHz和1 GHz,它们均采用了与DDR内存相同的Dual Pump技术,在一个时钟周期内可以传输两次数据,其上下行位宽均为16位,所以,800 MHz HyperTransport总线的数据带宽为6.4 Gbps,而1 GHz HyperTransport总线的数据带宽则达到了8 Gbps。

5) Execution Protection 防病毒技术

Execution Protection技术是AMD公司应用在Athlon 64中的又一项新技术,这项技术可以防止病毒进行缓冲区溢出攻击(指病毒使CPU因缓冲区溢出而执行恶意代码的攻击技术)。应用这项技术后,CPU缓冲区的数据将会只读而不能执行,可以有效地防止恶意代码的执行。目前,这项技术已经被封装在了Athlon 64系列CPU中,不过还需要Windows XP Service Pack 2的支持方能发挥作用。

6) Cool 'N' Quiet 节能技术

顾名思义, Cool 'N' Quiet 技术是一项旨在降低处理器功耗的技术, 它可以根据执行任务的情况动态调节 CPU 工作频率, 并通过搭配的测温器件自动控制 CPU 散热器的转速, 以达到节能降噪的目的。要实现 Cool 'N' Quiet 技术, 除了 CPU 之外, 相应的驱动程序以及主板 BIOS 的支持也是必不可少的。

(2) Athlon 64 系列 CPU 的核心类型

1) ClawHammer 核心

ClawHammer 核心是面向主流台式机市场的 Athlon 64 系列处理器使用的第一款核心, 它采用了 0.13 μm SOI 制造工艺, 具有 128 KB 一级缓存和 1 MB 二级缓存。早期采用 ClawHammer 核心的 Athlon 64 处理器只有 Socket 754 一种接口, 其 HyperTransport 总线频率为 800 MHz, 并且不支持双通道 DDR 内存。但是随着时间的推移, AMD 又将这款核心引入了高端市场, 推出了新款的 Athlon 64 FX-55、Athlon 64 FX-53 以及 Athlon 64 4000+。这三款产品均采用了 Socket 939 接口, HyperTransport 总线频率提升到了 1 GHz, 并且全部支持双通道 DDR 内存。

2) NewCastle 核心

基于 NewCastle 核心的有 Athlon64 3800+ 处理器。

在 ClawHammer 核心的 Athlon 64 处理器中, AMD 集成了 1 MB 的二级缓存, 这样做虽然有助于 CPU 性能的提升, 但是也带来了成本居高不下以及良品率太低的问题, 于是开发了 NewCastle 核心。NewCastle 核心仍然采用了 0.13 μm SOI 制造工艺, 但是却将二级缓存的容量降为 512 KB, 核心面积也由原来的 193 mm^2 缩小为 144 mm^2 , 制造成本远低于 ClawHammer 核心, 有利于 Athlon 64 处理器的推广。

采用 NewCastle 核心的 Athlon 64 共有两大系列: Socket 939 接口系列和 Socket 754 接口系列, 它们的主要区别在于: Socket 939 系列具有 1 GHz 的 HyperTransport 总线频率, 支持双通道 DDR 内存, 而 Socket 754 系列的 HyperTransport 总线频率为 800 MHz, 不支持双通道 DDR 内存。

3) Winchester 核心

Winchester 核心的最大改进在于将制造工艺由 0.13 μm 提升到了 0.09 μm 。目前, 采用 Winchester 核心的 Athlon 64 CPU 均为 Socket 939 接口, 具有 512 KB 的二级缓存和 1 GHz 的 HyperTransport 总线, 并且提供了对双通道 DDR 内存的支持。

(3) Socket AM2 处理器

新一代 Socket AM2 处理器包括代号 Windsor 的 Athlon 64/FX、Athlon 64 X2 双核心处理器、代号 Orleans 的 Athlon64 单核心处理器及主攻低端代号 Manila 的 Sempron 处理器, 它们还是采用 90 nm 制程技术, 据 AMD 的白皮书显示, Windsor 和 Orleans 最主要的改良只在于内存控制器, 由支持双通道 DDR400 提升为双通道 DDR2 800, 另外还加入 AMD 虚拟化技术, 它的用途类似 Intel 的 VT 功能, 有效提高了系统运作虚拟多作业系统的性能、可靠度及安全性, 除此之外, 核心的主要架构及规格并没有作出任何改良, 也意味着 Socket AM2 版本的 Athlon 64/FX、Athlon 64 X2 及 Athlon 64 与 Socket 939 相比, 但相同频率下并不会出现很明显的性能增长。相反, 新一代 Socket AM2 低端 Sempron 处理器则较令人注目, 上一代 Socket 754 只支持单通道 DDR400, 新一代 Manila 核心则提供至支持双通道 DDR2-800, 因此与上代在相同频率下, 会有较显著的性能增长。



一、选择题

- 下面列出的 Intel 系列 CPU 中支持 MMX 技术的是 ()。
 - 80286
 - 80386
 - 80486
 - Pentium II
- Pentium 4 和 Celeron D 两种 CPU 的主要区别是 ()。
 - 前者比后者时钟频率高
 - 前者比后者寻址范围大
 - 前者有多媒体指令而后者没有
 - 前者的 Cache 比后者大
- 下列关于二级高速缓存说法有误的是 ()。
 - 二级高速缓存比一级高速缓存大
 - 二级高速缓存的设置始于 80486 时代
 - 二级高速缓存最大限度减小了主内存对 CPU 运行造成的迟缓
 - 二级高速缓存的设置是为了解决主内存不足的问题
- 下列因素中, 与高速缓存的命中率无关的是 ()。
 - 主存的存取时间
 - 块的大小
 - 高速缓存的组织方式
 - 高速缓存的容量
- 下列关于 MMX 技术和 3D NOW! 技术说法正确的是 ()。
 - MMX 侧重浮点运算
 - 3D NOW! 侧重整数运算
 - MMX 技术是由 AMD 公司推出的
 - 3D NOW! 主要针对三维建模、坐标变换等三维应用场合
- 下列说法正确的是 ()。
 - 主频、外频和倍频系数是 CPU 工作频率的不同描述
 - PC 中各分系统的时钟频率都与系统时钟相同
 - CPU 外频提高后, 主频也相应提高
 - 数据传输最大带宽即同时传输的数据位宽度
- 缓冲的全写法是指 ()。
 - 不通过高速缓存来写入数据
 - 将高速缓存的内容清除
 - 同时写入高速缓存与系统内存
 - 不将数据写入系统内存
- 缓冲器在处理器运算上最大的优点是 ()。
 - 数据传输快
 - 具备大容量
 - 线路设计简单
 - 可以随时增加
- 下列因素与系统性能提高无关的是 ()。
 - CPU 主频
 - 指令结构
 - CPU 与内存之间的存取速度
 - CPU 芯片封装方式
- CPU 与内存之间的内容是 ()。
 - 高速缓存中的内容
 - 是主存储器中的部分地址的副本
 - 主存容量的扩充
 - 主存储器部分地址的副本

二、填空题

- 具有 16 位数据线、32 位地址线的 CPU 可以访问的物理空间为 ()。
- 所谓的超标量 CPU, 就是指集成了多个 ALU、多个 FPU、多个译码器和多个流水线的 CPU, 以 ()

处理的方式来提高性能。

3. 高速缓存的3种基本结构是:()、()和()。
4. 在高速缓存的地址映射中,若主存中的任意一块均可映射到高速缓存内的任意一块的位置上,则这种方称为()。

5. 计算机字长取决于()总线的宽度。

三、简答题

1. 决定CPU性能的主要技术参数有哪些?
2. CPU的主频、外频和倍频系数的关系如何?什么是超频?
3. 为什么处理器中必须有缓存器的设计?
4. 试列出第六代和第七代CPU的典型产品。
5. CPU与主板连接的接口有哪几种形式?它们互相兼容吗?决定CPU性能主要技术参数有哪些?
6. 计算机访问存储器有哪两种局部规律?什么是存储器的层次结构?
7. 简述高速缓存的作用是什么?

8. 什么是MMX技术?主要用于提高CPU哪方面的性能?

9. 解释下列名词。

Cache MMX SSE 3DNow! IA-32 EM64T

实 验

1. 使用测试软件(PC Wizard 或 HARDINFO 等),选择至少两台PC,检测其CPU信息,包括CPU的名称、生产厂商、时钟频率、核心电压、超频检测、CPU所支持的多媒体指令集、CPU的一级高速缓存和二级高速缓存的资料(大小、速度、技术)等。
2. 采用基准程序法测试CPU性能,包括CPU Dhrystone Benchmark(处理器整数性能测试)、CPU Whetstone Benchmark(处理器浮点性能测试)。
3. 测试完成后写出详细的实验报告,内容包括CPU工作频率对性能的影响,缓存对性能的影响同品牌(AMD、Intel)的差异,不同核心的差异,从性能、价格、用途多方面阐述你自己对选择CPU的观点。

内部存储器 第4章

内部存储器简称内存,又称主存,是PC系统中仅次于CPU的重要部件之一,它是用来保存CPU在运行过程中的程序、数据及中间结果的部件。本章主要讲述内部存储器的功能、组成、半导体存储器的的工作原理,并简要介绍PC机中常用的内存类型和内存管理方面的知识。

4.1 存储系统概述

存储系统是计算机的重要组成部分,计算机中的各种信息都存放在存储器中,存储器是根据容量大小、存取速度快慢、成本价格高低等因素按照一定的体系结构组织起来的。根据存储器在计算机中处于不同的位置,可分为主存储器和辅助存储器。在主机内部,直接与CPU交换信息的存储器称内存存储器或主存储器。在执行期间,程序的数存放在内存存储器内。各个存储单元的内容可通过指令随机读/写访问的存储器称为随机存取存储器(RAM)。另一种存储器叫只读存储器(ROM),里面存放一次性写入的程序或数据,仅能随机读出。RAM和ROM共同分享主存储器的地址空间。RAM中存取的数据掉电后就会丢失,而掉电后ROM中的数据可保持不变。因为结构、价格等原因,主存储器的容量受到限制。为满足计算的需要而采用了大容量辅助存储器或称外存储器,如磁盘、光盘和磁带等。外存通常是磁性介质或光盘,能长期保存信息,并且不依赖于电源状态。

图4-1显示了微型计算机系统存储器组织,它呈金字塔结构,越往上,存储器的速度越快,CPU的访问频率越高;同时,每位存储容量的价格也越高,系统的拥有量越小。从图中可以看到,CPU中的寄存器位于该塔的顶端,它有最快的存取速度,但容量极为有限;向下依次是CPU内的高速缓存、主板上的高速缓存(由高速SRAM组成)、主存储器(由DRAM组成)、外部存储器(半导体盘、磁盘)和大容量外部存储器(光盘、磁带);位于塔底的存储设备,其容量最大、每位存储容量的价格最低,但速度可能也是较慢或最慢的。本章主要讨论主存储器级的各种半导体存储器单元、辅存和外

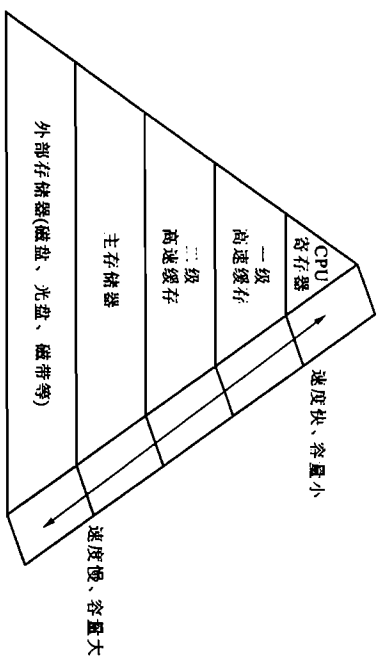


图 4-1 微型计算机存储系统的层次结构

存储器将在后面章节讲述。

4.2 内存的作用及其分类

4.2.1 内存的作用

当计算机启动后,必须执行由一连串由指令组成的程序,才能完成指定的功能。如果这些程序放在磁盘等速度缓慢的数据存储装置上,那么处理器执行程序的速度就会很慢。这就需要更快的数据存储区域,先把计算机要执行的程序放在这个区域,处理器直接从这个区域读取指令和数据。内存的功能就是提供这个快速数据存放区域。其作用是在慢速的外部存储设备和高速的处理器之间承担中间角色,如图4-2所示。

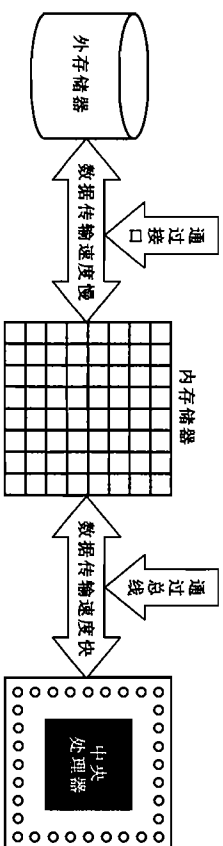


图 4-2 内存的作用

在微型计算机中,通常所说的内存主要指的是动态随机存储器(DRAM),即内存条,其主要

作用如下:

- 暂时存放在执行的程序、原始数据、中间结果和运算结果。
- 作为CPU运行程序的区域。
- 配合CPU与外设打交通。

4.2.2 内存的分类

内存实质是一组或多组具备数据输入输出和数据存储功能的集成电路。内存分为随机存储器和只读存储器两大类。随机存储器(RAM)中的内容可通过指令随机访问。只读存储器(ROM)中存放一次性写入的程序或数据,仅能读取这些程序或数据,不能进行修改。RAM和ROM共同分享主存储器的地址空间。RAM中存放的数据掉电后就会丢失,而掉电后ROM中的数据可保持不变。

现在的RAM多为MOS型半导体电路,它分为静态和动态两种。静态RAM是靠双稳态触发器来记忆信息的;动态RAM是靠MOS电路中的栅极电容来记忆信息的。由于电容上的电荷会泄漏,需要定时给予补充,所以动态RAM需要设置刷新电路。但动态RAM比静态RAM集成度高、功耗低,从而成本也低,适于做大容量存储器。所以主内存通常采用动态RAM,而高速缓冲存储器(Cache)则使用静态RAM。只读存储器又可分为不可改写只读存储器、可改写的只读存储器。不可改写的只读存储器中的信息只能被读出,不能被操作者修改或删除(如掩膜ROM、PROM)。可改写的只读存储器和一般的只读存储器不同点在于其内容可以擦除并重写(如EPROM、EEPROM)。图4-3是内存的分类。

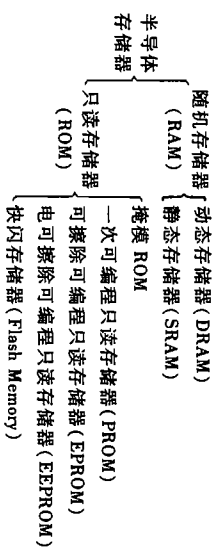


图 4-3 内存分类

4.2.3 内存的主要技术指标

内存的技术指标一般包括存储容量、存取周期、错误校验等。

1. 存储容量

存储器可以容纳的二进制信息量称为存储容量。存储容量的大小以字节数为单位来衡量,其单位为字节(B)、千字节(KB)、兆字节(MB)和吉字节(GB)等。它们之间的换算关系是:

$$1 \text{ KB} = 2^{10} \text{ B} = 1024 \text{ B}$$

$$1 \text{ MB} = 2^{20} \text{ B} = 1024 \text{ KB}$$

$$1 \text{ GB} = 2^{30} \text{ B} = 1024 \text{ MB} = 1048576 \text{ KB}$$

内存的容量一般为几十M到几百M字节。

2. 存取周期

存储器从接收读出命令到被读出信息稳定在存储器数据寄存器(MDR)的输出端为止的时间间隔,称为存取时间;两次独立的存取操作之间所需的最短时间为存取周期。内存的存取周期一般为 $60\text{ ns} \sim 120\text{ ns}$, $1\text{ ns} = 10^{-9}\text{ s}$ 。常见的存取周期有 60 ns 、 70 ns 、 80 ns 、 120 ns 等几种,相应地在内存条上标为 -6 、 -7 、 -8 、 -120 等字样。这个数值越小,存取速度越快,但价格也便随之上升。在选配内存时,应尽量挑选与 CPU 系统总线时钟周期相匹配的内存条。

3. 错误校验

常用的内存错误校验方式有奇偶校验、ECC 和 SPD。

对于内存的奇偶校验(Parity)要从比特概念说起。比特(bit)是内存中的最小单位,也称“位”。它只有两个状态,分别以 1 和 0 表示。我们将 8 个连续的比特叫做一个字节(byte)。非奇偶校验内存的每个字节只有 8 位,若它的某一位存储了错误的值,就会使其中存储的相应数据发生改变而导致应用程序发生错误。

而奇偶校验内存存在每一字节(8 位)外又额外增加了一位作为错误检测之用。比如一个字中存储了某一数值(10011110),把这每一位相加起来($1+0+0+1+1+1+1+0=5$)。若其结果是奇数,校验位就定义为 1,反之则为 0。当 CPU 返回读取存储的数据时,它会再次相加前 8 位中存储的数据,计算结果是否与校验位相一致。当 CPU 发现二者不同时就做出一定的反应。现在主板都可以使用带奇偶校验位或不带奇偶校验位两种内存条,但要注意两种不能混用。

差错校验与纠正(Error Checking and Correcting, ECC)内存也是在原来的数据位上外加一些位来实现的。若是 8 位数据,则需 1 位用于奇偶校验,5 位用于 ECC,这额外的 5 位用来重建错误的信息。当数据的位数增加一倍时,奇偶校验位也增加一倍,而 ECC 位只需增加一位;当数据为 64 位时,所用的 ECC 位和奇偶校验位数相同(都为 8)。

在那些奇偶校验只能检测到错误的地方,ECC 可以纠正绝大多数错误。计算机工作正常时,你不会发觉你的数据出过错,只有经过内存的纠错后,计算机的操作指令才可以继续执行。当然,在纠错时系统的性能明显降低。不过这种纠错对服务器等应用而言是十分重要的。ECC 内存的价格比普通内存要昂贵许多。

串行存在探测(Serial Presence Detect, SPD)是一个 256 字节的 EEPROM 芯片。型号多为 24LC01B,位置一般在内存条正面的右侧,里面记录了诸如内存的速度、容量、电压与行、列地址带宽等参数信息。

当开机时 PC 的 BIOS 将自动读取 SPD 中记录的信息。如果没有 SPD,就容易出现死机或致命错误的现象。它是识别 PC100 内存的一个重要标志。现在个别厂商一方面为了降低生产成本,另一方面又要从表面上迎合 PC100 标准,就在 PCB 板上焊上一片空的 SPD。这样就有可能导致在 100 MHz 以上外频不能正常工作。还应该注意的是一些厂商出的主板(如 Intel 原装主板)一定要 BIOS 检测到 SPD 中的数据才能正常工作,对于使用假 SPD 的内存来说,就会有兼容或死机的现象出现。

4.3 半导体存储器的组成及工作原理

4.3.1 随机存储器

1. 静态 RAM

(1) 静态 RAM 基本存储电路

静态 RAM 的基本存储电路通常由如图 4-4 所示的 6 个 MOS 管组成。在此电路中, $T_1 \sim T_6$ 管组成双稳态触发器, T_1 、 T_2 为放大管, T_3 、 T_4 为负载管。若 T_1 截止,则 A 点为高电平,它使 T_2 导通,于是 B 点为低电平,这又保证了 T_1 的截止。同样, T_1 导通而 T_2 截止,这是另一个稳定状态。因此可用 T_1 管的两种状态分别表示“1”或“0”。由此可知,静态 RAM 存储信息的特点是这个双稳态触发器的稳定状态密切相关。显然,仅仅保持这两个状态的一种是不够的,还要对状态进行控制以便写“1”或写“0”,于是就加上了控制管 T_3 、 T_4 。

当地址译码器的某一个输出线送出高电平到 T_3 、 T_4 控制管的栅极时, T_3 、 T_4 导通。A 点与 I/O 线相连,B 点与 I/O 线相连。这时,如要写“1”,则 I/O 为“1”,I/O 为“0”,它们通过 T_3 、 T_4 管与 A、B 点相连,即 A = “1”,B = “0”,使 T_1 截止, T_2 导通。而当写入信号和地址译码信号消失后, T_3 、 T_4 截止,该状态仍能保持。如要写“0”,I/O 线为“0”,I/O 线为“1”,这使 T_1 导通, T_2 截止,只要不掉电,这个状态会一直保持,除非重新写入一个新的数据。

当要读取所存储的内容时,仍需地址译码器的某一输出线送出高电平到 T_3 、 T_4 管栅极,即此存储单元被选中,此时 T_3 、 T_4 导通,于是 T_1 、 T_2 的状态被分别送至 I/O 和 I/O 线,这样就读出了所存储的信息。所存储的信息被读出后,所存储的内容并不改变,除非重写一个数据。

由于 SRAM 存储电路中,MOS 管数目多,故集成度较低,而 T_1 、 T_2 管组成的双稳态触发器必有一个是导通的,功耗也比 DRAM 大,这是 SRAM 的两大缺点。其优点是不需刷新电路,从而简化了外部电路。

(2) 静态 RAM 的结构

静态 RAM 芯片内部是由很多如图 4-4 所示的基本存储电路组成的,所有基本单元排列成 m 行乘 n 列的存储矩阵。芯片容量为单元数与数据线位数之乘积。为了选中某一个单元,往往利用矩阵式排列的地址译码电路。例如 1 024 个单元的内存芯片,需 10 根地址线,其中 5 根用于行译码,另 5 根用于列译码,译码后在芯片内部排列成 $2^5 = 32$ 条行选择线和 $2^5 = 32$ 条列选择线,这样可选中 1 024 个单元中的任何一个。而每一个单元中的基本存储电路个数与数据线位数相同,即每个存储单元由若干个基本存储电路组成。典型的 SRAM 芯片的功能框图如图 4-5

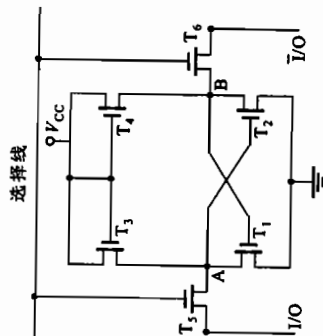


图 4-4 静态 RAM 的基本存储电路

所示。

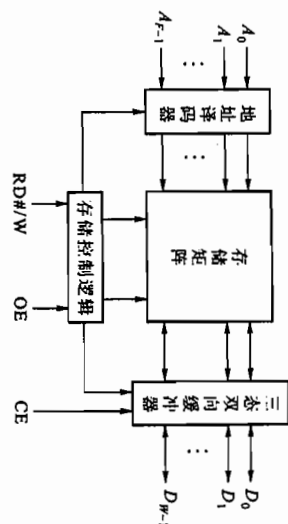


图 4-5 SRAM 芯片的功能框图

2. 动态 RAM

(1) 动态 RAM 的基本存储电路

静态 RAM 每个单元电路需 6 个 MOS 管，功耗大且集成度低。为减少 MOS 管数目，提高集成度和降低功耗，就出现了动态 RAM 器件，其基本存储电路为单管动态存储电路，如图 4-6 所示。由图可见，DRAM 存储信息靠的是电容器 C。电容器 C 有电荷时，为逻辑“1”，没有电荷时，为逻辑“0”。但由于任何电容器都存在漏电现象，因此，当电容器 C 存有电荷时，过一段时间由于电容器的放电过程导致电荷流失，信息也就丢失。解决的办法是刷新，即每隔一定时间（一般为 2 ms）就要刷新一次，使原来处于逻辑电平“1”的电容器的电荷又得到补充，而原来处于电平“0”的电容器仍保持“0”。

在进行读操作时，根据行地址译码，使某一行选择线为高电平，于是使本行上所有的基本存储电路中的 MOS 管 T 导通，使连在每一列上的刷新放大器读取对应存储电容器上的电压值。刷新放大器将此电压值转换为对应的逻辑电平“0”或“1”，并重写到存储电容器。而列地址译码产生列选择信号，所选中那一列的基本存储电路才受到驱动，从而可读取信息。

在写操作时，行选择信号为“1”，T 管处于导通状态，此时列选择信号也为“1”，则此基本存储电路被选中，于是由外接数据线送来的信息通过刷新放大器和 T 管送到电容器 C 上。

刷新是逐行进行的，当某一行选择信号为“1”时，选中了该行，电容器上信息送到刷新放大器上，刷新放大器对这些电容器立即进行重写。由于刷新时，列选择信号总为“0”，因此电容器 C 上的信息不可能被送到数据总线上。

(2) 动态 RAM 的结构

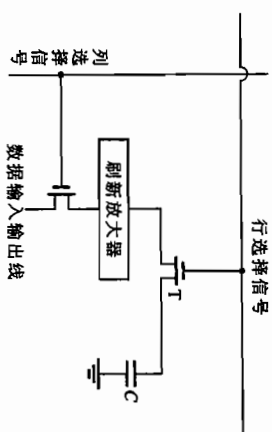


图 4-6 单管动态存储电路

图 4-7 所示是一个 2116 静态存储器的结构。动态 RAM 芯片由存储矩阵、读写控制及刷新控制电路构成。图中，片上的行地址选通 RAS 和列选通 CAS 分别选通和锁存行地址和列地址。在行地址中，包括了存储地址和刷新地址，因此要有刷新多路器，该多路器让刷新地址在刷新期间通过。行地址 $A_0 \sim A_6$ 和刷新地址 $RA_0 \sim RA_6$ 通过刷新多路器送往行/列多路器，再送往 2116 的对应引脚。通过 RAS、CAS 及 R/W 信号的配合，实现对存储器的刷新与读写。在 8086 系统中，动态 RAM 的刷新管理是由刷新控制电路来完成的。对于 CPU 来说，动态 RAM 和静态 RAM 在接口电路上没有多大区别。为了保证刷新电路的刷新持续时间，当刷新过程和 CPU 的存储读写同时出现时，刷新控制会产生一个申请信息，以便 8086 在总线周期中插入等待状态。

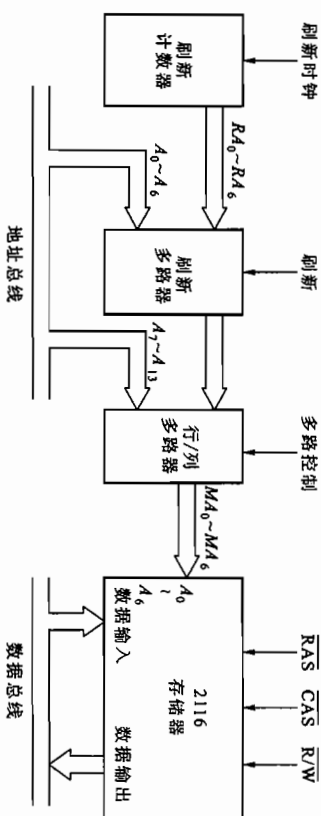


图 4-7 DRAM 芯片的内部结构框图

DRAM 刷新时是按行进行的，刷新时一次选中一行。当某一行选线被选中时，该行上所有存储电路中的电容器上的信息送到刷新放大器上，刷新放大器又对这些电容器立即进行重写。注意，刷新时，列选择信号总为“0”，因此电容器上的信息不会被送到数据总线上。

4.3.2 只读存储器

只读存储器 (ROM) 通常有掩模 ROM、一次可编程只读存储器 (PROM)、可擦除可编程只读存储器 (EPROM)、电可擦除可编程只读存储器 (EEPROM)、快闪存储器 (Flash Memory) 等几种类型。只读存储器的存储单元可以由二极管、晶体管或场效应晶体管构成。

1. 掩模 ROM

这种 ROM 是由制造厂家利用一种掩模技术写入程序的。掩模 ROM 制成后，用户不能修改，根据制造工艺可以将它们分为 MOS 型和双极型两种。MOS 型 ROM 功耗小、速度慢，适用于一般微型计算机系统；而双极型则速度快、功耗大，适用于速度较高的计算机系统。图 4-8 是掩模 ROM 的原理图。图中是简单的 4×4 位二极管 ROM，采用单译码结构， A_0, A_1 译码后可译出 4 种状态，输出 4 条选择线，可分别选中 4 个单元，每个单元有 4 位输出。图中所示的矩阵中，在行和列的交点，有的连有二极管，有的没有，这是制造厂家根据用户提供的数据对芯片（掩模）进行二次光刻所决定的，所以称为掩模 ROM。若地址线 A_0, A_1 为 00，即字线 0 为高电平，则选中 0 号

单元;若有二极管与其相连(如位线3、2和0),其相应的二极管导通,位线输出为1,而位线1没有管子与字线相连,输出为0,即0号单元的内容为1101。

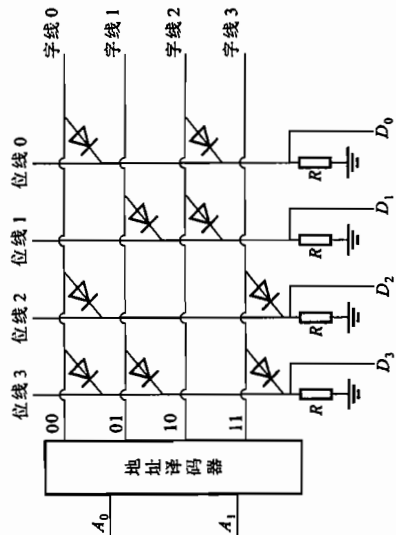


图 4-8 掩模 ROM 构成的原理图

2. 一次可编程只读存储器(PROM)

PROM 的总体结构与掩模 ROM 一样,同样由存储矩阵、地址译码器和输出电阻组成,不过在出厂时已经在存储矩阵的所有交叉点上全部制作了存储单元,就是在所有的存储单元都存储了 1。

图 4-9 是熔丝型 PROM 存储单元的原理图,它由一只晶体管和串联在发射极的快速熔丝组成。晶体管的 BE 结相当于接在字线和位线之间的二极管,熔丝多用低熔点合金制成。在写入数据时利用大电压产生大电流在需要写入 0 的单元将熔丝烧断就可以了。

3. 可擦除可编程只读存储器(EPROM)

EPROM 可反复使用多次,灵活、方便。目前多用叠栅注入型 MOS 管(称为 SIMOS 管)构成 EPROM 的存储单元。

EPROM 芯片封装时表面有一个石英玻璃透明窗口,如图 4-10 所示。用专门的设备(如紫外线擦除器)使芯片窗口受到紫外线照射时,所有电路中的浮动栅上的电荷获得能量会形成光电流泄漏走,使管子恢复初始状态,从而把原先写入的“0”信息擦去。经过照射后的 EPROM,还可以用专门的设备(EPROM 写入器)再次写入所需要的信息,然后用黑纸或黑胶布把小窗口贴上,以防紫外线把其中的内容擦掉。

要注意,编程后的芯片在阳光的影响和室内日光灯的照射下,经过 3 年时间浮动栅的电荷可泄漏完。若在太阳光直射

下,约一个星期电荷可泄漏完。所以,在正常使用和储藏时,应在芯片窗口上贴上黑色的保护纸。

4. 电可擦除可编程只读存储器(EEPROM)

EPROM 芯片的擦除需要将芯片取下,用紫外线照射十几分钟,而且是整片擦除,相对来说读写操作速度慢,不方便。随着电子技术的发展,又出现了一种称为电可擦除可编程只读存储器,简称 EEPROM 或 E²PROM,其存储信息的原理类似 EPROM,但擦除的原理不同。EEPROM 是通过对存储信息的 MOS 管的源极和漏极之间加一个较高的电压,释放浮栅上的电荷。它可以整片擦除,也可以擦除指定的单元。EEPROM 具有 EPROM 可重编程的特点,又具有擦除速度快、可按单元擦除的优点。

5. 快闪存储器(Flash Memory)

快闪存储器是在 EPROM 和 E²PROM 基础上发展起来的非易失性存储器,兼具两者的优点。快闪存储器的写入原理与 EPROM 相同,而擦除原理同 EEPROM。

快闪存储器既和 ROM 一样是一种非易失性存储器,又和 RAM 一样存储容量大;既和 EPROM 一样可以改写,又比它容易改写且价格便宜;既比硬盘存取速度快,又比它体积小、重量轻、功耗低和不易损坏。因此,快闪存储器发展速度快,应用范围宽。但是,它与 DRAM 相比,有两大缺点,一是寿命短,擦写次数仅 10 万次,而 DRAM 可达 1 015 万次;二是存取速度慢,一般为 45 ns~120 ns,而 DRAM 可达 10 ns~20 ns。所以,它无望取代 DRAM,不能成为主要的内部存储器。

快闪存储器的工作原理是利用“热电子”注入来完成写入过程,利用 Fowler-Nordheim 隧道效应来完成擦除,一次可擦除若干区域或整个存储器。

快闪存储器的工作方式有如下几种:

- 读工作方式:它的正常工作方式是只读方式。
- 写入工作方式:它的写入和擦除都要依靠电的方法来进行,因而要采用写命令的方法来管理各种写入功能。
- 擦除工作方式:它是依靠电的方法按区块或整片进行擦除操作。

快闪存储器的应用范围正在日益扩大。目前,快闪存储器主要用于微型计算机的 BIOS、MP3、手机、PDA、数码相机、各种类型的存储卡、U 盘等设备中。

4.3.3 内存的组成

内部存储器一般由地址译码器、存储体、控制逻辑组成,如图 4-11 所示。

1. 存储体

存储体(Memory Bank, MB)是能够寄存二进制信息的基本存储单元的集合。为了便于信息的读写,这些基本的存储单元应配置成一定的阵列,并进行编址,所以存储体也称为存储矩阵。

每个具有唯一地址的存储单元中,可存储一位或多位二进制数据。所以芯片的存储容量可以理解成芯片的存储单元数与每个单元存储位数之积。

2. 地址译码电路

地址译码电路的功能为接收 CPU 发出的地址信号,产生地址译码信号,以便选中存储矩阵

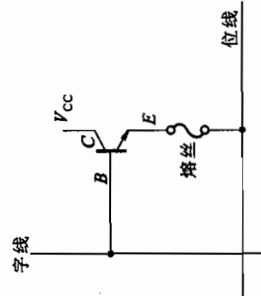


图 4-9 熔丝型 PROM 存储单元

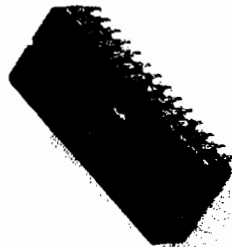


图 4-10 EPROM 存储器

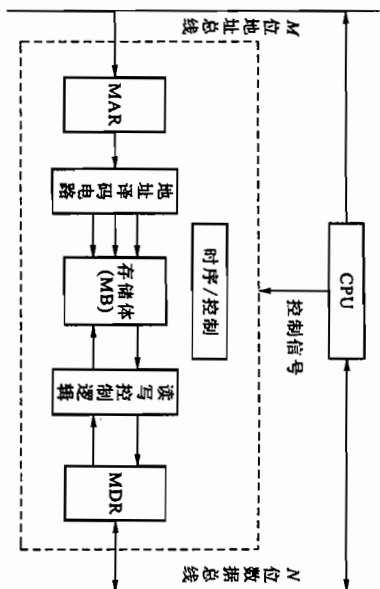


图 4-11 内存的基本组成

中的某个存储单元。存储矩阵中基本存储电路的编址方法有两种：单译码与双译码。

单译码方式较适用于小容量的字结构存储器，存储器中的存储单元呈线性排列，例如在图 4-12(a) 中，地址信号线 $A_5 \sim A_0$ 输入为 000101 时便选中了第 5 个存储单元。

双译码方式比较适用于大容量的存储器，如图 4-12(b) 所示。将地址线分为列线与行线两组，分别译码，当输入一个地址编码时，例如当输入 001010 时，行译码产生为 2，列译码产生为 1，则选中存储单元为第一列第二行的存储单元。

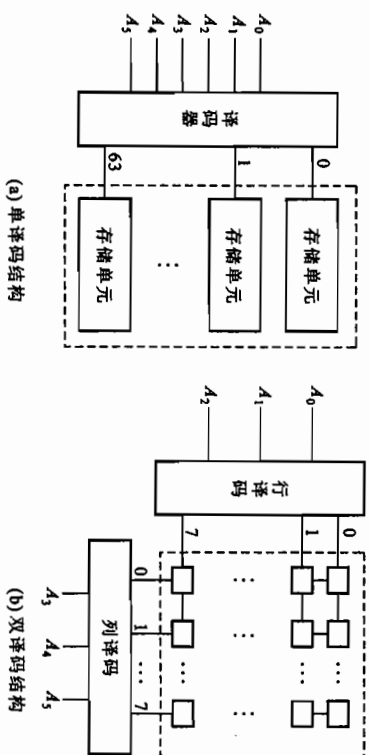


图 4-12 内存译码方式

3. 读写控制逻辑

读写控制逻辑通过发出片选信号、读信号、写信号来控制存储芯片的读写操作。存储芯片的片选一般用 CS 表示。有效时可对存储芯片进行读写操作，无效时芯片脱离总线。存储芯片的片选与系统高位相联系。一般用 OE(输出允许)和 WE(写允许)来表示。当芯片被选中时：

- OE 被用来控制读操作，有效时，允许芯片将寻址单元内的数据输出。该控制端一般与系统的读控制线 MEMR(或 RD)相连。

• WE 被用来控制写操作，有效时，引脚上的数据将被允许进入芯片，写入被寻址的单元。该控制端一般与系统的写控制线 MEMW(或 WR)相连。

近几年来内存技术一直在不断地发展，从最早使用的 DRAM 到后来的 FPM DRAM、EDO DRAM，到今天的 SDRAM 及最新的 RDRAM，出现了多种内存控制与访问技术。它们的共同特点是：如何缩短确定读写单元地址的时间和缩短输入输出数据的时间。

4.4 RAM 的基本工作方式

RAM(主要指动态 RAM)芯片设计得像一个个二进制的矩阵，每一个存储单元位于行与列的交叉点，并且具有一个行地址和一个列地址。芯片被读写前，主板上芯片组中的内存控制器要给出被读写的存储单元的地址。内存芯片内部的逻辑电路将地址转换成该单元的行、列值后，CPU 才能读写指定单元的数据。例如，一个标明为 70 ns 的内存芯片要用 70 ns 才能读或写该存储单元一个位的数据。

此外，还需要额外的时间从 CPU 得到地址信息，设置下一条指令。对系统而言，一个完整的内存读写周期，是从内存控制器发一个地址信息给内存芯片开始，到它再传送下一个地址信息给内存芯片为止这一段时间，一般需要用时 100 ns。对内存进行读操作主要经历如下过程：

- ① 选中地址。CPU 向内存控制器发出地址，内存控制器再向内存芯片传送这一地址，芯片内的逻辑电路将地址转换为存储单元的行地址和列地址。
- ② 把数据从选中的存储单元传送到内存芯片的输出电路。
- ③ 内存芯片输出数据。

存储器控制技术就是要减少其中的一步或多步所用的时间，以减少总的读写时间。

由于目前主存主要采用 DRAM 存储器，但 DRAM 的读取速度远低于 CPU 的运行速度，现今解决 CPU 与主存之间的速度匹配的主要方法是在 CPU 与 DRAM 间加上用读取速度更快的 SRAM 做成的二级高速缓存，这种内存系统可以承担系统 85% 的内存请求，而不需 CPU 增加额外的等待周期。

DRAM 需要恒定电流以保存信息，一旦断电，信息即丢失。它的刷新频率可达每秒几百次，但由于 DRAM 使用同一电路来存取数据，所以 DRAM 的存取时间有一定的时间间隔，这导致了它的存取速度并不是很快。另外，在 DRAM 中，由于存储地址空间是按页排列的，所以当访问某一页面时，切换到另一页面会占用 CPU 额外的时钟周期。

为了提高 DRAM 的读写速度，人们发明了一系列先进的技术，这些技术虽不能使 DRAM 与

CPU 同步,但对内存的读写速度已经有了很大的提高。

4.5 内存模组与基本结构

我们通常所见到的内存都是以模组的形式出现的,即由若干数量的内存颗粒封装成一根内存条。为什么要做成这种形式呢?这里需要了解两个基本概念:物理存储体与芯片位宽。

4.5.1 物理存储体

传统内存系统为了保证 CPU 的正常工作,必须一次传输完 CPU 在一个传输周期内所需要的数据。而 CPU 在一个传输周期能接收的数据容量就是 CPU 数据总线的位宽,单位是“位”。内存与 CPU 之间的数据交换通过主板上的北桥芯片进行,内存总线的数据位宽等同于 CPU 数据总线的位宽,这个位宽称为物理存储体(Physical Bank, P-Bank)的位宽。以目前主流的 DDR 系统为例, CPU 与内存之间的接口位宽是 64 位,也就意味着 CPU 在一个周期内会向内存发送或从内存读取 64 位的数据,这 64 位的数据集合就是一个内存条存储体。

P-Bank 其实就是一组内存芯片的集合,这个集合的容量不限,但这个集合的总位宽必须与 CPU 数据位宽相符。随着计算机应用的发展,一个系统只有一个 P-Bank 已经不能满足容量的需要。所以,芯片组开始可以支持多个 P-Bank,一次选择一个 P-Bank 工作,这就有了芯片组支持多少物理存储体的说法。而在 Intel 的定义中,则称 P-Bank 为行(Row),比如 845G 芯片组支持 4 个行,也就是说它支持 4 个 P-Bank。

一条内存条的物理存储体是由所采用的内存芯片的位宽决定的,各个芯片位宽之和为 64 位就是单物理存储体;如果是 128 位就是双物理存储体。因此,可以通过两种方式增加这种类型内存的容量。第一种就是通过增加每一个独立模块的容量来增加存储体的容量,第二种方法就是增加存储体的数目。由于目前内存芯片位宽的限制,一个系统只有一个物理存储体已经不能满足容量的需要。目前新一代芯片组可以支持多个物理存储体,最少的也能支持 4 个物理存储体。对于像 Intel 845D 这种支持 4 个存储体的芯片组来说,在选购内存时就要考虑一下插槽数与内存存储体的分配问题了。因为如果选购双存储体的内存,这意味着在 Intel 845D 芯片组上最多只能使用两条这样的内存,超出的话芯片组将无法识别。最好根据主板所提供的内存插槽数目来选购内存。如果主板只提供了两个内存插槽,那就不必为内存是单存储体还是双存储体而担心了;如果主板提供了 4 个内存插槽(同一种规格),那么应该尽量购买单存储体或大容量双存储体的内存,以免给日后升级留下不必要的麻烦。

注意:SDRAM 与 DDR 内存的物理存储体是一样的,不过在 RDRAM 内存规格中,物理存储体为通道(Channel)所取代。

4.5.2 逻辑存储体

逻辑存储体的英文全称为 Logical Bank,简称 L-Bank。如果将物理存储体说成是内存芯片阵列的话,那么逻辑存储体可以看作是数据存储阵列。

简单地说,SDRAM 的内部是一个存储阵列(图 4-13),阵列就如同表格一样,可将数据

“填”进去。因此逻辑存储体可以看成是一张二维表,内存中的数据以位为单位存储在每一个单元中。每个单元称为存储单元。只要指定一个行(Row),再指定一个列(Column),就可以准确地定位到某个存储单元。每个单元都可以存储数据,而且每个单元的存储空间相同,实际上与物理存储体的每个单元存储的数据量相同。这个具体的单元存储数据量即为逻辑存储体的位宽(实际上内存芯片的位宽就是逻辑存储体的位宽)。由于工艺上的原因,这个阵列不可能做得太大,所以一般内存芯片中都是将内存容量分成几个阵列来制造,也就是说内存芯片中存在多个逻辑存储体,随着芯片容量的不断增加,逻辑存储体数量也不断增加。较早以前是两个存储体,目前基本上都是 4 个存储体,这也是 SDRAM 规范中最高逻辑存储体数量。在 RDRAM 中,逻辑存储体最多可达 32 个。在最新的 DDR-II 标准中,逻辑存储体的数量也提高到了 8 个。

列地址 (Column)

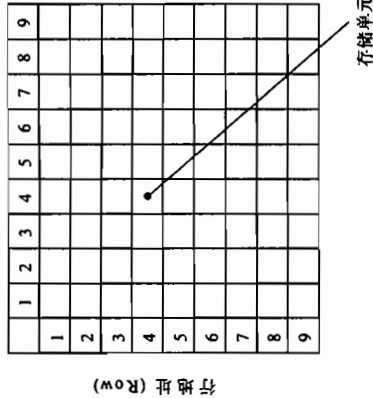


图 4-13 L-Bank 存储阵列示意图

这样,在进行寻址时就要先确定是哪个逻辑存储体,然后再在这个选定的逻辑存储体中选择相应的行与列进行寻址。可见对内存的访问,一次只能有一个逻辑存储体工作,而每次与北桥交换的数据就是逻辑存储体存储阵列中一个“存储单元”的容量。在某些厂商的表述中,将逻辑存储体中的存储单元称为 Word(此处代表位的集合而不是字节的集合)。

主板芯片组本身在设计时,一个时钟周期内只允许对一个逻辑存储体进行操作,而不是同时对所有逻辑存储体进行操作。逻辑存储体的地址线是通用的,只要有一个逻辑存储体编号加以区别就可以了(Bank0 ~ Bank3)。但是这个芯片的位宽决定了一次能从它那里读出多少数据。

对于 DDR 内存,逻辑存储体的作用、原理与在 SDRAM 中是一样的,主要区别是逻辑的容量、规格有所不同。SDRAM 中逻辑存储体存储单元的容量与芯片位宽相同,DDR 的逻辑存储单元的容量是芯片位宽的两倍:芯片位宽 $\times 2 =$ 存储单元容量。另外,DDR 中的真正行、列地址数也与同规格 SDRAM 不一样。这主要是由 DDR 的工作原理所决定的。DDR 内存的存储单元容量设计成两位预取(2-bit Prefetch),也称为 $2-n$ 预取(n 代表芯片位宽)。另外,当系统运行在双通道模式下时,由于两个通道内的相同物理存储体的内存芯片在同一时间的工作行地址相同,所以此时可同时打开的页面数量减半,但页面的容量翻倍。由于页面数量的多少与容量关系到内存控制器能力的发挥,所以关键就在于控制器怎么协调好页面数量与页面容量之间的关系。

4.5.3 内存条(模组)的结构及工作原理

SDRAM 是一种具有同步接口的动态随机存储器,目前微型计算机的主流 DDR 存储器就是由其演变而来。本节以 128 Mb (32 Mb \times 4) SDRAM 为例介绍内存的组成及工作原理,图 4-14 是其内部结构示意图。该存储器的 I/O 信号包括:

$A_0 \sim A_{11}$: 地址线 12 根(寻址空间 $2^{12} = 4096$)

RAS#: Row Address Strobe, 行地址选通脉冲

CAS#: Column Address Strobe, 列地址选通脉冲
CS#: 片选信号
WE#: 写允许
CKE: 时钟有效信号
CLK: 时钟信号

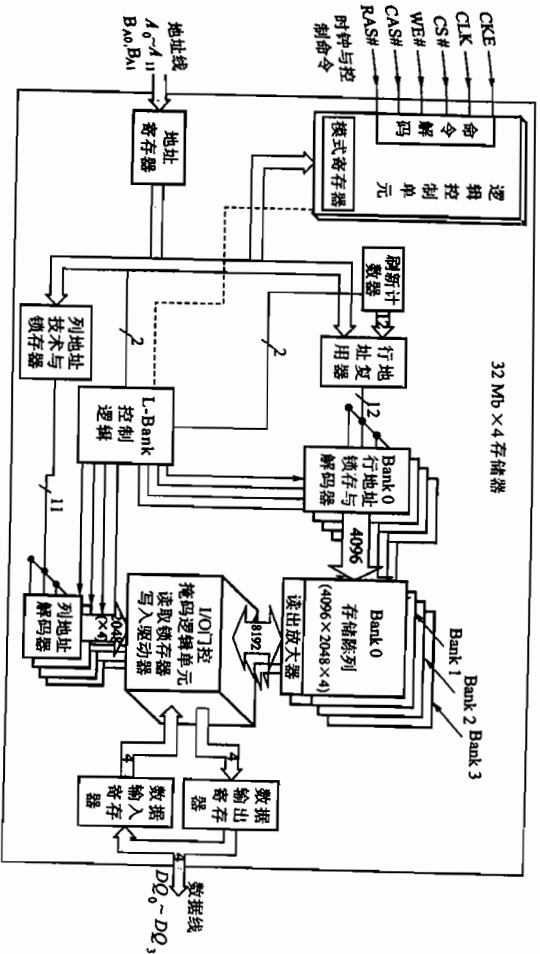


图 4-14 128 Mb (32 Mb x 4) SDRAM 内部结构图

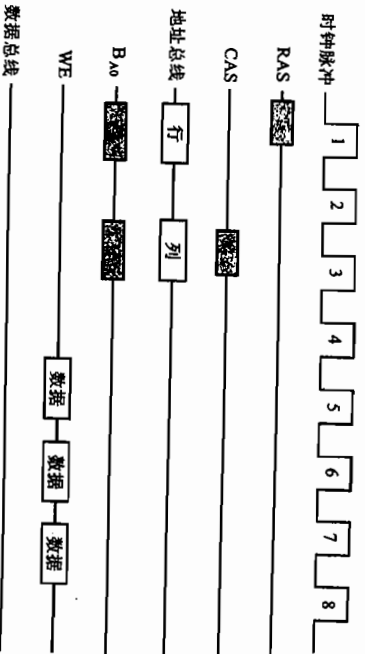


图 4-15 SDRAM 读取数据的时序图

图 4-15 是 SDRAM 读取数据的时序图, 基本工作过程如下。

(1) 芯片初始化

在 SDRAM 芯片内部有一个逻辑控制单元, 并且有一个模式寄存器为其提供控制参数。因此, 每次开机时 SDRAM 都要先对这个控制逻辑核心进行初始化。关键的阶段就在于模式寄存器 (Mode Register, MR) 的设置, 简称 MRS (MR Set), 这一工作由北桥芯片在 BIOS 的控制下进行, 寄存器的信息由地址线来提供。

(2) 行有效

初始化完成后, 要想对一个逻辑存储体 (L-Bank) 中的阵列进行寻址, 首先就要确定行 (Row), 使之处于活动状态 (Active), 然后再确定列。虽然之前要进行片选和 L-Bank 的定址, 但它们与行有效可以同时进行。

在 CS#, L-Bank 定址的同时, RAS (Row Address Strobe, 行地址选通脉冲) 也处于有效状态。此时 A₀ 地址线则发送具体的行地址。如图中是 A₀ ~ A₁₁, 共有 12 根地址线, 由于是二进制表示法, 所以共有 4 096 行 (2¹² = 4 096)。A₀ ~ A₁₁ 的不同数值就确定了具体的行地址。由于行有效的同时也是相应 L-Bank 有效, 所以行有效也可称为 L-Bank 有效。

(3) 列读写

行地址确定之后, 就要对列进行寻址了。但是, 地址线仍然是行地址所用的 A₀ ~ A₁₁ (本例)。读/写命令是通过对芯片的可写状态的控制来达到读写目的的。WE# 信号有效时即为写命令。WE# 无效时就是读取命令。

列寻址信号与读写命令是同时发出的。虽然地址线与行寻址共用, 但 CAS (Column Address Strobe, 列地址选通脉冲) 信号可以区分开行与列寻址的不同, 配合 A₀ ~ A₉, A₁₁ 来确定具体的列地址。

读取命令与列地址一块发出 (当 WE# 为低电平时即为写命令)。

在发送列读写命令时必须要有行有效命令有一个间隔, 这个间隔被定义为 t_{RCD}, 代表 RAS to CAS Delay (RAS 至 CAS 延迟)。大家也可以理解为行选通周期, 这应该是根据芯片存储阵列电子元件响应时间 (从一种状态到另一种状态变化的过程) 所制定的延迟。t_{RCD} 是 SDRAM 的一个重要时序参数, 可以通过主板 BIOS 经过北桥芯片进行调整, 但不能超过规定的范围。广义的 t_{RCD} 以时钟周期 (t_{CK}, Clock Time) 数为单位, 比如 t_{RCD} = 2, 就代表延迟为两个时钟周期。具体到确切的时间, 则要根据时钟频率而定。对于 PC100 SDRAM, t_{RCD} = 2, 代表 20 ns 的延迟, 对于 PC133 则为 15 ns。

(4) 数据输出 (读)

在选定列地址后, 就已经确定了具体的存储单元。剩下的事情就是数据通过数据 I/O 通道 (DQ) 输出到内存总线上了。但是在 CAS 发出之后, 仍要经过一定的时间才能有数据输出, 从 CAS 与读取命令发出到第一批数据输出的这段时间, 定义为 CAS 潜伏期 (CAS Latency, CL)。由于 CL 只在读取时出现, 所以 CL 又被称为读取潜伏期 (Read Latency, RL)。CL 的单位与 t_{RCD} 一样, 为时钟周期数, 具体耗时由时钟频率决定。

(5) 数据输入 (写)

数据写入的操作也是在 t_{RCD} 之后进行, 但此时没有了 CL (记住, CL 只出现在读取操作

中),行寻址与列寻址的时序图和读时序图一样,只是在列寻址时,WE#为有效状态。

(6) RAS 和 CAS 引脚停止激活,等待下一个读(写)命令

在内存的读取过程中,内存不可能在运行完一个读取操作之后就立刻进行第二个读取操作,因为 DRAM 的读取操作包括电容器的充电和放电以及把信号传出去的时间,所以在两个读取操作中间要留出足够的时间,让内存进行这些方面的操作,经 DRAM 读操作的延迟。

(7) SDRAM 刷新

之所以称为 DRAM,就是因为它要不断进行刷新(Refresh)才能保留住数据,因此刷新是 DRAM 最重要的操作。

那么隔多长时间需要刷新一次呢?目前公认的标准是,存储体中电容的数据有效保存期限是 64 ms,也就是说每一行刷新的循环周期是 64 ms。这样刷新速度就是:行数量/64 ms。内存规格中,经常会看到 4096 Refresh Cycles/64ms 或 8192 Refresh Cycles/64ms 的标识,这里的 4096 与 8192 代表这个芯片中每个逻辑存储体的行数。刷新命令一次对一行有效,发送间隔也是随总行数而变化,4 096 行时间间隔为 15.625 μ s,8 192 行时间间隔为 7.8125 μ s。

刷新操作分为两种:自动刷新(Auto Refresh,AR)与自刷新(Self Refresh,SR)。不论是何种刷新方式,都不需要外部提供行地址信息,因为这是一个内部的自动操作。对于自动刷新,SDRAM 内部有一个行地址生成器(也称刷新计数器),用来自动依次生成行地址。由于刷新是针对一行中的所有存储体进行,所以无需列寻址,或者说 CAS 在 RAS 之前有效。所以,自动刷新又称 CBR(CAS Before RAS,列提前于行定位)式刷新。由于刷新涉及所有逻辑存储体,因此在刷新过程中,所有逻辑存储体都停止工作,而每次刷新所占用的时间为 9 个时钟周期(PCI33 标准),之后进入正常的工作状态,也就是说在这 9 个时钟期间内,所有工作指令只能等待而无法执行。64 ms 之后则再次对同一行进行刷新,如此周而复始进行循环刷新。显然,刷新操作肯定会对 SDRAM 的性能造成影响,也是 DRAM 相对于 SRAM(无需刷新仍能保留数据)取得成本优势的同时所付出的代价。

自刷新则主要用于休眠模式低功耗状态下的数据保存,其最著名的应用就是 STR(Suspend to RAM,挂起内存)。在发出 AR 命令时,将 CKE 置于无效状态,就进入了 SR 模式,此时不再依靠系统时钟工作,而是根据内部的时钟进行刷新操作。在 SR 期间除了 CKE 之外的所有外部信号都是无效的(无需外部提供刷新指令),只有重新使 CKE 有效才能退出自刷新模式并进入正常操作状态。

4.6 主流内存条介绍

4.6.1 FPM DRAM

FPM DRAM(Fast Page Mode DRAM,快速页模式动态随机存储器)之所以能提高速度是基于这样一个事实——计算机中大量的数据是连续存放的。比如,若一个数据与前一个数据的行地址相同,内存控制器就不必再传一个行地址,只要再传一个列地址就可以了。当 CPU 选中某单元时,有理由认为下一个地址就是 CPU 要选的。FPM DRAM 的这种控制方式使 CPU 能用较少

的时钟周期读取较多的数据。这样存取同一“页”数据的速度与效率就大大提高了(行地址不变时,列地址可寻址的空间称为“页”;一页通常为 1 024 B 的整数倍)。显然,FPMD RAM 的这种工作方式需要内存芯片和内存控制器共同配合才能完成。在 586 级 PC 的主板上,内存控制器的被制作在主板的芯片组中。随着技术的成熟,FPM DRAM 的访问时间也在不断的缩短,从 120 ns 缩短到如今通用的 60 ns。60 ns 的 FPM DRAM 可用于总线速度(或外频)为 66 MHz 的 Pentium 系统(CPU 主频为 100 MHz、133 MHz、166 MHz 和 200 MHz)中。

FPM DRAM 的数据存取过程是:在 DRAM 阵列中,首先提供一个行地址,并置 RAS 为低电平,之后通过周期性的 CAS 实现多个列的存取。每一个 CAS 周期都提供一个列地址,并置 CAS 为低电平,等待合法的数据输出。之后系统锁住数据,再置 CAS 为高电平,并按照同样的顺序开始下一个周期。

注意:当一个合法的数据被系统锁住后,置 CAS 为高电平,数据不能输出。

FPM DRAM 内存条主要采用 72 线的 SIMM(Single Inline Memory Module,单列直插存储模块)封装如图 4-16 所示,其存取时间一般在 60 ns~100 ns 左右。

以上这种快速页模式的内存常用于视频卡,通常也叫做 DRAM。其中一种经过特殊设计的内存的存取时间仅为 48 ns,这种内存就叫做视频随机存储器(Video RAM,VRAM)。这种经过特殊设计的内存具有“双端口”,其中一个端口可直接被 CPU 存取,而另一个端口可独立地被 RAM 的直接存储器存取(Direct Memory Access,DMA)通道存取,这样存储器的 DMA 通道可以不必等待 CPU 完成存取就可同时工作,从而比一般的 DRAM 要快些。



图 4-16 72 线的 SIMM 内存条

4.6.2 EDO DRAM

EDO DRAM(Extended Data Output DRAM,扩展数据输出动态随机存储器)是在 FPM DRAM 基础上加以改进的存储器控制技术。它可以在输出一个数据的过程中,就准备下一个数据的输出。EDO DRAM 采用一种特殊的内存读出控制逻辑,在读写一个地址单元时,同时启动下一个(连续)地址单元的读写周期,从而节省了重选地址的时间,使存储总线的速率提高到 40 MHz。与 FPM DRAM 相比,由于增大了输出数据所占的时间比例,性能提高了将近 15%~30%,而其制造成本与 FPM DRAM 相近。EDO DRAM 的存取时间约为 50 ns~70 ns。

EDO 模式的存取有些像 FPM 模式,但有一点不同。当 CAS 为高电平时,并不阻止数据的输出,此时锁存器(latch)维持住合法的数据之后,CAS 又回到低电平。在 FPM 模式存取中,锁存器由 CAS 控制,在 CAS 变成高电平时,数据被锁住,可有效地获得数据;在提供一个新的地址后,内存阵列中新的数据被定位存取,此时并不影响上一次存取的数据输出。当连续存取数据时,因交错关系,EDO DRAM 能使内存控制器省去用来发出下一列地址信息的 10 ns 的等待周期,这就能提供比 FPM DRAM 更快的数据流,从而使计算机的效率提高 10%~15%。

FPM DRAM 和 EDO DRAM 的芯片制作技术其实是相同的,不同的是 EDO DRAM 所增加的机制必须在芯片组的支持下将发送的数据信号的处理时间缩短,以加快系统的整体执行效率。目前,EDO DRAM 内存条主要采用 72 线的 SIMM 形式封装,如图 4-17 所示,也有少部分采用

168 线的 DIMM (Double Inline Memory Module, 双列直插存储器模块) 封装。工作电压为 5 V。EDO DRAM 除了作主存之外, 还用于许多显示卡作为显存使用。



图 4-17 采用 72 线的 EDO DRAM 内存条

4.6.3 SDRAM

前面介绍的几种 DRAM 内存都属于非同步存储的存储器, 即它们的工作速度并没有和系统时钟同步, 它们的速度一般不能超过 66 MHz。

随着主流 CPU 的主频不断的提高, CPU 的外频已经远远超过了 66 MHz, 因此, 要求内存能适应 66 MHz 以上的总线速度, 这就出现了 SDRAM (Synchronous DRAM, 同步动态随机存储器)。SDRAM 与上述内存存取模式不同, 它可以与系统时钟同步工作。

SDRAM 支持高速总线时钟频率 (66 MHz 以上), 而不必插入指令等待周期, 它的基本原理是将 CPU 和 RAM 通过一个相同的时钟锁在一起, 使得 RAM 和 CPU 能够共享一个时钟周期, 以相同的速度同步工作。理论上其速度可与 CPU 的外频同步, 与 CPU 共享一个时钟周期。目前 SDRAM 能够搭配运行的时钟频率可以高达 66 MHz ~ 133 MHz, 存取时间只需 7 ns ~ 15 ns, 大大提高了 Pentium 系统的性能。以 Intel 430VX 芯片支持的 SDRAM 为例, 在 66.6 MHz 的内存数据总线速度下, 达到 7-1-1-1 的时序水准, 即第一次存取需要 7 个时钟周期, 但接下来的存取都能在 1 个时钟周期内完成 (如果总线速度更高, 则第一次存取所需的总线周期会更小), 就是说, SDRAM 在开始的时候要花费一些时间, 但在以后, 每一个时钟周期可以读写一个字节的数, 做到了所有的输入/输出信号与系统时钟同步。这已接近主板上的同步二级高速缓存 (流水线总线缓存) 的 3-1-1-1 水准。一般来说, 在系统时钟为 66 MHz 时, SDRAM 与 EDO DRAM 相比, 显示不出其优点, 但当系统时钟增加到 100 MHz 以上, SDRAM 的优点便很明显。

SDRAM 与系统时钟同步, 采用流水线处理方式, 当指定一个特定的地址, SDRAM 就可读出多个数据, 即实现数据的突发传送。具体来说, 第一步, 指定地址; 第二步, 把数据从存储地址传到输出电路; 第三步, 输出数据到数据总线。关键的是以上三个步骤是各自独立进行的, 且与 CPU 同步, 而以往的内存只有从头到尾执行完这三个步骤才能输出数据。这就是 SDRAM 高速的原因。SDRAM 的读写周期一般为 7 ns ~ 15 ns。

SDRAM 采用新的双存储体结构, 内含两个交错的存储阵列, 允许两个内存页面同时打开, 当 CPU 从一个存储体或阵列访问数据的同时, 在内存控制器作用下另一个存储体或阵列已准备好读写数据。通过两个存储阵列的紧密切换, 读取效率得到成倍提高。

现在已有 100 MHz 和 133 MHz 的 SDRAM 可以使整机性能大约提高 5% ~ 10%。由于 SDRAM 是双体结构, 原则上可用于所有双端口方式的地方。比如可代替显示卡用的双端口视频内存 (VRAM) 来提高带宽。为降低功耗, SDRAM 一般使用 3.3 V 电压, 部分产品 3.3 V 和 5 V

均可。

SDRAM 以其优良的特性, 成为当前内存市场的主流产品, 受到普遍的欢迎, 并且为各种流行的芯片组所支持。SDRAM 可以使所有的输入输出信号保持与系统时钟同步。而在以前, 这只有 SRAM (静态随机存储器) 才能办到。

SDRAM 不仅可用作主存, 在显示卡专用内存方面也有广泛应用。对显示卡来说, 数据带宽越宽, 同时处理的数据就越多, 显示的信息就越多, 显示速度和显示质量也就越高。

由于视频内存 (VRAM) 成本较高, 其应用受到很大限制。因此在一般显示卡上, 使用廉价的 SDRAM 或 EDO DRAM 比较普遍。由于早期的图形加速卡的 64 位带宽 (指内部总线带宽) 已扩大到 EDO RAM 所能达到的带宽的极限, 要达到更高的带宽和更高的分辨率 (如 1600 × 1200 像素以上), 而又尽量降低成本, 就只能采用频率在 66 MHz 以上、高带宽 SDRAM 了。

SDRAM 也可以应用于统一内存结构 (Unified Memory Architecture, UMA), 这是一种使用主存代替显示内存的结构。这种结构在很大程度上降低了系统成本。许多高性能显示卡价格高昂, 就是因为其专用显示内存成本极高, 而 UMA 技术利用主存代替显示内存, 不再需要增加专门的显示内存, 因而降低了系统成本。

目前 SDRAM 普遍采用 168 线的 DIMM 封装如图 4-18 所示, 同时需要主板芯片组的支持, 而且使用时要将工作电压调为 3.3 V, 这就需要主板具有支持 3.3 V DIMM 内存条插槽的功能。



图 4-18 168 线的 DIMM 封装的 SDRAM 内存条

4.6.4 DDR SDRAM 内存

DDR SDRAM (Double Data Rate SDRAM, 双数据速率同步动态随机存储器) 也可以说是传统 SDRAM 的升级版。VIA 公司主推的 DDR SDRAM, 目前标准为 PC266, 它的带宽为 2.12 Gbps。顾名思义, DDR 通过采用将一个时钟周期的前后沿都用于触发数据传输的方式, 将峰值数据传输率提高了一倍, 是 SDRAM 的两倍, 如图 4-19 所示。由于仅多采用了下降沿信号, 并不会造成能耗的增加。此外, 为了取得较高的速率, DDR 不再像 SDRAM 一样支持电压 3.3 V 的 LVTTL 标准, 而改为支持电压为 2.5 V 的 SSTL2 信号标准。就 DDR SDRAM 架构来看, 基本上可完全沿用 SDRAM 现有的生产体系。就生产工艺来说, DDR SDRAM 可与 SDR 共享同一个 DRAM 设计, SDRAM 生产通常需要 20 ~ 22 道工序, 生产 DDR SDRAM 大部分的工序与制程可完全相同, 仅需在最上一道金属层或在垫底 (Pad) 上做些更动便可生产出 DDR SDRAM 产品, 芯粒面积与 SDRAM 并无差别。在封装上, 仍是用传统的 TSOP II 封装 (但为 66 引脚), 测试则仍可沿用现有的 SDRAM 测试平台。模组部分采用 184 引脚, 印刷电路板 (PCB) 使用 4 ~ 6 层板, 其他零组件或组装上并无太大差别, 与 SDRAM DIMM 成本相差不大, DDR SDRAM 的物理大小和标准的 DIMM

一样:同样的长度、同样的引脚距离、同样的PCB厚度等。区别在于引脚数。标准的SDRAM有两个小缺口,168个引脚。而DDR DIMM多出16个引脚,共184个引脚,占用了空间,只有1个小缺口,如图4-20所示。多出的引脚包含新的时钟、电源和接地等信号,如图4-20所示。DDR除了带宽比SDRAM提高一倍外,最大的优势来自于可完全使用现有的SDRAM生产体系。DDR因其设计可直接改良目前的SDRAM架构,在价格上较有优势。更重要的是,由于DDR大部分的制作流程均与目前相同,所需要总生产成本仅比目前的SDRAM高出约3%。

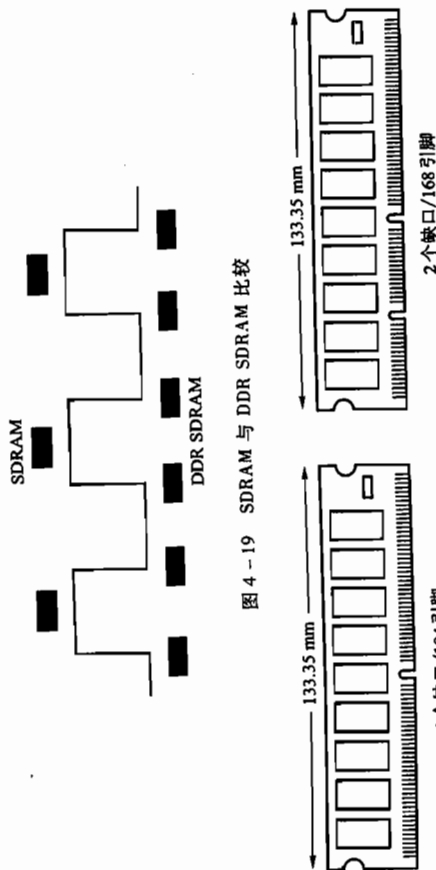


图 4-19 SDRAM 与 DDR SDRAM 比较

(a) PC 2100 DDR DIMM
(b) 传统的 PCI33 SDR DIMM

图 4-20 SDRAM 与 DDR 外部规格比较

4.6.5 DDR2 SDRAM 内存

DDR2(Double Data Rate 2) SDRAM 是由电子设备工程联合会(JEDEC)开发的新一代内存技术标准。与上一代DDR内存不同的是,虽然同是采用了在时钟的上升沿和下降沿同时进行数据传输的基本方式,但DDR2内存却拥有两倍于上一代DDR内存预取能力(即4位数据预取)。也就是说,DDR2内存每个时钟周期能够以4倍于外部总线的速率读/写数据,并且能够以内部控制总线4倍的速率运行(如图4-21所示)。

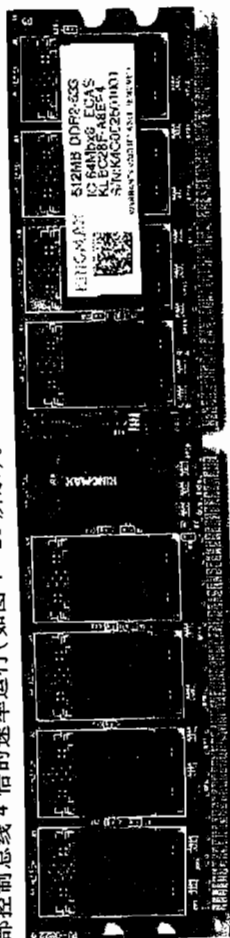


图 4-21 最新 DDR2 533 内存

1. 为何要 DDR2

从“速率=位宽×频率”的内存性能计算公式来看,提高内存性能有两种方式,增加内存总线的位宽或者是提高内存工作的频率。但问题是,由于发热量以及设计等方面的限制,导致存储单元无法提高频率,内存总线位宽也不能轻易增加。

尽管DDR400中存储阵列工作频率是200 MHz,目前最快的DDR SDRAM的频率(这里不包括那些超频的内存)达到了550 MHz,它的内部存储阵列工作频率达到275 MHz,这个频率已经很难再进一步提高。这就需要一个新的内存机制保证内存频率和性能可以稳定的提高,那就是DDR2内存。

DDR采用了2位预取,也就是2:1的数据预取。2位预取架构允许内部的队列工作频率仅为外部数据传输频率的一半。在SDRAM中,数据传输频率和时钟频率一样。DDR2采用了4位预取,这就是DDR2提高数据传输频率的关键,可以在不提高内部存储阵列频率的情况下提高数据传输带宽(如图4-22所示)。

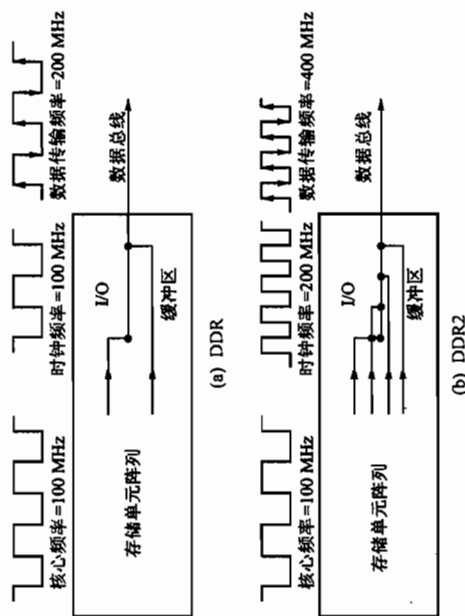


图 4-22 DDR 与 DDR2 工作原理

2. DDR2 性能优越

DDR2内存所带来的性能提升是明显的。在同等核心频率下,DDR2内存实际工作频率是DDR内存的两倍。这得益于DDR2内存拥有两倍于标准DDR内存的4位预取能力。也就是说,虽然DDR2与DDR一样,都采用了在时钟的上升沿和下降沿同时进行数据传输的基本方式,但DDR2拥有两倍于DDR的预取系统命令数据的能力。在同样100 MHz的工作频率下,DDR的实际频率为200 MHz,而DDR2则可以达到400 MHz。

另外,在同等工作频率下的DDR和DDR2内存中,DDR2内存的延时要高于DDR内存,比如说,DDR 200和DDR2-400具有相同的延时,而DDR2-400具有高一倍的带宽。实际上,DDR2-400和DDR 400具有相同的带宽,都是3.2 Gbps,但是DDR 400的核心工作频率是200

MHz, 而 DDR2 - 400 的核心工作频率是 100 MHz, 也就是说 DDR2 - 400 的延迟要高于 DDR 400。

3. 功耗更低的 DDR2

在内存技术中, 散热一直是急需解决的问题。目前的 DDR 内存并不需要太大的容量, 散热问题不太突出, 但对于未来的 DDR2 内存, 散热就不可忽视。例如, 在内存插槽中安装 4 GB 的内存时, 在峰值调用下内存的功耗将在 35 W ~ 40 W 之间。虽然这样容量的内存很少见, 但在 DDR2 平台中, 4 GB 容量已不是很稀奇的事。因此, 需要预先解决散热问题。而且内存的工作频率将会得到不断的提高, 就更需要尽快解决散热的问题(发热量和工作频率往往是同步增长的)。减少发热量的最好的方法就是采用新的技术。

4. DDR2 采用了更新的技术

与普通 DDR 内存相比, DDR2 内存采用了更新的技术, 其中最主要的是 OCD (On-Chip Driver)、ODT (On Die Terminator) 和 Post CAS。OCD 被称为离线驱动调整。通过 OCD 技术可以提高信号的完整性。DDR2 通过调整上拉 (pull-up) 和下拉 (pull-down) 的电阻值使两者电压相等, 使用 OCD 通过减少 DQ - DQS 的倾斜来提高信号的完整性; 通过控制电压来提高信号品质。在 I/O 驱动器上新增稳压线路, 令充电、放电动作的电压值误差减至最少, 以防止电压不稳定的时候引起数据丢失。

ODT 是内置核心的终结电阻器。DDR2 可以根据自己的特点内置合适的终结电阻, 这样可以保证最佳的信号波形。使用 DDR2 不但可以降低主板成本, 还得到了最佳的信号品质。

DDR2 通过引入 Post CAS 功能来解决指令冲突问题。Post CAS 是为了提高 DDR 2 内存的利用效率而设定的。Post CAS 是指将 CAS (读写命令) 提前几个周期, 直接插到 RAS 信号后面的一个时钟周期。

4.6.6 Rambus 内存

Rambus 公司于 1990 年 3 月成立, 不久就诞生了 Rambus 的核心专利——Rambus 信令电平 (Rambus Signaling Level, RSL) 技术。Rambus 内存最早于 1995 年 12 月推出, 当时它与任天堂 64 (Nintendo 64) 游戏机一起发售, 但名声不大。从 1996 年 12 月开始, Rambus 与 Intel 合作开发, 准备将 Rambus 推广到 PC 领域。Rambus 内存真正亮相于 PC 市场是在 1999 年 11 月。

1. RDRAM 简介

RDRAM 与 DDR SDRAM 一样, 也是一种采用双沿触发技术的内存, 但它在结构、控制体系方面相对于传统 SDRAM 有着不小的变化。首先我们来看看它与 SDRAM 之间的简单比较。

现在大多数 DDR 内存都使用 8 ~ 16 个内存芯片, 如图 4-23 所示。图 4-23 (a) 所示是一条 8 芯片的 SDRAM 内存条。8 个芯片都采用并行总线与 64 位总线连接。

SDRAM 内存的寻址信号和数据信号是完全区分开的。要注意, 图 4-23 仅仅是简化的示意图, 实际原理要复杂得多。最关键的是 64 位的数据总线被划分成了 8 个并行数据线, 用来连接这些内存芯片。

图 4-23 (b) 所示是一条 8 芯片的 RDRAM 内存条, Rambus 传输线穿过两条 (或 4 条) RIMM, 然后分别在两端终止。所有的 Rambus 芯片都是串联在传输线上的, 每个芯片都是 16 位的, 所以 RDRAM 通道只需要一个芯片就可以工作了。

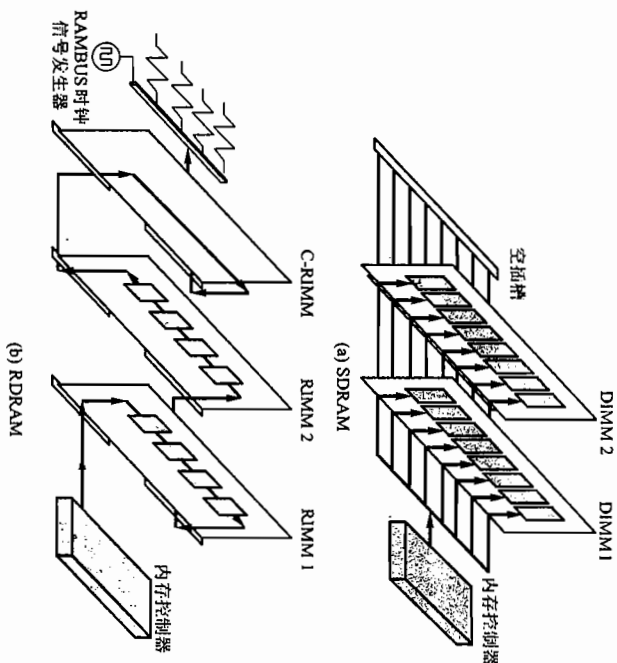


图 4-23 RDRAM 与传统 SDRAM 的比较

而 DDR 内存采用 8 个 8 位的并行数据线。第一条内存上的第一个内存芯片和第二条内存上的第一个内存芯片连接, 其余的依此类推。必须要有 4 ~ 16 个内存芯片, 内存条才能工作。

因此, RDRAM 虽然只有一条 16 位的数据线, 但是这一条数据线却可以连接全部的内存芯片。从图 4-23 中可以看出, RDRAM 在工作方式上与 SDRAM 有很大不同。SDRAM 需要多颗芯片并组成物理存储体与北桥沟通, 而在 RDRAM 架构中每个芯片就是一个可以单独工作的读写单元, 芯片的位宽就是与北桥接口的位宽, 所以如果用 ECC, 就要用专门的 ECC 型芯片, 也因此有了 16 位和 18 位两种规格。而芯片的位宽就是一个 RDRAM 通道的位宽 (本例以 16 位芯片进行介绍)。

为了达到更高的容量, 在一个通道中将多颗 RDRAM 芯片串联起来, 形成 Rambus 接口内存模组 (Rambus Interface Memory Module, RIMM)。如果主板允许, 完全可以设计一个超长的内存插槽与模组, 但现实中肯定不能这么做, 所以在主板上 Rambus 又把模组串联起来组成通道。由于是串联的形式, 所以要求起始端与终端形成一个完整的通路, 而 RIMM 就是这个通路的串联器, 因此 Rambus 要求所有的插槽必须插满。如果没有 RIMM, 则用 RIMM 续连器 (Continuity RIMM, C-RIMM) 代替, 以达到联通 RSL 信号并终结的目的。工作时, RDRAM 每次寻址一颗芯片, 所需的数据则由通道数据总线传送到北桥, 而不像 SDRAM 那样由所在模组直接通过 DIMM 接口传向北桥。因此, RIMM 的引脚定义几乎是左右对称的。

由于位宽的降低,为保证高带宽,RDRAM使用了更高的时钟频率(这就意味着它不可能与系统时钟同步,所以只能叫做RDRAM而不是RSDRAM),芯片的工作频率明显高于SDRAM/DDR,这样芯片的工作热量也急剧上升,为此 Rambus 在官方规范中规定 RIMM 必须配备散热片,如图 4-24 所示。



图 4-24 RAMBUS 的外观

2. RDRAM 的主要特点

RDRAM 的内部仍主要由逻辑存储体构成,但它的设计与 SDRAM 有很大的不同。逻辑存储体中存储单元的容量也并不等于 RDRAM 的接口位宽,而是它的 8 倍,因此可以说 RDRAM 是一种 8 位预取设计,这是它最主要的特点。对于 16 位芯片,其存储单元的容量为 128 位,这些数据分别通过 A 和 B 传输至逻辑存储体。由于预取为 8 位,所以 RDRAM 的突发长度也固定为 8,因为如果再高,对于 PC 应用将不太适合。不过需要特别注意的是,一个字节的传输不是由数据通道中的 8 条数据线进行并排传输,而是一个字节由一条数据线进行 8 次传输(即串行传输,这也是为什么有人说 RDRAM 是串行内存的原因),这一点也与 SDRAM 不同,这意味着北桥在进行数据读/写时,必须要等 8 个周期之后才能完成,中途不能停止。也就是说,读取时目前的北桥(如 850)一次接收 128 位(16 字节)的数据,然后再转换为两个 64 位数据分两次向 CPU 传送。

Rambus 架构引入了精简指令集计算机(RISC)技术,依靠其极高的工作频率,通过减少每个周期的数据量来简化操作。通过在上升沿和下降沿各可传送一次,使原有的 400 MHz 的频率变为 800 MHz。Rambus 架构之所以可以达到 400 MHz 的时钟频率,是因为它使用了铜线连接内存控制器和内存模块,并且通过减少铜线数量和长度,降低电磁干扰。单从技术的观点来看,作为一种全新的串行工作模式内存,若不考虑实际的难度与成本,Rambus 架构在其接口高速运行下,所能提供的带宽视 Rambus 通道的数量而定。若是 1 个通道,800 MHz 的 RDRAM 的带宽为 800 MHz × 16 b = 1.6 Gbps;若是两个通道,则可提升为 3.2 Gbps;若是 4 个通道的话,将达到 6.4 Gbps。而 PCI33 的带宽为 1.06 Gbps,PC266 的带宽则为 2.13 Gbps。RDRAM 的引脚数为 184,使用 1.8 V 电压。但由于是全新的设计,需要新的 RIMM 插槽及与芯片组配合。而且 Rambus 要求 RIMM 槽中必须全部插满,空余的 RIMM 槽要用专用的 RDRAM 连接板 Ram(也称 RDRAM 终结器)插满。

4.6.7 几种常见内存的带宽比较

前面已介绍过目前比较流行的内存主要是 PC800 RDRAM、PC266 DDR RAM、PCI33 SDRAM,表 4-1 是这几种内存的性能比较表。

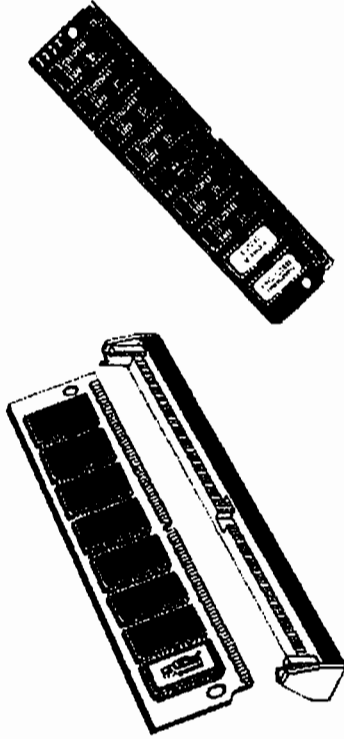
表 4-1 RDRAM、SDR/DDR 与 SDRAM 性能比较

内存种类	PC800 RDRAM	PC266 DDR RAM	PCI33 SDRAM
系统时钟频率	400 MHz	133 MHz	133 MHz
数据传输速率	800 MHz	266 MHz	133 MHz
数据带宽	1.60 Gbps(单通道)	2.13 Gbps	1.06 Gbps
电气接口	RSL	SSTL2	LVTTTL
内存总线宽度	16 位	64 位	64 位
内存槽	32	4	4
管脚数目	184	184	168
芯片封装方式	CSP	TSOP II	TSOP II
电压	1.8 V	2.5 V	3.3 V

4.6.8 内存接口类型

前面提到了 SIMM 接口类型和 DIMM 接口类型,下面简单介绍一下这两种内存接口。

SIMM 代表 Single Inline Memory Module,即单列直插内存模组,这是 586 及较早的 PC 机中常用的内存接口(如图 4-25 所示)。在更早的 PC 机中(486 以前),多采用 30 引脚的 SIMM 接口,而在 Pentium 中,应用更多的则是 72 引脚的 SIMM 接口,或者是与 DIMM 接口类型并存。



(a) 示意图

(b) 实物照片

图 4-25 SIMM 接口

DIMM 代表 Dual Inline Memory Module,即双列直插内存模组,也就是说这种接口的内存条的两面都有引脚,如图 4-26 所示。这种接口的内存广泛应用于现在的计算机中,通常每面有 84 个,所以一共有 $84 \times 2 = 168$ 个引脚。因此,人们经常把这种内存称为 168 引脚内存,而把 72 引脚的 SIMM 接口类型内存直接称为 72 线内存。DRAM 内存通常为 72 引脚,EDO - DRAM 内存

既有 72 引脚的,也有 168 引脚的,而 SDRAM 内存通常为 168 引脚的。

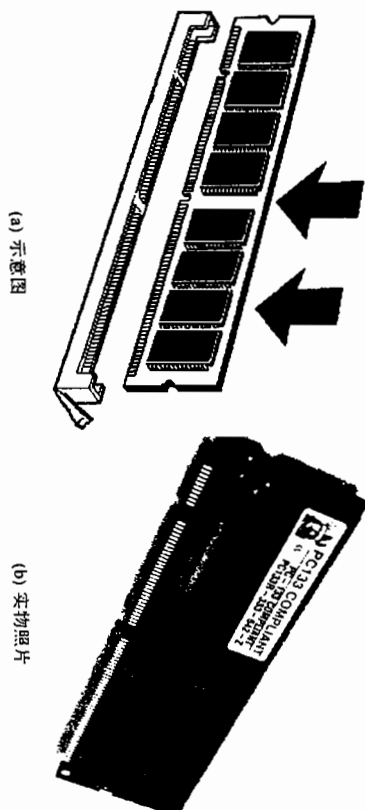


图 4-26 DIMM 接口

4.7 内存相关技术

4.7.1 内存双通道技术

随着处理器对内存带宽的要求越来越高,内存带宽成为系统越来越严重的瓶颈。只要提高内存的运行频率,就可以增加带宽,但是由于受到晶体管本身的特性和制造技术的制约,内存频率不可能无限制地提升,所以在全新的内存研发出来之前,双通道内存技术就成了一种可以有效地提高内存带宽的技术。它最大的优势在于只要更改内存的控制方式,就可以在现有内存的基础上带来内存带宽的提升。从理论指标来看,双通道内存技术具有相当的优势。双通道 DDR 400 的理论带宽为 64 Gbps,和 Intel 的前端总线为 800 MHz 的 Pentium 4 处理器及 i865、i875 芯片组完全匹配。前端总线为 800 MHz 的 Pentium 4 平台选用双通道 DDR 400,与双通道的内存控制和管理机制及高带宽有很大关系。

1. 双通道内存技术原理

双通道内存技术其实就是双通道内存控制技术(如图 4-27 所示),它能有效地提高内存总带宽,从而适应新的微处理器的数据传输,处理的需要。双通道 DDR 有两个 64 位内存控制器。双 64 位内存体系所提供的带宽等同一个 128 位内存体系所提供的带宽。

双通道体系包含了两个独立的、具备互补性的智能内存控制器,两个内存控制器能够并行运作。例如,当控制器 B 准备进行下一次读/写内存的时候,控制器 A 则正在读/写主内存,反之亦然。两个内存控制器的这种互补“天性”可以让有效等待时间缩减 50%,因此双通道技术使内存的带宽增加了一倍。它的技术核心在于:芯片组(北桥)可以在两个不同的数据通道上分别寻址、读取数据,达到 128 位的带宽。

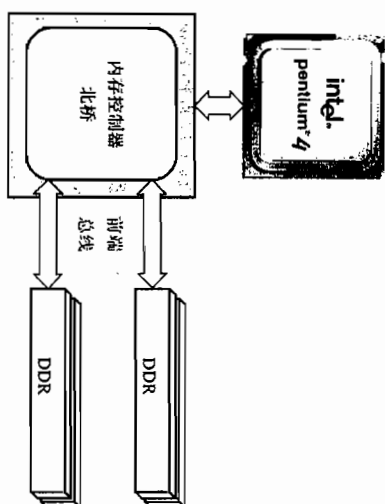


图 4-27 双通道内存技术示意图

2. 双通道内存技术的发展

双通道内存技术最初是从 Rambus 推出的 RDRAM 内存条开始的。Rambus 的内存速度非常快,但是总线宽度却比 SDRAM 内存还要小,因此它不得不结合 Intel 的双通道内存控制技术提高带宽,达到高速传输数据的目的。不过 Rambus 内存由于生产成本过高的原因,逐步被市场淘汰,但是双通道内存控制技术得到了发扬光大。如今 Pentium 4 采用的 NetBurst 架构对内存带宽要求非常高,如果内存无法提供相应的数据传输速率,处理器的总线速度再快也是无用的。

因此只有通过双通道内存控制技术才能够解决这个问题。最近金邦推出了 DDR 500 内存条,单条的数据带宽达到 4 Gbps;如果使用双通道技术,带宽将达到 8 Gbps。

3. 双通道内存技术的应用

双通道内存主要是依靠主板北桥的控制技术,与内存本身无关。双通道内存的安装有一定的要求。主板的内存插槽的颜色和布局一般都有区分。Intel 的 i865 和 i875 系列主板一般有 4 个 DIMM 插槽,每两个一组,每组颜色一般不一样,每组代表一个内存通道,只有当两组通道上都同时安装了内存条时,才能使内存工作在双通道模式下。另外,要确保对称安装,即第一个通道第一个插槽搭配第二个通道第一个插槽,依此类推。如果在相同颜色的插槽上安装内存条,则只能工作在单通道模式下。nFORCE2 系列主板同样有两个 64 位的内存控制器,其中 A 控制器只支持一个内存插槽, B 控制器则支持两个。A、B 插槽之间有一段距离,以便识别。A 控制器的内存插槽在颜色上也可能与 B 控制器两个内存插槽不同,只要将一根内存插入独立的内存插槽,而将另外一根插入另外两个彼此靠近的内存插槽,就能建成双通道模式。此外,如果全部插满内存,也能建立双通道模式,而且 nForce2 主板在组建双通道模式时对内存容量乃至型号都没有严格的要求,使用方便。如果安装方法正确,在主板开机自检时,屏幕显示内存的工作模式(如 DDR 333 Dual Channel Mode Enabled),则内存已经在双通道模式。

双通道内存技术并非 DDR 内存所独有, RDRAM 也应用了这种技术,如 Intel 的 i850E 芯片

组就支持双通道 PC1066 RDRAM。因此确切地说,双通道内存技术是双通道内存控制技术,是在当前内存技术的基础上开发的一种内存管理和控制技术。它的重点在于对内存的控制而不是内存本身,整合在芯片组北桥中的内存控制器承担了这项功能,因此说它是芯片组技术似乎更合适。

4.7.2 内存参数及其优化

为了使内存工作在最佳状态,通常需要在 BIOS 设置下设置内存参数。下面简单介绍一下影响内存性能的参数及其意义。

1. 内存延迟时序“CL - tRCD - tRP - tRAS”的设置

首先,需要在 BIOS 设置中打开手动设置。在 BIOS 设置中找到“DRAM Timing Selectable”, BIOS 设置中可能出现其他描述为 Automatic Configuration、DRAM Auto、Timing Selectable 或 Timing Configuring By SPD 等,将其值设为“Manual”(视 BIOS 的不同可能的选项有:On/Off 或 Enable/Disable)。如果要调整内存时序,应该先打开手动设置,之后会自动出现详细的时序参数列表。

(1) CL(CAS Latency)

内存读写操作前列地址控制器的潜伏时间,可能的选项:1.5/2/2.5/3。BIOS 中可能的其他描述为 tCL、CAS Latency Time 或 CAS Timing Delay。

这个参数很重要,内存条上一般都有这个参数的标记。在 BIOS 设置中,DDR 内存的 CAS 参数选项通常有“1.5”、“2”、“2.5”、“3”几种选择,SDRAM 则只有“2”、“3”两个选项。较低的 CAS 周期能减少内存的潜伏周期以提高内存的工作效率。因此只要能够稳定运行操作系统,应当尽量把 CAS 参数调低。反过来,如果内存运行不稳定,可以将此参数调高,以提高内存稳定性。

(2) tRCD(RAS - to - CAS Delay)

行寻址至列寻址延迟时间,可能的选项:2/3/4/5。BIOS 中的可能其他描述为 tRCD、RAS to CAS Delay 或 Active to CMD 等。数值越小,性能越好。

(3) tRP(RAS Precharge Time)

内存行地址控制器预充电时间,可能的选项:2/3/4。BIOS 中的可能其他描述为 tRP、RAS Precharge 或 Precharge to active。预充电参数越小,则内存读/写速度就越快。

(4) tRAS(RAS Active Time)

内存行有效至预充电的最短周期,可能的选项:1/2/3...15。BIOS 中的可能其他描述为 tRAS、Active to Precharge Delay、Row Active Time、Precharge Wait State、Row Active Delay、Row Precharge Delay 或 RAS Active Time 等。

一般情况下,可选的参数选项有 5、6 或 7;但是在一些 nForce 2 主板上,选择范围却很大,最高可到 15,最低达到 1。这个参数需要结合具体情况而定,一般最好设在 5~11 之间,并不是越大或越小就越好。具体的调整要遵循以下两个原则:

1) 当内存页面数为 4 时,tRAS 设置短一些可能会更好,但最好不要小于 5。另外,短 tRAS 的内存性能相对于长 tRAS 可能会产生更大的波动性,对时钟频率的提高也相对敏感;当内存页面数大于或等于 8 时,tRAS 设置长一些会更好。

目前的芯片组都具备多页面管理的能力,所以如果可能,请尽量选择双物理存储体的内存模组,以增加系统内存的页面数量。但怎么分辨是单物理存储体还是双物理存储体呢?就目前市场上的产品而言,256 MB 的模组基本都是单物理存储体的,双面但每面只有 4 颗芯片的也基本上是单物理存储体的,512 MB 的双面模组则基本都是双物理存储体的。

页面数量的计算公式为:物理存储体数量 \times 4,如果是 Pentium 4 或 AMD 64 的双通道主板,则还要除以 2。比如两条单面 256 MB 内存,就是 $2 \times 4 = 8$ 个页面,用在 i875 主板上组成双通道就成了 4 个页面。

2) 对于 i875 和 i865 主板,双通道时页面数达到 8 或者以上时,tRAS 设置长一些内存性能更好;对于非双通道 Pentium 4 与 AMD 64 主板,tRAS 长短之间的性能差异较小。

2. Bank Interleaving

Bank Interleaving 的意思是“存储体交错”,是加快内存访问速度的一种技术。可能的选项:Off/Auto/2/4。这里的存储体是指逻辑存储体。目前的 DDR RAM 内存芯片都是由 4 个逻辑存储体所组成,为了最大限度减少寻址冲突,提高效率,建议设为 4(设为 Auto 也可以,这样可根据 SPD 中的 L-Bank 信息来自动设置)。有人甚至认为,启用存储体交错对于系统性能提高的贡献,比将内存 CAS 延迟时间从 3 改成 2 还要大。

Intel 和 VIA 都支持存储体交错技术,主要模式有 2 路交错和 4 路交错两种。不过出于对系统的稳定性考虑,很多支持该技术的主板在默认情况下都关闭了存储体交错技术,或最多开启 2 路交错模式——虽然 4 路交错可以带来更大的性能提升。通过升级 BIOS、VIA 694X 以上芯片组都支持开启内存交错设置;即便 BIOS 不支持,也可以通过 WPCREdit 等专用软件来修改北桥芯片的寄存器,从而打开内存交错模式。

3. Burst Length

Burst Length 的意思是“突发长度”,可能的选项:4/8。Burst Length 用来决定内存读写命令能够访问的地址的最大数目。建议设为 8;如果是 Pentium 4 或 AMD 64 的双通道平台,建议设为 4。

4. Command Rate“首命令延迟”(可能的选项:1/2)

这个选项目前已经非常少见,一般还被描述为 DRAM Command Rate、CMD Rate 等。由于目前的 DDR 内存的寻址,先要进行物理存储体的选择(通过 DIMM 上 CS 片选信号进行),然后才是逻辑存储体/行激活与列地址的选择。这个参数的含义就是指在物理存储体选择完之后多少时间可以发出具体的寻址的逻辑存储体/行激活命令,单位是时钟周期。显然,也是越短越好。但随着主板上内存模组的增多,控制芯片组的负载也随之增加,过短的命令间隔可能会影响稳定性。因此,当内存条插得很多而出现不太稳定时,才需要将此参数调长。目前的大部分主板都会自动设置这个参数。

5. DRAM Clock

DRAM Clock 是指“内存时钟频率”,可能的选项:Host Clock/Hclk - 33 MHz/Hclk + 33 MHz。Host Clock 表示总线频率和内存工作频率同步,Hclk - 33 MHz 表示总线频率减去 33 MHz,Hclk + 33 MHz 表示总线频率加上 33 MHz。

在内存同步工作模式下,内存的运行速度与 CPU 外频相同。内存异步则是指两者的工作频

率可存在一定差异。该技术可令内存工作在高出或低于系统总线速度 33 MHz 的情况下(也有采用 3:4:5 的倍频模式的)。

通过调整该参数,不仅可以让“老”内存发挥余热,更重要的是可以充分挖掘内存的潜力并获得更宽泛的超频空间。Intel 的 810~875 系列芯片组和 VIA693 以后的产品,都支持内存异步。

注意:在 BIOS 中对内存进行优化设置可能会对计算机运行的稳定性造成不良影响,所以建议内存优化后一定要使用测试软件计算机运行的稳定性和速度进行测试。如果对内存的性能没有信心,那么最好采取保守设置,毕竟稳定性是最重要的。如果因内存优化而出现经常死机、重新启动或程序发生异常错误等情况,只要清除 CMOS 参数,再次设置成系统默认的数值就可以了。

4.7.3 内存技术规范及标注格式

内存技术规范统一的标注格式,一般为 PCx - xxx - xxx,但是不同的内存规范,其格式也有所不同。

1. PC66/100 SDRAM 内存标注格式

(1) 1.0~1.2 版本

这类版本内存标注格式为:PCa - bcd - eefgh,例如,PC100 - 322 - 622R。其中,a 表示标准工作频率,如 66、100、133 分别表示工作频率为 66 MHz、100 MHz、133 MHz 等;b 表示最小的 CL(即 CAS 纵列存取等待时间),用时钟周期数表示,一般为 2 或 3;c 表示最小的 tRCD(RAS 相对 CAS 的延时),用时钟周期数表示,一般为 2;d 表示 tRP(RAS 的预充电时间),用时钟周期数表示,一般为 2;e 表示最大的 tAC(相对于时钟下沿的数据读取时间,单位为 ns),一般为 6 或 5,越短越好;f 表示 SPD 版本号,所有的 PC100 内存条上都有 EEPROM,用来记录此内存条的相关信息,符合 Intel PC100 规范的为 1.2 版本以上;g 代表修订版本;h 代表模块类型;R 代表 DIMM 已注册,256 MB 以上的内存必须经过注册。

(2) 1.2b + 版本

其格式为:PCa - bcd - eefghR,例如,PC100 - 322 - 54122R。其中,a 表示标准工作频率,单位为 MHz;b 表示最小的 CL(即 CAS 纵列存取等待时间),用时钟周期数表示,一般为 2 或 3;c 表示最少的 tRCD(RAS 相对 CAS 的延时),用时钟周期数表示;d 表示 tRP(RAS 的预充电时间),用时钟周期数表示;ee 代表相对于时钟下沿的数据读取时间,表达时不带小数点,如 54 代表 tAC 为 5.4 ns;ff 代表 SPD 版本,如 12 代表 SPD 版本为 1.2;g 代表修订版本,如 2 代表修订版本为 1.2;h 代表模块类型;R 代表 DIMM 已注册,256 MB 以上的内存必须经过注册。

2. PC133 SDRAM(2.0 版本)内存标注格式

VIA 和 Intel 都提出了 PC133 SDRAM 标准,VIA 力推的 PC133 规范是 PC133 CAS = 3,沿用了 PC100 的大部分规范,例如 168 引脚的 SDRAM,3.3 V 的工作电压以及 SPD;Intel 的 PC133 规范要严格一些,是 PC133 CAS = 2,要求内存芯片至少 7.5 ns,在 133 MHz 时最好能达到 CAS = 2。

PC133 SDRAM 标注格式为:PCab - cde - ffg,例如,PC133U - 333 - 542。其中,a 表示标准工作频率,单位为 MHz;b 代表模块类型(R 代表 DIMM 已注册,U 代表 DIMM 不含缓冲器);c 表示最小的 CL(即 CAS 的延迟时间),用时钟周期数表示,一般为 2 或 3;d 表示 RAS 相对 CAS 的延

时,用时钟周期数表示;e 表示 RAS 预充电时间,用时钟周期数表示;ff 代表相对于时钟下沿的数据读取时间,表达时不带小数点,如 54 代表 tAC 为 5.4 ns;g 代表 SPD 版本,如 2 代表 SPD 版本为 2.0。

3. PC1600/2100 DDR SDRAM(1.0 版本)内存标注格式

其格式为:PCab - cde - ffg,例如,PC2100R - 2533 - 750。其中,a 表示内存带宽,单位为 Mbps,a/16 = 内存的标准工作频率,例如 2100 代表内存带宽为 2100 Mbps,对应的标准工作频率为 2100/16 = 133 MHz;b 代表模块类型(R 代表 DIMM 已注册,U 代表 DIMM 不含缓冲器);cc 表示 CAS 延迟时间,用时钟周期数表示,表达时不带小数点,如 25 代表 CL = 2.5;d 表示 RAS 相对 CAS 的延时,用时钟周期数表示;e 表示 RAS 预充电时间,用时钟周期数表示;ff 代表相对于时钟下沿的数据读取时间,表达时不带小数点,如 75 代表 tAC 为 7.5 ns;g 代表 SPD 版本,如 0 代表 SPD 版本为 1.0。

4. RDRAM 内存标注格式

其格式为:aMB/b c d PCe,例如,256MB/16 ECC PC800。其中,a 表示内存容量;b 代表内存条上的内存芯片数量;c 代表内存支持 ECC;d 保留;e 代表内存的数据传输速率,e/2 = 内存的标准工作频率,例如 800 代表内存的数据传输速率为 800 Mbps,对应的标准工作频率为 800/2 = 400 MHz。



一、选择题

- 下列有关存储器的读写速度排列正确的是()。
 - RAM > 高速缓存 > 硬盘 > 软盘
 - 高速缓存 > 硬盘 > RAM > 软盘
 - 内存 > 硬盘 > 软盘 > RAM
 - 硬盘 > 软盘 > RAM > 高速缓存
- 内部存储器的功能不包括()。
 - 暂存正在运行的程序
 - 作为 CPU 运行程序的区域
 - 配合 CPU 与外设打交
 - 具有非易失性,且解决了存储容量不足的问题
- 内存可分为()两类。
 - 动态内存和静态内存
 - 只读存储器和随机存储器
 - 高速内存和低速内存
 - 高速缓存和 ROM
- ROM 与 RAM 的主要区别在于()。
 - 在掉电后,ROM 中存储的信息不会丢失,RAM 信息会丢失
 - 掉电后,ROM 信息会丢失,RAM 则不会
 - ROM 是内存存储器,RAM 是外存储器
 - RAM 是内存存储器,ROM 是外存储器
- 断电会使存储数据丢失的存储器是()。
 - ROM
 - EEPROM
 - DRAM
 - 硬盘

二、填空题

- ROM 和 RAM 共同分享()的地址空间,但其区别在于:前者存放一次性写入的数据或程序,只能()不能();而后者则可以通过指令()。

2. 为保证动态 RAM 中的内容不消失,需要进行()操作。
3. RAM 多采用 MOS 型半导体集成电路芯片制成,根据其保存数据的机理的不同,RAM 可分为 DRAM 和()两大类。

4. 计算机的主存储器用来存储数据和指令,为了实现按地址访问,每个存储单元必须有一个唯一的地址,PC 机主存储器的编址单位是()。

5. EPROM 是可改写的,()作为读写存储器。(填“能”或“不能”)

三、简答题

1. 试给出 DRAM 优于 SRAM 的两个优点。
2. 如何区别存储器和寄存器? 两者是一回事的说法对吗?
3. SDRAM、DDRAM 和 RDRAM 的主要特点和区别是什么? 存储器是如何分类的?
4. 内存的主要性能指标有哪些?
5. 动态 RAM 为什么要刷新?
6. 解释下列名词:
SDRAM, EEPROM, SPD, RDRAM, ECC, DIMM



1. 在网上查找测试计算机内存的软件,下载并安装,然后对所用计算机进行测试。
2. 对测试结果进行分析,写出试验报告。

总线与芯片组

第5章

总线是计算机各个部件之间传输信息的通道。计算机的主要部件,如 CPU、存储器以及各种接口电路通过总线进行连接。由于总线是公用的,总线的性能对计算机的性能具有举足轻重的影响。在现在的计算机中,往往同时存在多种总线,分别担任不同的传输任务。

计算机的主机系统包含 3 个核心部件:CPU、存储器和芯片组。芯片组确定了计算机主机系统除 CPU、存储器以外的大部分功能,集成了总线、各种外部接口和常用功能模块。

本章介绍总线和实现总线的芯片组。

5.1 总线

总线是各种信号线的集合,是计算机各部件之间传送数据、地址和控制信息的公共通路。总线往往满足以下 3 个特性:高速性、公用性和标准性。

多个部件之间的信号传输都是通过一个总线通道进行的,而总线在某一时刻,只能有一个点对点传输,在不同的时间段完成不同的部件之间的信号传输。

5.1.1 总线的概念

在微型计算机系统中,CPU 必须与一定数量的部件和外围设备连接,但如果将各部件和每一种外围设备都分别用一组线路与 CPU 直接连接,那么连线将会错综复杂,甚至难以实现。为了简化硬件电路设计及系统结构,常用一组线路,配置以适当的接口电路,与各部件和外围设备连接,如图 5-1 所示,这组共用的连接线路被称为总线。

在计算机系统中采用总线连接,可以简化系统结构,降低成本,而且容易扩展。为了使不同厂商研制的不同计算机部件能通过总线实现连接,要求总线必须符合一定的标准,而且是开放的技术标准。

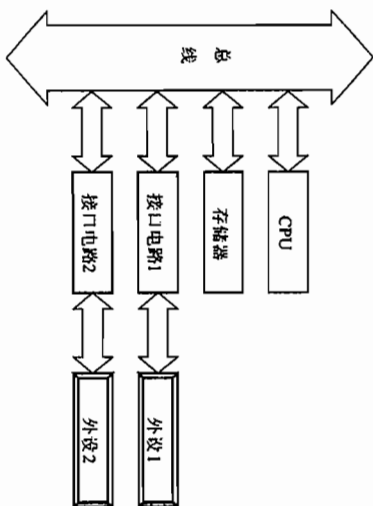


图 5-1 通过总线连接 CPU、存储器和接口电路

由于在总线上传输的信息可以分为 3 类,即数据信息、地址信息和控制信息,而且这 3 类信息在总线上有独立的传输通道,所以一般的传统总线都由 3 部分信号线组成,即数据总线、地址总线和控制总线。

- 数据信号:是信息本身。
- 地址信号:说明信息的存储位置,无论是存储器,还是端口的寄存器,信息的存放都是有地址的。
- 控制信号:说明当前的总线状态,控制信号主要包括存储器读/写控制、端口读/写控制、中断申请/应答信号、DMA 控制信号等。

大多数总线都有时钟信号、电源等。
需要说明的是并非所有的总线都存在独立的数据总线、地址总线和控制总线。如 USB,它虽然在应用上更像接口,如果按概念分类,USB 却属于总线,通过 USB 的信号线传输的信息同样也包括数据、地址和控制信息,但 USB 却没有独立的数据总线、地址总线和控制总线,它们是集中在一个信号线中传输的。由于 USB 的应用属于接口,将在本书关于接口的内容中介绍。

随着计算机,尤其是微型计算机技术的发展,总线技术也经历了一个发展的过程。传统的总线在结构上与 CPU 有关,某一种总线的出现,可能和某一一系列的 CPU 的结构有密切的关系,这就使得总线结构和 CPU 密切相关,这种总线一般不具有通用性,在总线所连接的所有部件中,各个部件的地位不同,总线由 CPU 完全控制。

现代的总线力求在设计上与 CPU 无关,同时,在总线上连接的部件(包括 CPU)都可以控制总线,也就是说各个总线部件的地位是平等的,这样的总线支持多 CPU 系统。但由于没有了集中的总线控制中心,现代的总线必须具备总线仲裁机制,以确定当前的总线由哪一个部件控制。

5.1.2 总线的类型和性能

从广义上看,总线可能有很多种类型。

1. 片内总线

是指在 CPU 内部的总线,用于连接 CPU 内部的各个功能部件,如运算器、译码部件、寄存器、控制器等。片内总线一般不对外开放。

2. 处理器总线

指 CPU 引脚总线,所以叫前端总线(Front Side Bus,FSB),有时称为系统总线。前端总线的提法是在多总线系统中才有的,8086、80286 系统一般为单总线系统,80386 以后的系统为多总线系统。前端总线是现代微型计算机中核心部分的总线,速度最快,用于连接高速缓存、存储器、北桥芯片等高速部件。

3. I/O 总线

和前端总线相对应的是计算机外部的总线,其传输速度低于前端总线。I/O 总线是对外开放的,在计算机主机板上形成扩展插槽,用于连接扩展部件。计算机外部设备的连接和扩展要通过 I/O 总线完成。

4. 外部总线

用于连接计算机外部设备、符合总线特征的连接通道,在现代微型计算机中最典型的就是通用串行总线——USB。因为在逻辑上计算机的 USB 可以扩展,能连接多个 USB 设备,但是 USB 在应用上更接近接口,所以 USB 的内容在有关接口的内容中介绍。

5. 片间总线

芯片一级的总线,一般线数目很少,使用串行传输,速度不高,用于连接在一个系统中的多块芯片,常见的有 i²C(Integrated Circuit)、SPI(Serial Peripheral Interface)等。

总线的最主要参数是总线带宽——单位时间内传输的数据量,单位是 Bps(字节每秒)。影响总线带宽的因素主要有两个,即总线位宽和总线的时钟频率。

总线的位宽是指一次(一个总线时钟周期)可以传送的位数,可以是 8 位、16 位、32 位、64 位等。

$$\text{总线带宽} = (\text{总线位宽} \div 8) \times \text{总线时钟频率}$$

对于串行传输的总线,描述其性能的指标是传输速率,单位是 bps(位每秒)。

需要指出的是,总线带宽在大多数系统中描述的是总线传输的极限速率,在实际的系统中,总线为了完成数据传输,还需要传输地址和控制信号,实际的数据传输速率是达不到标称的总线带宽值的。

5.1.3 PC 机总线的演变

在微型计算机中,总线经历了从单总线到多总线的结构变化,从最初的 8086、80286 的使用的单总线结构,发展到之后的多总线结构。

1. 8086/80286 时代

这一时期的微型计算机采用工业标准体系结构(Industry Standard Architecture,ISA)总线,计算机中只有一条总线。按现代的观点来看,ISA 是一种简单的总线,它几乎就是处理器总线的延伸,存储器和所有的接口电路都通过 ISA 实现连接。

ISA 总线有 8 位和 16 位两种,分别在 8086 和如 80286 微型计算机中使用。ISA 总线完全由

CPU 控制,总线时钟和 CPU 时钟是一致的。

2. 80386/80486 时代

随着计算机运算速度的不断提高,ISA 总线逐渐不能满足要求了,主要原因是 CPU 时钟频率不断提高,由于 ISA 的设计上的原因,ISA 的时钟频率不能大幅度提高,从而限制了 IAS 总线传输速率的提高。其次,从 80386 开始进入 32 位 CPU 时代,16 位的 IAS 总线成为计算机系统性能的瓶颈。再次,外部设备,尤其是计算机的显示适配器(显卡)、音频适配器(声卡),要求有传输速率更高的总线。在这一时期的微型计算机中,除了保留了 ISA,又增加了新的 32 位总线。主要有 MCA、EIAS 和 VL-Bus。

(1) MCA

微通道体系结构(Micro Channel Architecture, MCA)是 IBM 公司为取代 ISA 而设计的全新的 32 位总线,采用了很多新技术。它的吞吐量比 ISA 大得多,支持总线主控,并且增加了即插即用功能。其缺点是不能兼容 ISA。

由于 MCA 是 IBM 公司的专有技术,不能免费使用,从而阻止了 MCA 的推广应用,MCA 只是在 IBM 公司当时生产的微型计算机上使用。

(2) EISA

扩展工业标准体系结构(Extended Industry Standard Architecture, EISA)是 ISA 的 32 位版本,是由 Compaq 组织的 EISA 协会开发的。EISA 采用了外观尺寸和 ISA 一样的扩展插槽,但增加了两排连接器,将原来的 16 位 ISA 扩展为 32 位总线,同时地址线也扩展为 32 条,以满足 80386 CPU 的要求。扩展后的 EISA 和 ISA 兼容,ISA 插卡可以插到 EISA 插槽,但 EISA 在普通桌面计算机上的应用很少,主要是应用在当时的高档服务器上。

EISA 的时钟频率是 8.33 MHz,最大带宽 $= 8.33 \times 32 \div 8 = 33 \text{ MB/s}$ 。

(3) VL-Bus

VESA 局部总线(VESA Local Bus, VL-Bus)是由视频电子标准协会(Video Electronics Standards Association, VESA)制定的总线标准。为了改善 PC 机的视频性能,VESA 在 1992 年引入了 VL-Bus 这一 32 位总线,使数据传输速率更接近 CPU 的频率,VL-Bus 的最大带宽为 132 MB/s。

VL-Bus 是在 ISA 的扩展插槽的基础上新增一段插槽,增加了主板的占用面积。但由于设计方面的不足,不支持总线主控,即插即用等新特性,很快便被后来的 PCI 所取代。

3. Pentium 时代

Pentium 时代前期主要是指 Pentium 到 Pentium 4 中期这一段,这时的微型计算机总线特征,是:FSB + PCI + ISA 多总线模式。

外设部件互连(P peripheral Component Interconnection, PCI)标准的出现淘汰了其他的扩展总线,成为 Pentium 时期标准的 I/O 总线。PCI 在技术上和指标上都有提高。由于 PCI 时钟不依赖于 CPU,使得 CPU 时钟有较大的提升空间,同时也使得 PCI 有较长的生命周期。

在这一时期,FSB 时钟频率和 CPU 的外频同步,不同的 CPU 对应不同的前端总线频率,使得总线的性能跟上 CPU 的发展。

由于当时在用的 ISA 扩展卡很多,新的系统保留 ISA 总线可以充分利用现有的硬件设备,有利于总线标准的平稳升级。

随着多媒体和三维的应用的逐渐增多,显示适配器的刷新速度也随之提高,从而对 PCI 总线形成压力。为了缓解显示适配器严重占用 PCI 总线的情况,出现了显示适配器专用接口——加速图形端口 I(Accelerated Graphics Port, AGP)。AGP 的出现使得 PCI 总线设备中负载最重的设备从 PCI 总线设备中分离出去,均衡了整个系统的负载。

这一阶段的微型计算机主机系统的典型结构如图 5-2 所示。这是 Pentium II CPU 和 Intel 的芯片组成的系统,CPU 和北桥芯片之间通过 64 位前端总线连接,前端总线带宽为 800 MBps。北桥芯片连接存储器和 AGP 接口,存储器通道和 AGP(AGP 2×为普通 AGP 的速度的 2 倍)的最大带宽分别为 800 MBps 和 533 MBps。北桥芯片的另一个重要功能是为提供 PCI 总线,形成 PCI 扩展插槽。

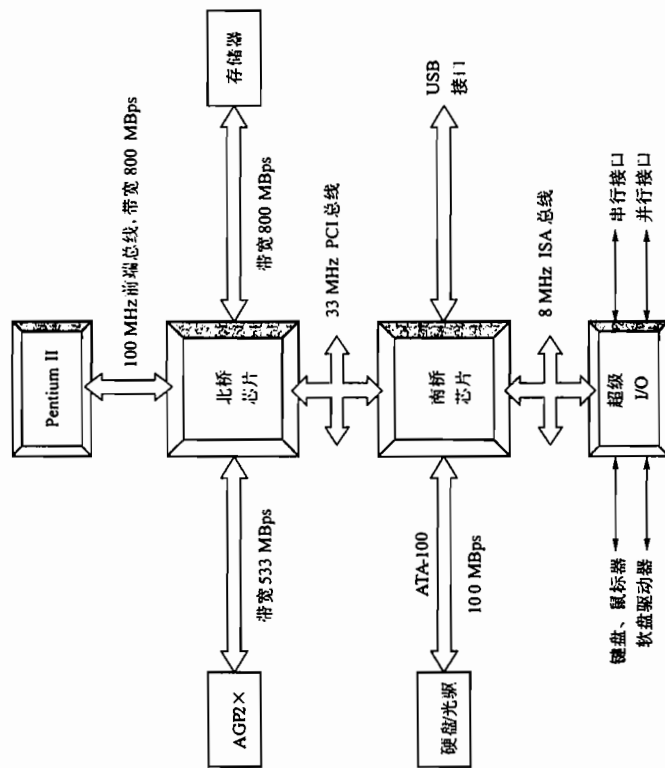


图 5-2 Pentium II 计算机的系统结构

北桥芯片和南桥芯片通过 PCI 总线连接,总线带宽为 133 MBps。南桥芯片生成 USB 和 ATA 接口,再通过 ISA 总线连接超频 I/O 电路,由超频 I/O 电路生成键盘、鼠标、软盘驱动器等设备的专用接口和串行、并行通用接口。

这一阶段的计算机主机板上同时存在 3 种不同的扩展插槽,分别是 ISA 总线扩展槽、PCI 总线扩展槽和 AGP 插槽,三种扩展槽在主板上的位置是固定的。

4. Pentium D/EE 时代

这一时期的微型计算机总线的特征是:FSB + PCI Express + PCI 的多总线模式。

由于速度更快的 PCI Express 总线的出现,成为 PCI 的更新换代的总线标准。但在新的

PCI Express出现之后, PCI还将存在很长一段时间。

5.1.4 ISA 总线

ISA 总线结构简单,如图 5-3 所示,对于想了解总线结构的人来说,应该先了解 ISA 总线。ISA 总线有 8 位和 16 位两种,它们分别是 8086 和 80286 CPU 的处理器总线的延伸。16 位的 ISA 总线是在 8 位 ISA 总线的基础上扩展而来的,其兼容 8 位的 ISA 总线。

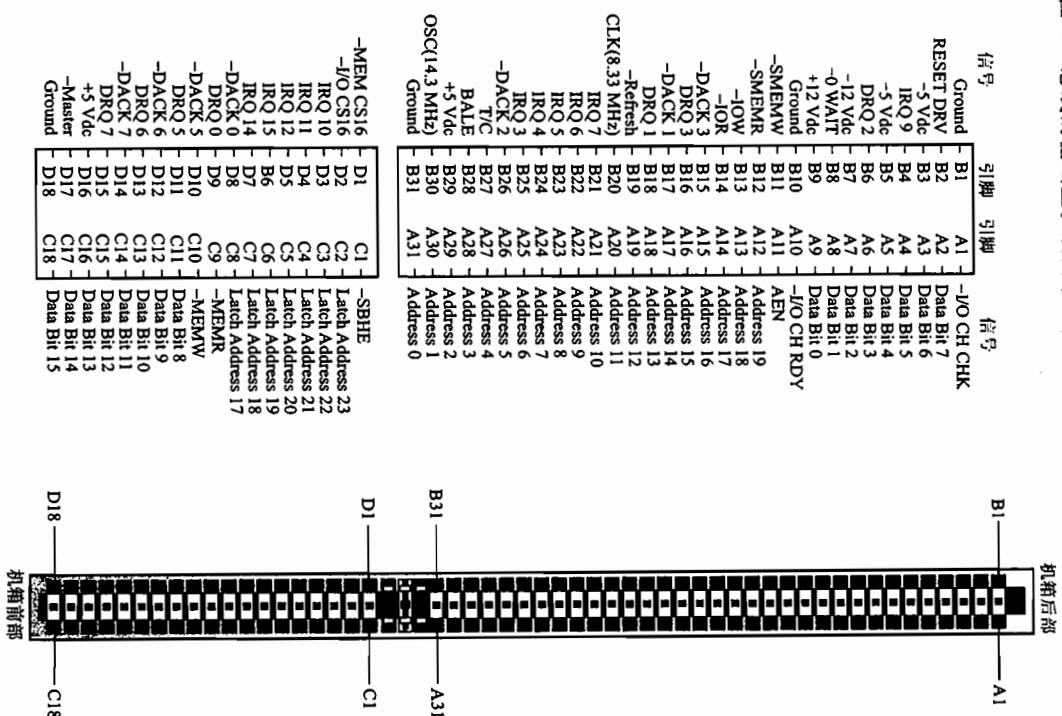


图 5-3 ISA 总线插槽

16 位 ISA 总线的基本指标如下:

- 数据线宽 16 位, 一次传输需要两个时钟周期。
- ◆ 地址线宽 24 位, 存储器寻址空间 16 MB。
- 最高时钟频率 8.33 MHz, 最大传输带宽 8.33 Mbps。
- 中断功能, 总线中有 16 个中断通道中的 10 条中断请求信号线。
- 支持 DMA, 总线中有 8 个 DMA 通道的请求和应答信号线。
- 电源部分包括 +5 V、-5 V、+12 V 和 -12 V 四组直流电源。
- 全部 98 条引脚。

由于 ISA 总线的信号大多数直接来自 CPU, 信号的功能比较直观。当 CPU 执行指令时, 不同的指令会在 CPU 总线上出现不同的信号组合。因此, 可以说 CPU 总线的状态是 CPU 执行指令的外部表现。

根据ISA的信号线命名,可以很容易分辨出信号线:

- Data Bit 0 ~ 15, 16 位数据线。
- Address 0 ~ 23, 24 位地址线。
- IRQx: 中断请求线。
- MEMR、MEMW, 存储器读、写控制线。
- IOR、IOW, 端口读、写控制线。
- DRQx, DMA 请求信号线。
- DACKx, DMA 应答信号线。
- CLK, 总线时钟;
- +5 Vdc, -5 Vdc, +12 Vdc, -12 Vdc, 直流电源线。

5.1.5 PCI总线

随着图形处理技术和多媒体技术的广泛应用,要求计算机的显示系统有高速的图形处理能力和 I/O 处理能力。这不仅要求显示适配卡要改善其性能,原有的 ISA、EISA 总线也已远远不能适应要求,需要更高速度的总线。1991 年,Intel 公司首先提出了 PCI 总线的概念,并联合多家计算机公司和厂商成立了 PCI 集团,旨在推广 PCI 总线。

PCI 总线主要是为了满足对高速的图形图像的传输与显示要求而开发的一种先进的高性能 32 位局部总线标准,它已经成为目前最流行的 PC 机扩展总线标准。

1. PCI总线的基本指标

PCI 总线的地址总线与数据总线均为 32 位,总线时钟频率为 33 MHz,基本数据传输速率为 $33 \text{ MHz} \times 32 \text{ b}/8 = 133 \text{ MBps}$ 。

PCI 总线的数据带宽可升级为 64 位。总线频率为 33 MHz 时, 升级后传输速率可达 266 MBps。

PCI 总线经历了几次版本升级,表 5-1 列出了 PCI 的不同版本。

表 5-1 PCI 总线版本

PCI 版本	发布年代	主要特征
PCI 1.0	1992 年 6 月	最初的 32/64 位标准
PCI 2.0	1993 年 4 月	定义了连接器和扩展板
PCI 2.1	1995 年 6 月	扩充 66 MHz 操作,事务顺序,反映时间变化
PCI 2.2	1999 年 1 月	电源管理,机械净化
PCI-X 1.0	1999 年 9 月	扩充 133 MHz 操作
Mini-PCI	1999 年 11 月	小型扩展板

Mini-PCI 是 PCI 规格的小型化版本。如果使用转接器,可以将 Mini-PCI 设备插在标准的 PCI 插槽中。Mini-PCI 有几种不同的规格:Type I、Type II 和 Type III。目前我们所说的 Mini-PCI 是指 Type III。Mini-PCI 提供与 PCI 总线相同的 32 位、33 MHz 的性能。

对于工作站来说,它的快速运行需要更加快速的总线。因此,从 PCI 2.1 版开始,相关厂商推出了 64 位的 PCI,频率在 33.3 MHz 时它的传输速率可以达到 266.6 MBps,而在 66.7 MHz 时则可以达到 533.3 MBps。对于服务器而言,需要比普通 64 位 PCI 更高的带宽,这主要是为了满足高速 I/O 接口的需求,因此就有了扩展版的 PCI-X (PCI eXtended)。PCI-X 仍然是 64 位,但是传输速率却有很大的提高,PCI-X 1.0 包括 PCI-X 66 与 PCI-X 133 (66、133 指的是总线频率,66 代表 66.7 MHz,133 代表 133.3 MHz),传输速率分别可以达到 533.3 MBps 和 1066.6 MBps;而在这之后的 PCI-X 2.0 包括 PCI-X 266 与 PCI-X 533,传输速率更是达到了 2133.3 MBps 和 4266.6 MBps,如表 5-2 所示。

表 5-2 PCI 总线类型

PCI 类型	总线宽度 (位)	总线时钟频率 (MHz)	带宽 (MBps)
PCI	32	33	133
PCI 66M	32	66	266
PCI 64 位	64	33	266
PCI 66M/64 位	64	66	533
PCI-X 1.0	64	66/133	533/1066
PCI-X 2.0	64	266/533	2133/4267

2. PCI 总线扩展槽

PCI 总线采用了多路复用技术,即其地址线和数据线引脚共用,这样可以减少引脚数量。ISA 总线的地址总线与数据总线是物理上分开的两组信号线,地址总线传输地址码,数据总线传输数据。而 PCI 总线采用多路复用技术,地址总线和数据总线共用一组信号线,分时传送数据和地址信息。

同 ISA 扩展槽相比,PCI 扩展槽的长度要短得多,而且颜色一般也是白色,如图 5-4 所示。32 位 PCI 插槽也分为长槽和短槽两部分,全部引脚数为:49 × 2 + 11 × 2 = 120。64 位 PCI 插槽是在 32 位 PCI 插槽的基础上扩展而来的。



图 5-4 PCI 总线扩展插槽

PCI 有 5 V 规范、3 V 规范和通用规范,这 3 种规范都有 32 位和 64 位两种版本,其中 5 V 规范用于工作站级系统,3 V 规范用于便携式系统,通用规范用于上述两种系统中使用的主板和扩展卡。

3. PCI 总线与 CPU 无关

PCI 总线是一种不依附于某个具体处理器的局部总线。从结构上看,PCI 是在 CPU 和原来的系统总线之间插入的一级总线,由一个桥接电路实现这一层的具体管理,并实现上下之间的接口以协调数据的传送。管理器提供了信号缓冲,使之能支持 10 种外设,并能在高时钟频率下保持高性能。

4. PCI 支持总线主控

总线主控技术就是允许智能设备在需要时取得总线控制权,以便加速数据传输并对高速专用性任务提供支持。PCI 总线标准定义了两种设备:目标设备和主设备。目标设备接受主设备的命令并响应主设备的请求,主设备则是智能化的,能独立于总线和其他设备运行。总线主设备、处理器和目标设备可以共享总线,主设备也能成为其他主设备的目标设备。

5. PCI 总线仲裁

总线仲裁是现代总线系统的核心问题之一。为了解决多个主设备同时竞争总线控制权的问題,必须具有总线仲裁部件。它通过采用优先级策略或公平策略,选择其中一个主设备作为总线的下一个主方,接管总线控制权。

按照总线仲裁电路的位置不同,总线仲裁分为集中式仲裁和分布式仲裁。集中式仲裁方式必有一个中央仲裁器,它受理所有功能模块的总线请求,按优先原则或公平原则进行排队,然后仅给一个功能模块发出授权信号。分布式仲裁不需要中央仲裁器,每个功能模块都有自己的仲裁器和仲裁器。通过分配优先级仲裁号,每个仲裁器将由仲裁总线上得到的仲裁号与自己的仲裁号进行比较,从而获得总线控制权。

PCI 总线采用集中式仲裁方式,每个 PCI 主设备都有独立的 REQ# (总线请求) 和 GNT# (总线授权) 两组信号线与中央仲裁器电路相连。由中央仲裁器根据一定的算法对各主设备的请求进行仲裁,决定把总线使用权授予谁。

6. 高性能的突发数据传输模式

PCI 提供的通道能满足多媒体和高速网络的要求,其重要的一点是 PCI 能支持一种名为线性突发的数据传输模式 (Burst Mode)。在这种模式下,PCI 能以很高的速度访问地址连续的数据。

据,这种特别适合于高分辨率且多达数百万种颜色的快速图像显示,而且可确保总线不断满载数据。这种线性的或顺序的寻址方式,意味着当由一个地址开始读/写大量数据时,每次只需将地址自动加1,便可接收数据流内下一个字节的数据。线性突发传输能更有效地动用总线的带宽传输数据,减少无谓的寻址操作。

7. 支持“即插即用”功能

PCI总线具有自动配置和自动识别外设的能力。当外设或外设接口卡接入系统时,能自动为其分配系统资源,如端口地址、中断号、DMA通道号配置等。PCI总线的这种“即插即用”(Plug & Play)技术主要是通过扩展卡中存储有关PCI外设的权限信息来完成的。

5.1.6 PCI Express 总线

计算机总线大约每隔3年在性能上提高一倍,从原先8位、16位ISA总线,到32位EISA、MCA、VL-Bus、PCI,到64位的PCI-64/66 MHz,再到如今的PCI-X 1.0和2.0。而处理器的性能提升则要快得多,差不多是每18个月就提升一倍。这种发展的不平衡就造成了I/O系统成为瓶颈,无法满足处理器、内存子系统的需要。新的高速CPU的出现促使新的总线标准PCI Express诞生。

从名称上来看,PCI Express似乎同PCI有着很大的关系,但实际上它并不是PCI技术的延续,就如同PCI取代ISA一样,PCI Express的出现将改变计算机的架构。PCI Express属于第三代总线,是由Intel公司提出的,得到了众多知名计算机厂商的支持。这种新的总线解决方案提供的是一个智能化的总线架构,它可以有效地解决多个设备共享总线的问题。

1. PCI Express 总线的技术特点

(1) 高性能

PCI Express总线只需要从芯片组中引出很少的引脚,使得主板布线难度大大降低,但是却具有比现在的PCI高得多的带宽,传输速率最高可以达到8 Gbps。

(2) 配置灵活

在配置的灵活性方面,PCI Express也优于PCI。它可以根据所连接的硬件设备的不同,使用不同的时钟频率与设备通信。

(3) 多种连接方式

这是同PCI总线非常不同的地方,PCI Express总线可以“走出机箱”。也就是说,PCI Express可以如同USB或者Firewire接口一样,通过计算机上的一定接口同采用符合PCI Express标准的接口的外部设备进行连接和通信。

(4) 点对点总线

对于PCI这种“总线式”的连接方式,一旦PCI总线成为瓶颈,将会影响所有连接其上的PCI设备。PCI Express总线采用了点对点技术,每个PCI Express设备都是直接同系统芯片进行通信,而不再存在带宽问题。

(5) 高级功能

PCI Express可以使用多种不同的信号协议,包括它本身的协议。它还具有高级电源管理和监视功能,所有的PCI Express设备都支持热插拔。在PCI Express中,诸如内存纠错等功能成为标准功能。

(6) 跨平台的兼容性

PCI Express最大的优点之一就是它的跨平台兼容性。现在的符合PCI 2.3规范的板卡可以在低带宽的PCI Express插槽上使用(软件级兼容PCI)。

2. PCI Express 物理串行连接

一个单独的基本的PCI Express串行连接就是两个独立的通过不同的低电压驱动信号实现的连接,一个接收对和一个发送对(共4条线路)。一个差分信号在两个接口之间传递,如图5-5所示。第一代的PCI Express连接信号传输速率为每对线单向2.5 Gbps。

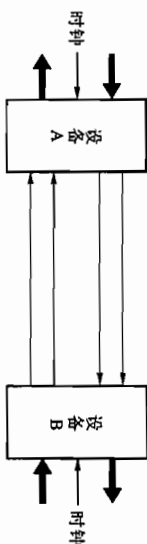


图 5-5 PCI Express 的基本串行连接

PCI Express基本串行连接的另一项关键技术是通过8b/10b的编码,植入了时钟信号。时钟信号被直接植入数据流中,而不是作为独立信号存在。8b/10b编码需要每个字符占据10b,也就是比通常的字符的传输位数多出20%。

PCI Express连接结构可以由一个或多个基本串行连接构成,构成 $\times 1$ 、 $\times 2$ 、 $\times 4$ 、 $\times 8$ 、 $\times 12$ 、 $\times 16$ 和 $\times 32$ 几种不同形式。一个 $\times 1$ 的连接具有4根信号线,一个 $\times 16$ 连接在每个方向上具有16个差分信号对。最多的连接是 $\times 32$ 的连接,可以每个方向上进行10 Gbps的传输(2.5 Gbps $\times 32/8b$)。不过由于8b/10b编码,事实上的带宽为8 Gbps(扣除20%的植入时钟信号)。

3. PCI Express 插槽

PCI Express规范中定义了各种各样的接口(如图5-6所示),从 $\times 1$ 到 $\times 16$ 总线宽度,而 $\times 2$ 模式则被保留用于其他类型的PCI Express内部互连。较小的PCI Express卡可以插入较大的插槽中。PCI Express卡可以支持热插拔,采用的3个电压分别是+3.3 V、+3.3 V aux和+12 V。目前,取代AGP插槽的是PCI Express $\times 16$,带宽为5 Gbps,有效带宽4 Gbps。



图 5-6 在主板上的PCI-E x1、PCI-E x16和PCI总线插槽

5.2 芯片组

在现代的 PC 机中,芯片组决定主机的众多技术指标,它的重要性甚至超过 CPU 和存储器。

5.2.1 芯片组的概念

芯片组包括处理器接口、存储器控制器、总线控制器、I/O 控制器等。目前,芯片组包含了除 CPU 和存储器之外的构成微型计算机的几乎所有集成电路。在 PC 机中,芯片组用来实现处理器与其他部件的联络。不经过芯片组,处理器无法与存储器、适配器、外围设备等进行连接。芯片组是 PC 机的通信中心。当决定一个系统时,应该首先选择的是芯片组,之后才会决定处理器、存储器、I/O 及扩展功能。

如果仔细看看现在的 PC 机主板就会发现,除 CPU 和存储器之外,主板上主要的半导体部件就是芯片组了,如图 5-7 所示。大多数情况下,芯片组由两块芯片组成。

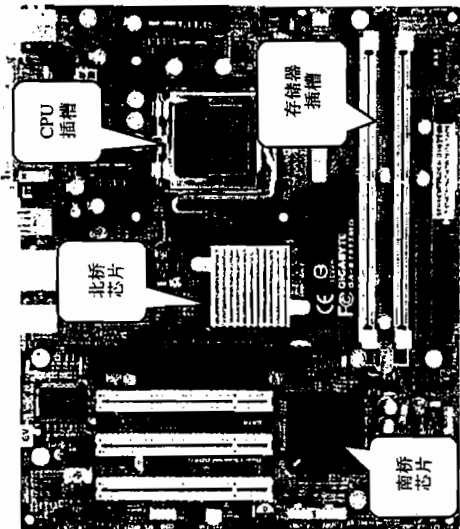


图 5-7 由芯片组构成的主板

当 IBM 生产第一块 PC 机主板时,它使用了很多分离的芯片完成设计。除处理器及可选的数学协处理器外,还要求有许多其他部件以完成整个系统。这些部件包括时钟生成器、总线控制器、系统定时器、中断及 DMA 控制器、CMOS RAM 及时钟、键盘控制器。为完成整个主板线路,还要有大量其他简单逻辑芯片。

早期的芯片组相当于这些专用芯片的简单集成。在以后的发展中,芯片组的功能越来越强大。在现代的 PC 机结构中,芯片组还嵌入了更多的功能模块,包括处理器接口、存储器控制器、

I/O 控制器等。

计算机采用芯片组的主要原因如下:

- 计算机电路非常复杂,必须使用集成度更高的半导体器件。
- 可以大大缩短从推出新的 CPU 到开发出采用新 CPU 的计算机的时间。
- 提高计算机的可靠性,降低成本。
- 是计算机产业细分的结果。

由芯片组构成 PC 的主机系统的方式是从 80386 时代开始普遍采用的,由此也出现了一些 CPU 厂商之外的其他专门开发和生产芯片组的厂家。目前生产 CPU 的厂家主要有 Intel 和 AMD,生产芯片组的厂家除了 Intel 和 AMD 之外,主要还有 VIA(威盛)、NVIDIA、ALI(扬智)、SIS(矽统)、ATI 等。

5.2.2 Intel 公司早期的芯片组

Intel 早期的芯片组主要是指 Pentium 开始、Pentium 4 之前的产品,表 5-3 列出了从 1993 年到 2000 年这一时期 Intel 的主要芯片组系列。在这一时期,使用的 CPU 主要是 Intel 第五代 CPU Pentium、Pentium MMX 和第六代 CPU Pentium Pro、Pentium II、Pentium III 等。

表 5-3 Intel 早期的芯片组

芯片组系列	年代	结构	CPU	系统总线频率	存储器	PCI	AGP
430	1993 ~ 1997	南北桥	Pentium	66	FPM/EDO/SDRAM	2.0/2.1	无
440	1996 ~ 1998	南北桥	Pentium II / Pentium III	66/100	EDO/SDRAM	2.1	2 x
810	1999	Hub	Pentium II / Pentium III	66/100/133	SDRAM	2.2	2 x
815	2000	Hub	Pentium II / Pentium III	66/100/133	SDRAM	2.2	4 x
820	1999 ~ 2000	Hub	Pentium II / Pentium III	66/100/133	RDRAM	2.2	4 x
840	1999	Hub	Pentium III Xeon	66/100/133	RDRAM	2.2	4 x

从表 5-3 中可以看出,这一时期的芯片组有两种结构,即南北桥结构和 Hub 结构。两种结构的主要区别是连接南桥芯片和北桥芯片之间的通道不同。在南北桥结构中,南北桥之间通过 PCI 总线进行通信,PCI 总线由北桥控制,而南桥控制的所有设备,必须和整个 PCI 总线共享总线带宽。当时 PCI 总线的数据线宽度为 32 位,工作频率为 33 MHz,所以总带宽为 133 MB/s。这样 PCI 总线就容易成为数据传输的瓶颈。典型的 Intel 440 系列芯片组采用的是南北桥结构,如图 5-8 所示。

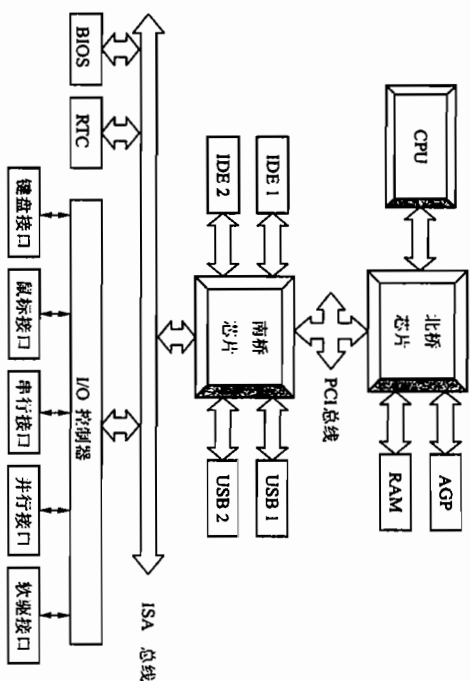


图 5-8 南北桥结构的芯片组

为了让主板芯片的架构更加明确, Intel 在开发 18xx 芯片组的时候, 提出比较正式的主机板芯片组的结构——IHA (Intel Hub Architecture), 简称 Hub 结构, 如图 5-9 所示。

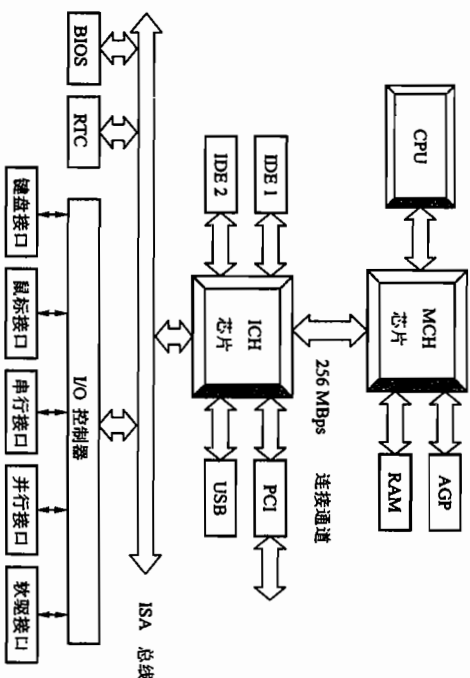


图 5-9 Hub 结构的芯片组

在 Hub 结构中, 两块芯片不是通过 PCI 总线进行连接, 而是利用能提供两倍于 PCI 总线带宽的专用通道进行连接。这样, 每个设备, 包括 PCI 总线, 都可以与 CPU 直接通信, 而不需要通

过 PCI 总线。在这一结构中, 计算机的各子系统都直接和芯片组相连, 显示卡、显示缓存、声卡、硬盘、内存、USB 设备等都有自己的通道, 整个系统呈放射状结构。而以前的主板芯片组的设计体系中, 各种设备都是通过 PCI 总线连接到主板芯片组, 这种连接方式产生了带宽瓶颈问题, 这一问题也是导致显示卡专用的 AGP 总线出现的直接原因, 810 芯片组的独特结构很好地解决了这一问题。

从表 5-3 中可以看出, Intel 早期芯片组支持的前端总线频率, 从 66 MHz、100 MHz, 最后发展到 133 MHz。存储器的类型也经历了较大的变化, 从 Intel 430 芯片组系列支持的 FRM DRAM (Fast Page Mode DRAM, 快速页模式动态随机存储器)、EDO DRAM (Extended Data Out DRAM, 扩展数据输出动态随机存储器), 到当时比较先进的 SDRAM (Synchronous DRAM, 同步动态随机存储器)。Intel 820 芯片组系列开始支持 RDRAM (Rambus DRAM)。Rambus 公司研制的新型存储器得到了 Intel 的支持, 但由于 RDRAM 的价格较高, 而且性能不及 SDRAM 的下一代产品 DDR SDRAM (Double Data Rate SDRAM, 双倍数据速率同步动态存储器), 没有得到市场的认可。

5.2.3 用于 Pentium 4 的 8xx 系列芯片组

用于 Pentium 4 的 8xx 系列芯片组的大部分技术参数和 Pentium 4 之前的 8xx 芯片组差别不大, 如表 5-4 所示。由于 Pentium 4 CPU 的前端总线带宽有较大的提高, 前端总线的工作频率为 400 MHz 以上。在 Intel 850 芯片组之后, Intel 公司放弃了对 RDRAM 的支持, 转向支持 SDRAM 的下一代存储器 DDR SDRAM。DDR SDRAM 的性能超过了 RDRAM, 而且保持较低的成本。在 Intel 865/875 芯片组开始支持双通道的 DDR SDRAM, 提高了存储器通道的传输带宽。与此同时, AGP 的版本由原来的 AGP 2.0 升级为 AGP 3.0, 即由 AGP 4 × 升级为 AGP 8 ×。USB 接口开始普及使用 USB 2.0, 将 USB 的传输速率由 USB 1.1 的 12 Mbps 提高到 480 Mbps。作为一个外部接口, 这样的高速率是空前的。同时, Intel 865/875 开始支持 SATA, 作为外存储器的新型接口, 它将逐步取代传统的 ATA-100/133 接口。初级的 SATA 的传输带宽 150 Mbps, 但由于 SATA 采用的新技术, 使得 SATA 在传输速率方面有较大的发展空间。

表 5-4 用于 Pentium 4 的 Intel 8 × × 系列芯片组的主要技术指标

芯片组	845	850	848	865	875
前端总线频率	400 MHz	400 MHz	400/533/800 MHz		
前端总线带宽	3200 MBps	3200 MBps	3200/4200/6400 MBps		
存储器	DDR 200/266	双通道 RDRAM	DDR SDRAM 266/333/400	双通道 DDR SDRAM 266/333/400	
PCI	6 个 PCI 2.2				
AGP	AGP 4 ×		AGP 8 ×		
USB	4 个 USB 1.1		8 个 USB 2.0		
外存储器接口	2 个 ATA-100		2 个 ATA-100 2 个 SATA 150		
音频	AC'97 6 通道				

Intel 865 芯片组的结构如图 5-10 所示。

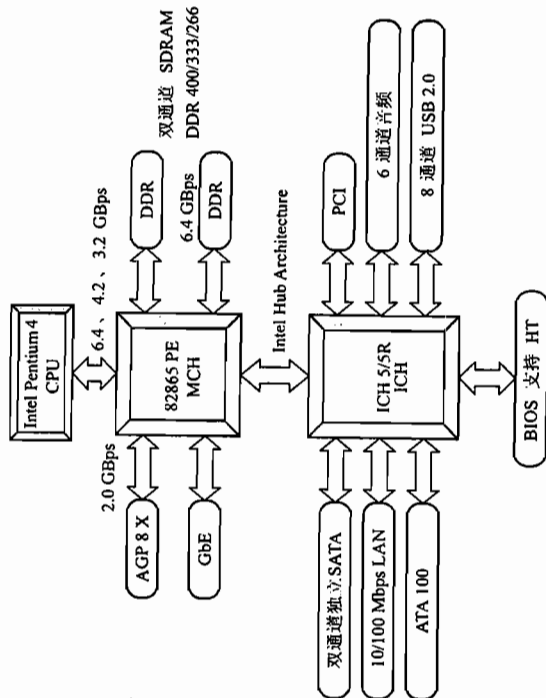


图 5-10 Intel 865 芯片组的结构

在 Intel 865 芯片组中,原来的北桥芯片被称为 MCH (Memory Controller Hub),南桥芯片被称为 ICH (I/O Controller Hub)。配合 865 PE 使用的 ICH 的型号是 ICH5 或 ICH5R,其中 ICH5R 支持 RAID。

5.2.4 Intel 9xx 系列芯片组

从 CPU 进入 Pentium 4 时期,芯片组的发展基本上可以分为两个阶段,第一个阶段的芯片组包括 Intel 845、848、850、865 和 875 等,这一阶段对应的 CPU 主要是 Pentium 4 的初级和中级产品。第二个阶段的芯片组主要有 Intel 910、915、925、945、955、965 和 975 等,对应的 CPU 主要是 Pentium 4 的后期产品和 Pentium D/EE。从功能上分析,8xx 系列和 9xx 系列芯片组的区别主要体现在 4 个方面,如表 5-5 所示。

表 5-5 Intel 8xx 和 9xx 芯片组的主要差别

	CPU	存储器	总线	音频
8xx 系列芯片组	Socket 478	DDR SDRAM	PCI	AC'97 96 kHz/20 b
9xx 系列芯片组	LGA 775	DDR2 SDRAM	PCI, PCI Express	HD Audio 192 kHz/24 b

9xx 系列芯片组与之前的 8xx 系列芯片组相比有多项改进,主要包括支持 LGA 775 封装的 CPU,支持 PCI Express 总线,支持双通道 DDR2 SDRAM 存储器,同时提高了音频的质量。MCH (南桥芯片) 和 ICH (北桥芯片) 的连接通道采用新的 DMI (Direct Media Interface),提供双向 2 GB/s 的传输速率。

1. Intel 915/925 系列芯片组

Intel 的代号分别为 Alderwood 和 Grantsdale 系列芯片组。这两款芯片组的出现,象征着近十年来计算机硬件平台的最大的换代工程: LGA 775 接口的 CPU 插座、双通道 DDR2 内存技术、PCI Express 显卡接口等,如图 5-11 所示。PCI Express 总线标准将取代使用超过 10 年的 PCI 总线,计算机技术进入了一个新的时代。

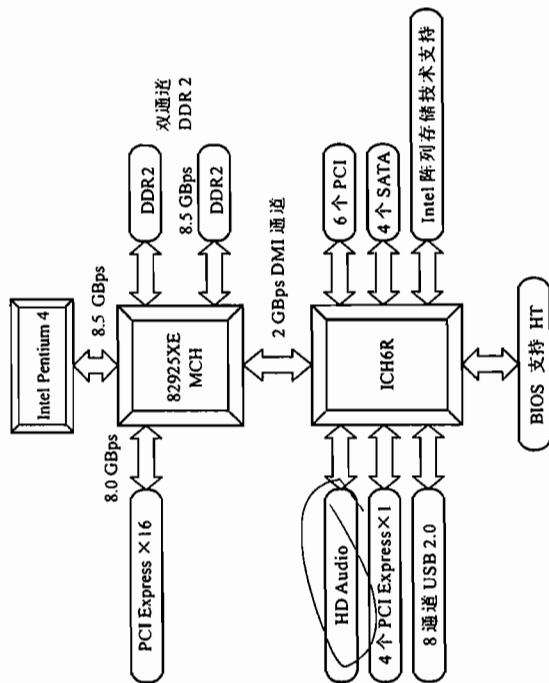


图 5-11 Intel 925 芯片组的结构

915/925 系列芯片组可以被认为分别是现有的 865/875 系列芯片组的升级版本。因此发布的 915 芯片组如 865 系列芯片组一样,有 915-P 和 915-G 两款,再加上后来的 915-GV 芯片组,一共是 3 款。915-G 就是 915-P 芯片组的内置显卡型号,而 915-GV 与 915-G 的区别是 915-GV 省掉了 915-G 上面的 PCI Express x16 显卡接口,从而使得其价格更低廉。925 芯片组则相当于 875 芯片组的地位,只有 925-P 一种。Intel 915 与 925 系列芯片组的技术参数如表 5-6 所示。

表 5-6 Intel 915/925 芯片组的技术参数

	Intel 925X	Intel 915G	Intel 915P	Intel 915GV
代号	Alderwood	Grandsdale-G	Grandsdale-P	Grandsdale-GV
处理器	Pentium 4 (Prescott) 和 Celeron LGA 775 / Socket 478			
系统总线频率	800 MHz			
超线程技术	支持			
800 MHz 时	DDR2-533/400	DDR2-533/400、DDR 400	DDR2-533/400、DDR 400	DDR2-533/400、DDR 400
533 MHz 时		DDR 400/333	DDR 400/333	DDR 400/333
ECC 内存校验	支持	不支持		
外接显卡	外接显卡	外接/集成显卡	外接显卡	外接/集成显卡
外置显卡接口	PCI Express x16			
内置显卡		Intel Extreme Graphics 3		Intel Extreme Graphics 3

在 CPU 支持方面,由于 Intel 同时推出了 LGA 775 接口的全新的 Prescott 核心的处理器,使得这两款芯片组均支持 800 MHz 前端总线的 LGA 775 接口的 Intel 处理器。但 925X 只支持 LGA 775 接口的 Pentium 4 处理器,并不支持 Socket 478 接口的 Pentium 4 与 Celeron 处理器;而 915 系列芯片组仍然支持 Socket 478 接口的处理器,所以我们会见到 Socket 478 接口的 915 主板。作为 Intel 的新一代芯片组,这两款产品理所当然地支持 Intel 的重要专利技术——超线程技术。LGA 775 插槽与现在的 Socket 插槽有很大的不同,CPU 采用了顶盖固定方式,可以更稳妥地固定 CPU 在插槽上的位置。

在内存支持上面,915 与 925 芯片组都支持双通道 DDR2-533 内存,可以提供 8.5 GBps 的带宽,但这两款芯片组还是较大区别的。作为高端平台芯片组的 925X 系列芯片组仅支持 DDR2 内存,而且支持内存的 ECC 校验功能;而作为中低端平台芯片组的 915 系列芯片组则同时支持 DDR 和 DDR2 内存,并且不支持内存的 ECC 校验功能。这样做的目的是为了确保持有平台向 PCI Express 顺利过渡,减轻用户负担。

DDR 和 DDR2 内存外形尺寸完全相同,但 DDR2 的金手指上采用了 240 引脚的边缘连接线,而 DDR 只有 184 个引脚。如果仔细观察,会发现两者边缘处的凹口位置也不一样。不一样的凹口位置可以防止将 DDR2 误插进 DDR 内存插槽中。DDR2 的工作电压为 1.8 V,因此 DDR 插槽的 2.5 V 电压显然不能让 DDR2 正常工作。

915/925 系列芯片组的问世带来的一个重要的改变就是 PCI Express 插槽的出现。这两款芯片组都支持一个 PCI Express x16 显卡插槽,另外还支持 4 个 PCI Express x1 插槽。

在硬盘的支持方面,这两款芯片组仍然保留了已经使用多年的标准 ATA 接口,但是只保留了一个 Ultra ATA 接口,仅支持两个 ATA 设备。而对于现在越来越流行的 SATA 接口,这两款芯片组都支持 4 个 SATA 接口,是 865/875 芯片组提供的 SATA 接口的两倍。915/925 系列芯片组支持的 SATA 仍然是 SATA-150 标准,并不支持更高端的 SATA-300 标准。

SATA 在问世之初,就与 RAID 功能紧密地联系在一起了。当芯片组发展到 915/925 的时候,RAID 的功能也得到了进一步的开发。通过 Intel 的矩阵存储(Matrix Storage)技术,这两款芯片组除了支持传统的 RAID 0 与 RAID 1 这两种磁盘阵列以外,还支持全新的类似于传统的 RAID 0 + 1 的矩阵式 RAID。但是在 RAID 配置方面,仍然不支持 ATA 与 SATA 硬盘混合建立磁盘阵列的方式。

音频系统的改进是 915/925 芯片组相对于 865/875 芯片组的又一大改进。915/925 的音频系统名为 Azalia,是一种高保真的音频解决方案。ICH6 南桥具备“Intel High Definition Audio”技术,使得 915/925 芯片组可以直接提供杜比 7.1 声道输出。此外,Azalia 的音频功能还支持 DVD-Audio、96 kHz/24 b 多声道和 192 kHz/24 b 双声道的音频输出,用户以后完全可以在 PC 上利用集成声卡享受到影院般的音响效果。

915-G/915-GV 芯片组则搭配了 Intel 的第三代图形处理器核心 Intel Extreme Graphics 3 (Intel GMA 900)。915 集成的显卡 Intel Extreme Graphics 3 使用一个 PCI Express x1 接口,这是 Intel 首次实现了在集成图形处理器核心直接支持 DX9 与 OpenGL 1.4。当与 DDR2-533 内存配合使用时,与现有 Intel 865G 芯片组相比,915 芯片组有较大的性能提升。除性能方面的改进以外,集成的图形引擎还支持双显示功能,但是大部分主板将不会直接提供两个图形输出接口。

北桥芯片集成了 875 主板的 CSA 通信架构,为用户提供更快的网络连接速度。CSA 技术是为千兆网卡所特别设计,为主板上的物理层网络通信设备提供了直接和 MCH 相连的通道,其通道达到了 266 MB/s 的带宽,完全满足千兆网卡的带宽要求。另外,CSA 技术大大减少了 CPU 的占用率,还可以更好地管理突发的大容量数据等。相对于以往的 32 位 PCI 插槽来说,CSA 具有带宽大、不占插槽和不占用其他设备带宽的优点,在搭配上千兆网卡芯片后就能实现千兆网络。

915/925 的又一改变是引入了全新的 ICH6 南桥芯片。ICH6 南桥芯片一共有 4 款不同的型号:ICH6、ICH6R、ICH6W、ICH6RW。末尾的“R”代表具有 RAID 功能,“W”代表具有无线局域网 (IEEE 802.11b/g) 功能,“RW”则代表同时具有以上两种功能。南桥芯片 ICH6 能够支持多达 4 个 PCI Express x1 接口,增加了两个 SATA 接口,一共提供 4 个 SATA 接口。Ultra ATA 接口则被缩减为一个通道。ICH6 南桥芯片的技术参数如表 5-7 所示。

表 5-7 Intel ICH6 南桥芯片技术参数

	ICH6	ICH6R	ICH6W	ICH6RW
PCI 总线	6 个			
SATA-150	4 个端口			
ATA-100	1 个通道			
USB 2.0	8 个			
千兆网	支持,占用一个 PCI Express x1			
4+1 Audio	支持			
内置 Wi-Fi 支持		支持	支持	支持
RAID (SATA)		支持		支持

2. 945/955 系列芯片组

915/925 和 945/955 这两代芯片组之间的差别并不是很大。不过 945/955 芯片组还是有其

优势所在,945 系列芯片组的开发代号为 Lakeport,是 915 系列的升级版本,分为外置显示核心的 945P 和简化的 945PL 以及整合 GMA950 显示核心的 945G 三种。

945/955 芯片组的推出就是为了支持 Intel 新推出的双核心处理器。945 和 955 芯片组之间的主要区别是对双核心处理器的支持有所不同。945 芯片组只能支持 Pentium D 双核系列处理器,并不能支持规格更高的 Pentium EE 系列处理器,而 955 芯片组则没有这方面的限制。

945/955 芯片组都支持双通道存储器,频率最高达到 667 MHz,其中 955X 芯片组支持 ECC 内存,最高容量可达到 8 GB,而 945 最多支持 4 GB 的内存。此外,955X 芯片组还有一个值得关注的优点,那就是提供了一种叫做 Turbo Mode 的内存模式。其实早就在 865/875 芯片组时代,Intel 就提供了一种内存加速技术,导致 875 芯片组的价格升高。后来各大主板制造商通过改造,让 865 芯片组也支持该技术,875 芯片组似乎就变成鸡肋了。945/955 芯片组都支持 PCI Express x16 显示卡接口,其中 945G 芯片组还集成了 GMA950 显示核心。945 和 955 芯片组的技术参数如表 5-8 所示。

表 5-8 Intel 945/955 芯片组参数表

芯片组	955X	945P	945G	945PL
处理器	Pentium EE Pentium D Pentium 4	Pentium D Pentium 4 Celeron D	Pentium D Pentium 4 Celeron D	Pentium D Pentium 4 Celeron D
前端总线频率 (MHz)	1066/800	1066/800/533	1066/800/533	800/533
内存规格	双通道 DDR2			
内存总线频率 (MHz)	667/533	667/533/400	667/533/400	533/400
内存容量 (GB)	8	4	4	2
错误校验	ECC	No	No	No
图形接口	PCI Express x16			
南桥搭配	ICH7, ICH7R			
外存接口	4 × SATA2 1 × ATA-100			
网络	千兆以太网			
音频	HD Audio, 24 位			

在国内特别值得一提的是 945PL 芯片组。由于国内对支持 DDR2 内存、SATA2 设备的低价 LGA 775 主板的需求比较大,所以 Intel 推出了面向低端市场的 945PL 芯片组。为了拉近 945PL 与 915 芯片组之间的价格差距,Intel 省去了其中的几个高端功能。下面我们来看看 945PL 芯片组的具体细节。

实际上,945PL 属于 945P 芯片组的简化版本,两者的主要区别如下:

• 945PL 最高只支持 800 MHz 的系统总线。

• 945PL 最高只支持 DDR2 - 533 内存,省去了对 DDR2-667 内存的支持。

• 945PL 最大只支持 2 GB 的内存。

可以看出,这 3 项改动都不会对 945PL 芯片组的性能造成很大影响,这种细微的规格变化也并没有影响 945PL 成为一款如同当年 Intel 845 系列芯片组一样性价比很高的产品。低价、实用是 945PL 主板最大的优势。

虽然 945PL 芯片组最高只能支持 800 MHz 的系统总线,但根据当时处理器的价格,显然 800 MHz/533 MHz 前端总线的处理器系列更适合普通用户,毕竟仅 Pentium EE 处理器才支持 1066 MHz,但其价格却昂贵得多。

在内存支持方面,945PL 不支持 DDR2-667,内存容量方面也只有 2 GB,不过 945PL 的功能完全可以满足大多数用户的需要。要知道 DDR2-533 是 Intel 平台当时的标准内存频率,对于支持 800 MHz 系统总线的处理器而言,双通道 DDR2-533 已经能够满足需要,DDR2-667 内存并不一定会给系统性能带来明显的提升,而且在以后的一二年中 DDR2-533 也不会过时。至于内存容量,1 GB 内存目前足够满足绝大多数用户的要求。从目前软件发展趋势来看,在两年内主流 PC 机的内存容量也不会超过 2 GB。

此外,相对于 915 和第三方芯片组,945PL 在传统优势项目(比如音视频解码、编码)方面也保持领先,同时支持 PCI Express 和 SATA2 等新特性,适应性强,具有升级的潜力。

从上面几点分析我们可以发现,945PL 的这几点简化之处对于大多数用户而言是“无疼痛”的,正是因为这种精简,造就了 945PL 的超值。

5.3 主机板

主机板(mainboard),简称主板,有时也称为系统板(systemboard)或母板(motherboard),它安装在机箱内,是微型计算机最基本的也是最重要的部件之一。图 5-17 是一块典型的主机板外观图。主机板一般为矩形电路板,上面安装了组成计算机的主要电路系统,除了 CPU 和内存条是可拆卸的部件之外,还固定安装有芯片组(一般是两块芯片)、BIOS 芯片以及一些专用的 I/O 控制芯片。

5.3.1 主机板的构成

除了 CPU 和存储器插座外,主机板上有很多用于连接其他设备和部件的总线或接口。目前的主机板上主要有:

- 总线扩展槽:一个 PCI Express x16,多个 PCI Express x1,多个 PCI。
- AGP 显卡插座:一般是 AGP 4 × 或 AGP 8 ×。
- USB 接口:一般有 4 ~ 8 个,包括在主机板上的接口和引到前面板上的接口。
- ATA 接口:一个或两个 ATA - 100 接口(如图 5-12 所示)。
- SATA 接口:2 ~ 6 个 SATA,可能是 SATA 或 SATA2(如图 5-13 所示)。

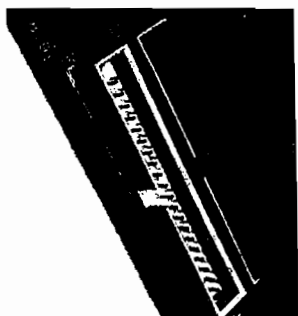


图 5-12 ATA 接口



图 5-13 SATA 接口



图 5-14 ATX 标准电源接口

- 电源接口: ATX 标准电源接口(如图 5-14 所示)。
- CPU 电源接口: 一个 4 芯的接口。
- 软驱接口: 大多数主板暂时还保留。
- 并行接口: 在后面的 25 针 D 型接口。
- 串行接口: 在后面的 9 针 D 型接口(如图 5-15 所示)。
- 网络接口: 集成网卡的主板上的 RJ-45 标准的局域网接口。
- 显示器接口: 集成显卡的主板上的 15 针 D 型 VGA 接口。

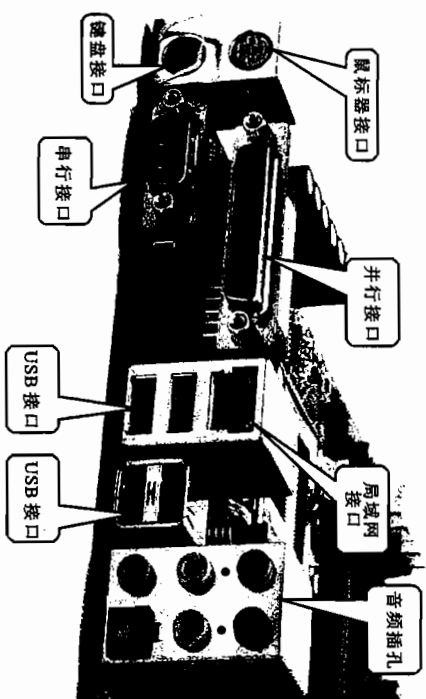


图 5-15 ATX 主机板后面的接口

- 音频插孔: 多个直径 2.5 mm 的音频输入/输出插孔。
- 键盘接口: 圆形的 PS/2 接口。
- 鼠标器接口: 圆形的 PS/2 接口。
- MIDI/游戏杆插座等。

5.3.2 主板标准

1. 印刷电路板

主板所用的印刷电路板(Printed Circuit Board, PCB)是由几层树脂材料黏合在一起的,内部采用铜箔走线。一般的主板印刷电路板分为 4 层,最上和最下的两层是信号层,中间两层是接地层和电源层,如图 5-16(a)所示。将接地层和电源层放在中间是为了便于对信号线进行修正。好的主板印刷电路板可达到 6 层,这是由于信号线必须相隔足够远的距离,以减小电磁干扰。6 层的印刷电路板可能有 3 个或 4 个信号层、一个接地层以及一个或两个电源层,如图 5-16(b)所示,以提供足够的电力供应。为使系统正常工作,信号迹线的布局与长度是至关重要的因素,它的设计宗旨是尽量避免由于信号线之间的干扰,造成信号失真,要求在相邻的两条迹线之间,留出足够大的间距。必须限制有些迹线的最大长度,以确保信号的最小衰减等。



图 5-16 印刷电路板

2. 常见主板结构规范

主板结构规范是指主板上元器件的布局排列方式。不同的主板类型通常要求不同的机箱与之相配套。各主板结构规范之间的差别包括尺寸大小和形状、元器件的放置位置、电源供应器等。常见的主板结构规范主要有 AT、Baby AT、LPX、ATX、Mini ATX、Micro ATX、NLX、Flex ATX 等。其中 AT、Baby AT 和 LPX 属于较早期的结构标准,现在已经很少见到了。下面着重介绍现在正在使用的几种主要规范。

关于主板结构的详细说明可以从网站 www.formfactors.org 找到。

(1) ATX/Mini-ATX 结构

ATX(AT eXtend)结构是 Intel 公司于 1995 年 7 月提出的,如图 5-17 所示。它是一种创新的机构设计,能够更好地支持电源管理。ATX 是将 Baby AT 与 LPX 两种结构规范的优点整合在一起,它在 Baby AT 的基础上逆时针旋转了 90 度,直接提供串行接口、并行接口、PS/2 键盘接口、PS/2 鼠标器接口等。

另外,在主板设计上,由于横向宽度增加,内存插槽可以紧靠最右边的 I/O 槽,使 CPU 插槽位置远离扩充槽,安放在内存插槽的右侧或下部,以允许插入长卡而不会挡到 CPU,内存条更换

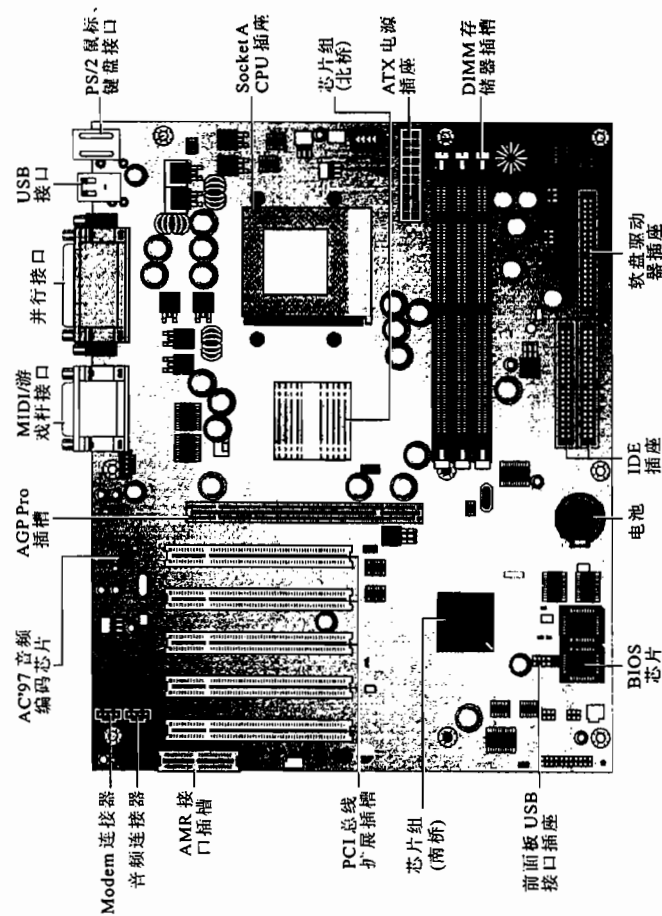


图 5-17 ATX 主板

也更加方便。软、硬盘连接插座并排放置,可缩短排线长度,方便安装。电源位于 CPU 插槽的右侧,利用电源单边托架风扇,可以直接给 CPU 及机箱内元件散热。外设接口面板插座直接集成在主板上,减小了电磁辐射,节省空间。

ATX 结构的底板尺寸是 12 英寸 \times 9.6 英寸,Mini ATX 结构的底板尺寸是 11.2 英寸 \times 8.2 英寸。两种结构的机箱是相同的,如图 5-18 所示。

ATX 结构主要有下列好处:

- 方便硬件的安装与拆卸。
- 支持现在各种多媒体卡与未来的新型设备。
- 降低系统的整体造价。
- 改善系统通风设计。

(2) Micro ATX 结构

Micro ATX 是依据 ATX 规格改进而成的一种新标准,并且已成为广泛使用的标准。Micro ATX 规格被推出的最主要目的是为了降低个人计算机系统的整体成本与能耗,让使用者可以购买到更便宜且更小型化的个人计算机。Micro ATX 结构规范的主要特性:支持目前的 CPU 技术、更小的主板尺寸、更小的电源供应器,减小主板与电源供应器尺寸的直接效果是计算机系统

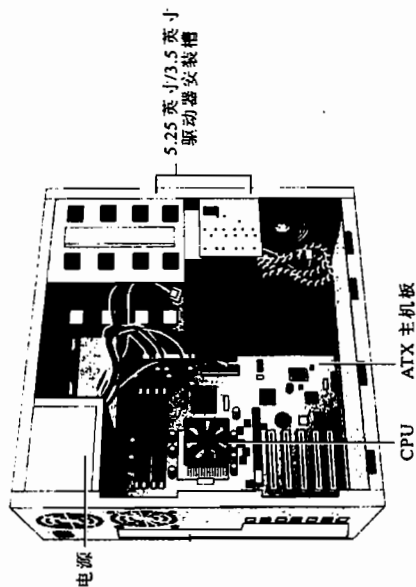


图 5-18 ATX 结构的机箱

的成本下降。然而,虽然减小主板的尺寸可以降低成本,但是主板上可以使用的 I/O 扩展槽也相对的减少了,Micro ATX 支持最多到 4 个扩展槽,这些扩展槽可以是 ISA、PCI 或 ACP 等各种规格的组合,视主板制造厂商而定。Micro ATX 规范的主板也可以用于 ATX 机箱与电源供应器,但是特别为 Micro ATX 所设计的机箱及电源供应器却具有较便宜的价格与较小的尺寸。为了完全使用到 Micro ATX 规格所带来的好处,必须使用一种叫做 SFX 1.1 的新式电源供应器或使用符合标准 ATX 2.01 规格的电源供应器。

Micro ATX 结构的主板尺寸是 9.6 英寸 \times 9.6 英寸

(3) Flex ATX 结构

1993 年 3 月,Intel 继 Micro ATX 之后又公布了 Flex ATX,为主板增加了一个更小的结构规范。Flex ATX 采用更小的设计是为了增加 PC 设计中的灵活性,尤其是廉价的、小型的、面向消费者的应用系统。

Flex ATX 定义了尺寸仅为 9.0 英寸 \times 7.5 英寸 (229 mm \times 191 mm) 的底板,是 ATX 主板族系中最小的。除了较小的尺寸之外,Flex ATX 外形同 Micro ATX 相比,另一个最大区别是 Flex ATX 只支持插座处理器 (Socketed Processor)。这意味着 Flex ATX 主板没有用于 Pentium II/III 或 Athlon 处理器盒式版本的 Slot 1、Slot 2 或 Slot A 类型的连接器。

除了尺寸较小以及要求插座处理器外,Flex ATX 的其余部分同标准 ATX 兼容,使用同一系列的螺丝孔设置及相同的 I/O 与电源连接器规范。



1. 名词解释

FSB, ISA, EISA, MAC, VESA, PCI Express, PCI, ACP, IHA, MCH, ICH, SATA, HD Audio, AC'97, DVB, RAID, ATX, Flex ATX, PCB.

2. 说明总线的主要特征。
 3. 总线有哪几种类型?
 4. 在现代微型计算机系统中, I/O 总线和系统总线的作用有什么不同?
 5. 总线可以分为哪几部分? 在总线传输过程中这几部分各起什么作用?
 6. 总线带宽和总线时钟频率、总线位宽是什么关系? 怎样理解总线带宽和实际传输速率之间的差别?
 7. PCI 总线在什么时代出现的? 微型计算机采用 PCI 总线的原因是什么?
 8. 简述 PCI 的基本技术指标。
 9. PCI 总线和其之前的 ISA 相比较采用了一些新技术, 主要有哪些新技术?
 10. 简述 PCI 总线的总线仲裁。
 11. 简述 PCI 的版本。
 12. 微型计算机采用 ACP 的原因是什么? 简述 ACP 的性能。
 13. 简述 PCI Express 的基本指标。
 14. 说明 PCI Express 总线在传输中是怎样同步的。
 15. 简述芯片组在现代微型计算机系统中的地位和作用。
 16. 北桥芯片(或 MCH)确定了主机系统的哪几个重要的速度指标?
 17. 简述 Intel 芯片组中南北桥结构和 Hub 结构的主要区别。
 18. 在 CPU 和存储器支持方面, Intel 915/925 与 Intel 945/955 芯片组有什么不同?
 19. 简述 Intel ICH5 和 ICH6 的功能方面的区别。
 20. 简述 Mini ATX、Micro ATX 和 Flex ATX 三种主机的区别。
2. 使用软件“优化大师”测试一台计算机的以下参数:
芯片组型号、主板型号、主板制造商、PCI 总线位宽和插槽数目、主板提供的接口。
3. 使用软件 CPU-Z(1.37 版)测试一台计算机的以下参数:
芯片组北桥(或 MCH)和南桥(或 ICH)型号、主板型号、主板制造商、BIOS 厂商和版本。

实 验

接 口 第6章

PC 机中的接口按其用途可分为通用接口和专用接口两类。本章主要介绍 PC 机中常用的通用接口, 专用接口在其所连接的外部设备的相关内容中介绍。在介绍通用接口之前, 首先介绍接口的基本知识, 然后依次介绍串行接口、并行接口、USB 接口、IEEE-1394 接口、IDE 接口、SCSI 接口和 InDA 接口等。

6.1 接口的基本知识

本节介绍 PC 机中接口的基本知识, 包括接口与主机系统及外设的关系、接口的构成和作用、接口的操作以及 PC 机中接口的 I/O 地址分配等。

6.1.1 接口的构成

计算机系统所配置的外部设备, 类型繁多, 数量不同。它们不仅在工作速度上与中央处理器差别极大, 而且在数据表示的形式上与计算机内部形式也不一致。因此, 要实现外部设备与主机之间的连接和信息交换, 必须经过一个数据转换和传输的设备。这种设备叫做 I/O 接口(input/output interface), 有的叫做 I/O 适配器(input/output adapter)或适配卡, 例如, 键盘适配卡、打印机适配卡、CRT 适配卡、磁盘适配卡等。

主机和外部设备之间的连接关系如图 6-1 所示。主机和任何外部设备的连接都是通过接口电路实现的。接口电路实现计算机总线线和外部设备之间物理连接, 是接口的硬件部分; 接口的另外一部分是软件部分, 接口电路的工作是在接口软件的控制下完成的。

外部设备的存在形式可以不同, 主要分为 3 种, 第一种是计算机主机箱之外的独立设备, 如打印机等; 第二种是安装在计算机主机箱内部的独立设备, 如硬盘等; 第三种是集成在主板上的电路单元, 如实时时钟电路等。

接口电路主要包括以下几个组成部分。

5. 信号转换驱动电路

有些外部设备的信号电平与主机不同,需进行电平转换、信号形式也可能不同,如有些外部设备送出的是模拟信号,这样就要进行模数(A/D)或数模(D/A)转换。

传送的通路可能有许多连接电路,而且有一定的传输距离,因此要求I/O接口能提供足够的驱动能力。同时,接口自身的负荷也应当限制在一定的水平上。

如上所述,I/O接口一边是面向主机控制,另一边是面向外部设备,不同的控制和不同的外部设备需要有不同的I/O接口。它们之间的功能划分是十分不同的,有的功能放在I/O适配器上,有的放在I/O设备上。一般情况下,联系紧密和关系复杂的功能应尽可能地安排在一起;联系不多,关系简单的功能可以分开。如果把设备控制器放在I/O适配器上,也就很难区分哪些是属于I/O接口,哪些是属于设备控制器了。例如,打印机适配器只完成接口功能,常作为通用并行接口,而具体的打印控制(设备控制器),则放在打印机中。在CRT适配器上,则包括接口和CRT控制器两部分功能,用来把显示数据转换为视频信号,这时外部设备只剩下CRT本身了。磁盘驱动器适配器中,则包含了磁盘控制器的大部分功能(磁盘控制程序、数据编码和译码以及错误检验等)。

一个完整的I/O接口,不仅包含一些硬件电路,还包含相应的驱动程序。这些程序的存放方式可能有多重,可以存放在接口电路的ROM芯片中,这种方式一般是总线扩展板方式的接口电路;另一种方式是集中存放在系统板的ROM芯片中,这些接口程序被称为基本I/O系统,即BIOS,应用程序可以通过调用BIOS来操作I/O接口,一般在主机板上的各种接口电路的接口程序就是这样存放的,这样I/O接口通过BIOS程序可以提供一个易于标准化的软件接口;最后一种是在存放在磁盘上,可以通过操作系统或厂商提供的驱动程序光盘获得。

6.1.2 接口信号

接口信号的编/译码是为了满足接口信号而设立的电路机制,不同的接口对信号波形的要求不同,尤其是串行方式的接口,如串行接口、USB接口、红外接口等,将数据在发送端转换成要求的波形,在接收端再转换回来。波形的编/译码是在接口电路硬件中完成的。

1. 单极性信号与双极性信号

单极性信号是指数字信号的0和1两种状态中有一个状态是零电平,而另一个状态是一个固定的电平。如果逻辑0对应零电平,逻辑1对应一个固定的正电平,就是单极性正逻辑信号。

双极性信号是指数字信号的0和1两种状态分别对应两个固定的非零电平,多数双极性信号的两个电平是一正一负而且绝对值相等。双极性信号的两个电平差可能比单极性信号更大,所以可以提高信号在传输过程中的抗噪声干扰能力。

最常见串行接口RS-232C就是双极性信号方式,它的逻辑0和逻辑1信号电平分别是+3V~-12V和-3V~-12V,在不同的系统中常常采用+3V、+5V、+12V和-3V、-5V、-12V三种电平。

2. 归零与不归零信号

不归零(Non Return to Zero)信号是指在一个位传输时间内,信号的电平是恒定不变的,而归零信号的有效电平时间只占位传输时间的一部分,如图6-2所示。

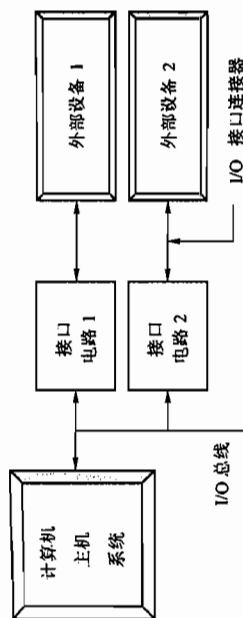


图6-1 主机、外部设备和I/O接口间的关系

1. 地址译码单元

一个计算机系统有多种或多种外部设备,如键盘、打印机、CRT、磁盘等,因此就有多个I/O接口。为了能够对各设备进行选择,必须给它们分配不同的地址码。需要指出的,在一个I/O接口中,有输入数据、输出数据,还有控制和状态信号。这样,一个I/O接口就需要有几个地址码,每个地址码要配一个寄存器。

每一个寄存器都需要不同的地址,地址译码单元的作用是识别特定的一个或几个地址。

2. 数据、控制和状态信息寄存器

接口电路中包括3种作用的寄存器:数据寄存器,用于输入/输出数据;状态寄存器,用于说明电路工作的状态,软件通过读状态寄存器掌握接口电路的工作状态;控制寄存器,用于实现对接口的控制,软件通过写控制寄存器完成对接口的控制。

主机可以把数据和状态信息送给外部设备,外部设备也可以把数据和状态信息送给主机。应当指出的,一般控制信息和状态信息都以数据形式交换,它们分别组成控制字和状态字。可以用I/O接口中某个寄存器的若干位表示主机的“启动”、“送数”等控制命令,另一个寄存器的若干位表示设备的“请求”、“设备忙”等工作状态。这样,主机就可以通过执行针对特定地址的输入/输出指令,完成对特定接口的操作。

3. 缓冲、暂存机制

主机和外部设备一般是按各自的节拍(时序)工作的。因此,为了协调它们之间的信息交换,通常需要缓冲或暂存,以满足各自的时序要求。尤其是现代计算机系统中,CPU和总线的速度大幅度提高,为了提高接口的传送效率,一般要求接口电路具备缓冲存储或暂存能力,有存储单元。

4. 数据转换

包括简单的格式转换和编码转换。若CPU字长16位,而外部设备按位串行传送,则I/O接口需进行串-并数据格式转换;若CPU字长32位,而外部设备数据格式为一个字节(即8位),这时则需进行组装或分解,即将4个字节拼成一个32位字或将一个32位字分解为4个字节。

编码/译码转换可以在软件中实现,也可以采用硬件实现方案。编码的目的是多种多样的,可以是压缩编码、纠错编码等,但本质上就是将数字信息按一定规则进行转换。

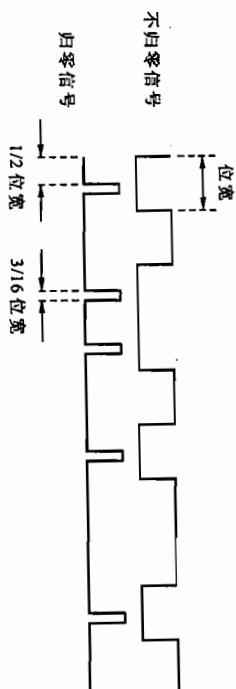


图 6-2 不归零归零信号波形

例如，IrDA 红外线接口的标准就规定使用归零信号驱动红外发送二极管，如图 6-2 下图所示。

3. 平衡与非平衡信号

非平衡信号是指所有的接口信号电平都是以地线作为参考点的。如果接口的每一个信号都具有自己的参考点，而且这个参考点是信号电平的反向信号，这就是平衡信号。

在 RS-485、RS-422、USB 等接口中采用的都是平衡信号。在平衡传输中，每一个信号需要两条接口线传输，如图 6-3 所示。

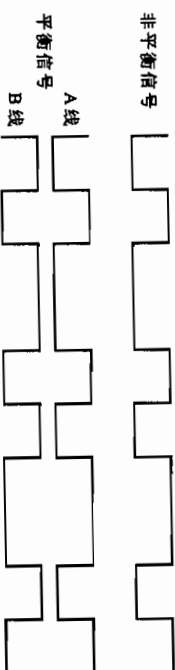


图 6-3 非平衡和平衡信号波形

和非平衡信号相比，平衡传输具有很好的抗干扰特性，适合在长距离或高速传输中使用，如 RS-422/RS-485 接口采用平衡传输，其传输距离比采用非平衡传输的 RS-232C 要远得多。在短距离传输中采用平衡传输则可以大大提高传输速度，如 USB 接口采用平衡传输，其传输速率远远高于采用非平衡传输的 RS-232C 等传统接口。

6.1.3 接口的分类

接口的分类方法是多种多样的，而且随着时代的不同，划分的方法可能不同。

1. 传输距离

从传输距离上划分，接口可以分为近距离、中距离和远距离三类。近距离是指一般设备内部连接和桌面范围的连接，连接电缆长 $1\text{ m} \sim 2\text{ m}$ ，如 USB 接口；中距离接口指一个房间内的连接，连接电缆最长为数千米，如 RS-232C 接口；远距离接口可以满足一个建筑内的连接，如 RS-

422 和 RS-485。在可以采用多种传输速率的接口中，传输的距离都和传输速率有关。

2. 信息位传输方式

从信息位的传输方式上划分，接口可以分为串行接口和并行接口两类。传统的概念中串行接口是慢速度接口，而并行接口是高速接口。但随着技术的发展，新型的高速接口逐步转向串行方式。

3. 定时方式

从信息传输的定时方式上划分，接口可以分为同步接口和异步接口。同步接口在传输信息的同时传送时钟信号，而异步接口则是从信息本身提取同步信息。

4. 传输媒介

从传输媒介上划分，接口可以分为有线接口和无线接口（包括红外线接口）。

5. 传输方式

从传输方式上划分，接口可以分为全双工、半双工接口。全双工接口可以同时进行双向的信息传输，如 RS-232C；半双工接口在某一时刻只能单向传输，如 USB 接口。

6. 物理连接

从物理连接关系上划分，接口可以是点对点连接、点对多点连接和多点连接三种。如 USB 接口在物理连接上只能是点对点方式；RS-485 可以实现点对多点和多点连接。

7. 标准

从标准的内容上划分，接口可以分为简单接口和复杂接口，这主要是指接口的标准规范中有没有规定高层协议，如通信规范、完善的差错控制方式、信息的编码和压缩方式以及一些现代接口的新特征。老式的接口如 RS-232C 属于简单接口，在 RS-232C 规范中基本上只规定了接口的机械规范、电气规范和功能规范，并没有规定高层协议。在使用这样的接口时必须在软件上考虑更多的问题，如差错控制等。复杂接口的标准中有很完善的规范，如 USB 接口。

6.1.4 接口的操作方式

对接口的操作是通过软件完成的，通过不同的软件控制方式实现对不同的接口电路的控制。评价一个接口软件的优劣，主要是看它占用计算机资源的多少，这些计算机资源包括 CPU 时间、存储器、中断和 DMA 通道等，在达到同样的接口传送速率的前提下，占用的计算机资源越少越好。

CPU 是计算机的核心部件，它不仅要承担数据的加工和运算，还要控制数据的输入和输出。可以通过对数据寄存器的读写来实现数据的输入和输出，也可以通过 DMA 方式完成数据的传送。由于计算机和外部设备是完全独立工作的，软件可以通过读取状态寄存器来掌握外部设备的状态变化，也可以通过中断方式，在外部设备发生状态变化时主动通知接口软件。

对于配备大量外部设备的计算机系统，还可以设置单独的计算机，专门用于 I/O 处理，这种方式普遍在中、大型计算机中采用。

本节主要叙述在计算机系统中常采用的 I/O 控制方式，包括查询控制方式、中断控制方式、直接存储器存取（DMA）方式和通道（通信处理机）方式 4 种，其中通道方式一般用于中、大型计

算机,就是使用一台专用的计算机作为 I/O 专用通信处理机,所有的外部设备的连接都通过通信处理机实现。

下面介绍 4 种接口控制方式的前 3 种。

1. 查询控制方式

在程序控制下进行信息传送,可分为无条件传送方式(亦称同步传送方式)与条件传送方式两种。无条件传送的方法是:在程序中的恰当位置直接插入 I/O 指令,当程序执行到这些指令时,外部设备马上与 CPU 进行数据交换。这种传送方式使用方便,所需的软、硬件都少。但它要求外部设备总是在数据交换前做好接收或发送数据的准备,因而只适用于操作时间为已知或变化十分缓慢的外部设备。当外部设备的操作时间未知或与 CPU 不同步时,可采用条件传送方式。这一方式的特点是:在执行 I/O 操作之前先用程序对外部设备的状态进行检测,只有当检测到所选择的外部设备已做好输入或输出状态信息后,才能开始执行 I/O 操作。因此该方式也可称为状态查询方式。

用查询方式进行数据交换的工作流程如图 6-4 所示。对输入设备而言,有效状态信息为 READY(输入数据准备好),对输出设备而言,有效状态信息为 BUSY(忙),均用 1 位二进制数表示。

查询方式的优点是简单,对硬件的要求低,编写软件也简单。由于状态的查询占用了大量的 CPU 时间,这种方法缺点是效率低,尤其是在 Windows 这样的多任务操作系统下,查询程序被分配在空闲的时间间隙上运行,没有办法实现及时的查询,很可能造成数据丢失。所以,在现代计算机中这种方式已经很少使用了。

2. 中断控制方式

查询方式的缺点是,每次输入输出一个数据,CPU 要检查一次外部设备状态。如果外部设备尚未准备就绪,程序便进入查询循环,许多 CPU 时间白白花费在状态查询中,浪费了宝贵的资源。因此,“中断”机制便引入到接口电路中,产生了由中断控制的 I/O 方式。

所谓“中断控制”,就是当 CPU 执行程序时,允许外部设备用“中断”信号中止 CPU 正在执行的程序。具体地说,当外部设备需要与 CPU 进行数据交换时,便由接口电路向 CPU 发出一个中断请求信号,待 CPU 响应这一中断请求后,便调用中断服务程序完成一个或多个字节的信息交换。这种方式不需要 CPU 主动查询,而是由接口电路主动通知 CPU,使得接口软件效率比较高。

中断控制方式的接口软件比较复杂,软件的代码包括两部分:主控程序部分和中断服务部分。主控程序部分设置变量和参数,包括输入输出缓冲区,并将中断服务的代码加载到内存,设置中断向量(将接口电路使用的中断号和这部分中断服务程序联系起来),这样在接口电路产生中断后,CPU 将根据中断号码,自动调用执行中断服务程序。中断服务程序和主控程序的联系

就是输入输出缓冲区。

因产生中断而进入中断服务程序后,程序首先查询产生中断的原因。一般产生中断的原因有多种,包括输入(接口电路收到了外部设备的数据)、输出(上次传送到接口电路的数据已经发送完毕)、异常等中断。再根据中断的原因,执行不同的处理代码,如图 6-5 所示。

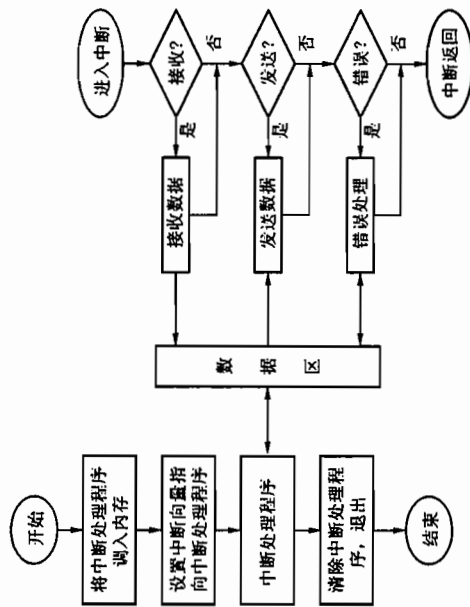


图 6-5 中断控制方式的软件实现

每产生一次中断,CPU 就会暂停原来执行的程序去调用中断服务程序,每次中断传送一个字节或字,有大量数据需要传送时将产生大量的、密集的中断。为了使得中断服务程序的效率更高,一般要在接口电路中设立缓冲存储机制,这样可以在每次中断传送一批数据,而不是一个字节或字,从而避免了产生大量的、密集的中断。

对速度较高的外部设备,设立缓冲存储机制是必需的,因为从中断的产生到中断处理可能需要一定的时间,如果没有缓冲存储就可能丢失信息,因此无缓冲存储机制的中断控制方式一般仅用于低速外部设备与 CPU 之间的信息交换。

3. DMA 方式

用带有缓冲存储机制的中断方式控制 I/O 传送时,每次中断需要传送一批数据,而这一批数据的传送是逐个单元(字节或字)完成的,传送是在主控程序的输入输出缓冲区和接口电路之间实现的,每个单元的传送都不是一条指令可以完成的,如输出至少需要 3 个步骤,先将一个单元取到 CPU 内部,再将这个单元送到接口电路,再更新输出缓冲区的指针。

采用直接存储器存取(Direct Memory Access, DMA)方式就是为了提高成批数据的传送效率,在高速 I/O 设备(如硬盘、光盘、声卡等)传送成批数据时,最好采用 DMA 方式。这种方式完全由硬件即 DMA 控制器(DMA Controller, DMAC)来完成 I/O 传送。数据交换不必经过 CPU 的寄存器,可直接在接口电路和存储器之间进行。

在 DMA 控制方式下,当某一外部设备需要输入或输出一批数据时,可向主板上的 DMA 控

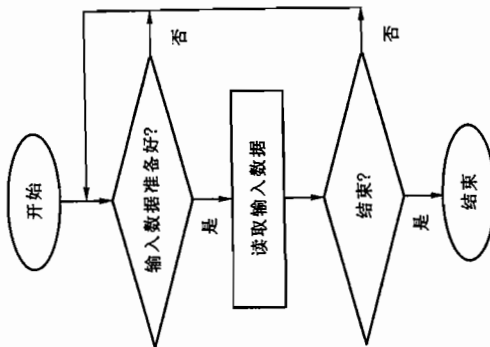


图 6-4 查询方式的读数据流程

制器发出请求。DMA 控制器接收到这一请求后,向 CPU 发出总线请求。若 CPU 响应 DMAC 的请求,就把总线使用权交给 DMA 控制器(因为 DMA 传送需要使用总线),此后数据传送不再通过 CPU,而是在 DMA 控制器操纵下直接在接口电路与存储器之间进行,当这批数据传送完毕后,DMA 控制器再向 CPU 发出结束请求,CPU 响应这一请求,立即收回总线使用权。

由此可见,采用 DMA 控制方式时,CPU 只需在数据传送开始前和结束时各响应一次中断,从而减轻了 CPU 的负担,但 DMA 控制器一般只能对一台或几台同类型的外部设备完成控制功能。建立一个 DMA 传送需要相应的硬件设备,故使用 DMA 传送少量的数据可能是不合算的。大多数的 DMA 传送都是用来移动大块的(至少几十个字节)数据。

如果要使用 DMA 控制方式,接口电路需要有相应的支持。

从查询方式到中断方式,再到 DMA 方式,是接口的控制方式从低级到高级的过程。针对一个特定的接口电路,并不是 3 种控制方式都可以选择的。如果使用中断控制方式,要求接口电路必须具备中断机制;同样,如果使用 DMA 控制方式,要求接口电路必须具备中断机制和 DMA 机制。

6.2 RS-232C

串行接口的特点是数据信息以串行方式逐位传送。在计算机中有很多接口都是串行方式的,如 USB 接口、SATA 接口、键盘接口和鼠标接口等。串行接口又可分为同步串行和异步串行两类。同步串行接口在连接线路中有时钟信号线,而异步串行接口没有时钟信号线。

按照习惯的叫法,串行接口是指 RS-232C 接口。本节介绍的就是 RS-232C,它属于异步串行接口,是 PC 机中常用的通用接口,可用于连接鼠标、调制解调器(Modem)等。所谓串行接口,就是所传送的数据是以串行(逐位)的方式传送的。发送时先将并行的字节转换成串行的位并逐位发送,接收时再将依次收到的数据位组装成字节。

6.2.1 RS-232C 串行接口标准

RS-232C 标准是美国电子工业协会(Electronic Industries Association, EIA)与贝尔公司一起开发并于 1969 年公布的通信协议。字母 RS 表示 Recommended Standard(推荐标准),232 是识别代号,C 是标准的版本号。

RS-232C 标准最初是为远程通信连接数据终端设备(Data Terminal Equipment, DTE)与数据通信设备(Data Communication Equipment, DCE)而制定的,但目前更广泛地应用于计算机与终端或外部设备之间的近距离连接。这个标准对串行通信接口的有关问题,如信号功能、电气特性和机械特性都做了比较明确的规定。由于通信接口与设备制造厂商都生产与 RS-232C 兼容的通信设备,因此它已成为微型计算机串行通信接口中广泛采用的标准接口。

1. 连接器

由于 RS-232C 标准只规定了采用一对物理连接器,但对连接器本身的物理特性没有任何定义,因而出现了 DB25、DB9 等各种类型的连接器,其引脚的定义也各不相同,使用时应特别注意。

在实际应用中,DB25 和 DB9 这两种连接器使用得较多。下面将这两种连接器做简单介绍。

早期的 PC 和 PC/XT 采用的是 DB25 型连接器。虽然 RS-232C 定义了 25 个引脚信号标准,但实际上进行异步串行通信时,只需 9 个电压信号、2 个数据信号(RXD、TXD)、6 个控制信号和 1 个信号地线。由于 PC 及 PC/XT 计算机除了支持 EIA 电压接口外还支持 20 mA 电流环接口,另需 4 个电流信号,故而采取 DB25 型连接器作为 DTE 与 DCE 之间通信电缆的连接。

DB9 型连接器,其信号引脚如图 6-6 所示。由于 AT 计算机串行接口取消了电流环接口,采用 DB9 型连接器。如果将配接 DB9 型连接器的 DTE 与配接 DB25 型连接器的 DCE 相连接,必须使用专门的电缆线或转接头。

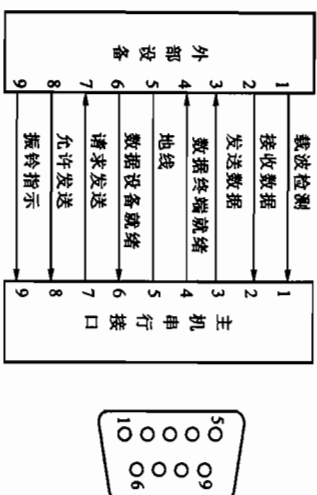


图 6-6 DB9 型连接器

2. 电缆长度

传输电缆的长度与传输线的线间电容有关。EIA 标准规定被驱动电路(终端)的电容(包括电缆连接电容)必须小于 2 500 pF。对于一个多芯电缆来说,每英尺(0.305 m)电容为 40 pF ~ 50 pF,所以满足电容特性的电缆长度最长为 50 英尺(15.24 m)。

如果电容特性不满足,则明显的反应是从 0 到 1 或从 1 到 0 的跳变时间将超过 RS-232C 标准规定的 4% 位码时间的最大允许值。超过 50 英尺所增加的电容和上述充电时间的差异会使接收端信号产生畸变,这些畸变会引起字符接收错误。如果在时钟频率有畸变或信号状态发生跳变时出现噪声,则更易出错。

RS-232C 标准允许的连接电缆不超过 50 英尺,但若能保证电缆总电容小于 2 500 pF,则电缆长度可超过限定值。同时,RS-232C 标准所允许的信号传输速率在 75 bps ~ 20 000 bps 的范围之内,而在传输距离稍长的实际应用中常被限制在 19 200 bps 以内。

3. RS-232C 信号电平

RS-232C 标准对信号的逻辑电平、最高数据传输速率和各种信号功能都做了规定。RS-232C 连接器任一引脚上的信号状态如表 6-1 所示。RS-232C 选择 -15 V ~ -3 V 和 +3 V ~ +15 V 这个范围而不采用 TTL 逻辑(0 V ~ 5 V)的原因,是为了提高抗干扰能力和增加传送距离。由于传号和空号状态用相反的电压表示,其间至少有 6 V 的差异,从而极大地提高了数据传输

输的可靠性。

表 6-1 RS-232C 信号电压的含义

状态	信号电压
二进制逻辑	-15 ~ -3 V
信号状态	1
	传号 (MARK)
	0
	空号 (SPACE)

RS-232C 用正、负电压来表示逻辑状态,而 TTL 以高、低电平表示逻辑状态,因此,为了能够同计算机接口或终端的 TTL 器件连接,必须在 RS-232C 与 TTL 电路之间进行电平和逻辑关系的变换。

4. 接口信号参数

由于 RS-232C 属于异步串行接口,它的信号特点之一就是每一个二进制的传输时间严格一致,以便于在接收端确定每一个位的开始和结束时间。这一点确定了异步串行通信必须确定传输速率,而且传输速率在整个通信结束之前是不能改变的。除此之外,RS-232C 标准还规定了一个字节的传送波形,包括起始位、终止位和校验位等,如图 6-7 所示。

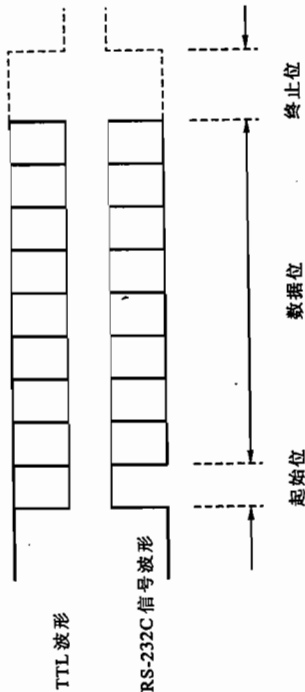


图 6-7 RS-232C 的信号波形

实际上,标准规定了 RS-232C 通信速率的一系列可选值,在传送之前首先选定传输速率,而且通信双方的传输速率必须一致。设定的参数如下:

- ① 可设定通信速率为以下值之一:150、300、600、1 200、2 400、3 600、4 800、7 200、9 600、14 400、19 200、28 800、33 600、38 400、57 600、115 200,单位是“位每秒(bps)”。
- ② 可设定数据位的个数,5~8 个数据位。
- ③ 可设定停止位长度,1、1.5 或 2 个停止位。
- ④ 可设定字节校验为有校验(奇校验、偶校验)或无校验等。选择有校验后,发送电路将自动产生校验位,并将其插入到数据位和终止位之间发送出去。

6.2.2 RS-232C 的接口信号

RS-232C 标准规定了在串行通信时,数据终端设备(如计算机)和数据通信设备(如调制解调器)之间的接口信号。所谓“发送”和“接收”是从数据终端设备的角度来定义的。表 6-2 列出了 RS-232C 的 DB9 和 DB25 互接线对应表。下面就 DB9 连接器引脚号介绍信号线,按照它们的功能可分类介绍如下。

表 6-2 DB9 和 DB25 互接线对应表

DB9 引脚	DB25 引脚	信号名	描述
1	8	CD	载波检测
2	3	RxD	接收数据
3	2	TxD	发送数据
4	20	DTR	数据终端就绪
5	7	SG	信号地
6	6	DSR	数据设备就绪
7	4	RTS	请求发送
8	5	CTS	清除发送
9	22	RI	振铃指示

1. 地线

引脚 5:信号地,这是其他各信号电压的参考点。无论电缆如何连接,这条线是必不可少的。

2. 接收、发送数据线

引脚 3:发送数据(TxD),由 DTE 发至 DCE 的信号。在数据未发送时,一直保持负电压。

引脚 2:接收数据(RxD),由 DCE 发至 DTE 的信号。

3. 常用控制信号

引脚 7:请求发送(RTS),由 DTE 发至 DCE 的信号,表示它要向 DCE 发送数据。当数据设备就绪(DSR)、数据终端就绪(DTR)为正电平时(表示接通),RTS 就应为正电压(表示接通)。

引脚 8:清除发送(CTS),由 DCE 发至 DTE 的信号,表示 DCE 已准备好接收来自 DTE 的发送数据。如果数据设备就绪(DSR)为断开状态(负电压),则 CTS 也应该是断开状态(负电压),以表明 DTE 不应发送数据。CTS 接通的条件是数据设备就绪(DSR)、请求发送(RTS)为正电压。

引脚 6:数据设备就绪(DSR),由 DCE 发至 DTE 的信号,表示 DCE 已与通信信道相连接。

引脚 4:数据终端就绪(DTR),由 DTE 发至 DCE 的信号,表示 DTE 准备发送数据至 DCE。数据终端就绪(DTR)必须先接通,然后数据设备就绪(DSR)才能变为接通状态。

4. 与调制解调器有关的信号

引脚 9:振铃指示(RI),由 DCE 发至 DTE 的信号,RI 为正电压时,指示 DCE 正在接收振铃

信号。在每次振铃期间 RI 为接通状态,而在两次振铃之间,则为断开状态。

引脚 1:载波检测(CD),当 DCE 接收到满足要求的载波信号时,CD 便是正电压,这个信号可用来驱动载波检测发光二极管。

在上述各信号线中,发送数据、接收数据和信号地这 3 条线是最基本的。DSR、DTR、CD 和 RI 是针对电话网络设计的。在本地互连的微型计算机系统,最常用的联络信号是 DTR、DSR、RTS、CTS。

对数据终端而言,若要发送数据,其必要条件是 DTR、DSR、RTS、CTS 应接通(正电压)。有时 RTS 不一定是必要条件,但 DTR 必须接通,并且要求 DSR、CTS 这两条输入线必须接通。当 DTE 与 DCE(或 DTE)互连时,这两个信号来自互连的设备,也可以用模拟的方法产生,例如通过软件设置或者将 DTR、CTS 直接连接到正电压。

6.2.3 RS-232C 串行接口的实现

在最早的 PC 机中,串行接口是由一块独立的 IC 芯片实现的,如 Intel 8250,实现串行通信的功能部件被称为通用异步收发器(Universal Asynchronous Receiver/Transmitter, UART),如图 6-8 所示。在后来的 PC 机中,将 UART 和其他的标准接口电路集成在一起,被称为超级 I/O 芯片。在现代的芯片组结构中,超级 I/O 芯片被集成到了南桥芯片中。

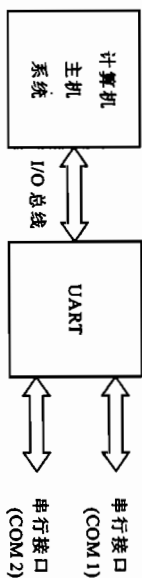


图 6-8 串行接口的实现

一般情况下,PC 机提供两个串行接口,分别被称为 COM1 和 COM2。如果需要更多的串行接口,可以使用扩展卡扩展得到。串行接口的接口电路支持中断,所以编写串行通信软件大多使用中断控制方式,在 Windows 操作系统下一般不提倡采用查询控制方式编写接口软件。

在标准配置中,COM1 使用 4 号中断通道 IRQ4,COM2 使用 IRQ3。当一个串行接口装入系统时,必须配置使用专用的 I/O 地址(端口)和中断(中断请求 IRQ)。最好的方案是遵循这些设备的默认配置,建议使用表 6-3 所示的地址和中断。

表 6-3 标准串行接口地址和中断

COMx	I/O 端口地址	中断号
COM1	3F8~3FFh	IRQ4
COM2	2F8~2FFh	IRQ3
COM3	3E8~3EFh	IRQ4*
COM4	2E8~2EFh	IRQ3*

每一个接口都使用了多个寄存器,每一个寄存器都是 8 位的,通过对这些寄存器的读或写,可实现对接口的操作。根据设计的接口电路的不同,端口在读写特性方面可以分为只读、只写和读写三种。串行接口电路占用 8 个端口,这些端口地址被寄存器分配使用,主要的寄存器如表 6-4 所示。表中端口地址以偏移量表示,如 COM1 使用 3F8 位基址,表 6-4 中的端口地址为 +0,表示 3F8, +1 表示 3F9,依此类推。

表 6-4 RS-232 接口的寄存器

寄存器	端口地址	端口操作	功能
发送寄存器	+0	只写	发送数据
接收寄存器	+0	只读	接收数据
速率设置寄存器	+0/1	读写	设定通信速率
中断允许寄存器	+1	读写	设定哪些事件可以产生中断
中断标志寄存器	+2	只读	产生中断后指示中断事件编码
FIFO 控制寄存器	+2	只写	控制 FIFO 缓冲区
线路控制寄存器	+3	读写	设置数据位、停止位长度、校验位等
Modem 控制寄存器	+4	读写	完成对 RS-232C 外接的 Modem 的控制
线路状态寄存器	+5	只读	用以软件查询接口电路的状态
Modem 状态寄存器	+6	只读	检测 RS-232C 外接的 Modem 的状态

数据传送使用两个寄存器——发送寄存器和接收寄存器,这两个寄存器共用一个端口地址,分别为只写和只读。

速率设置寄存器用于设置通信速率,使用与数据寄存器(相对地址为 +0)和中断允许寄存器(相对地址为 +1)地址相同的端口地址,由线路控制寄存器中的最高位来确定当前这两个端口是否用于设置通信速率。

中断允许寄存器和中断标志寄存器对于中断控制方式的程序设计是必不可少的,前者确定哪些事件可以引发中断,这些事件可能包括收到新数据、前一个字节发送完毕、线路状态发生变化等。当中断发生,进入中断处理程序后,可以读取中断标志寄存器,用以确定是什么事件引起的中断。

FIFO(First In First Out,先进先出)控制寄存器是集成在 UART 芯片中的存储区,不同的芯片可能容量不同,多数为 64B。FIFO 的作用是作为接收和发送的缓冲存储区。FIFO 机制的引入使得软件每次发送到 UART 的数据和从 UART 接收的数据不是一个字节,可以多达几十个字节,这样做的目的是提高软件的执行效率,使得即使在高通信速率的情况下,产生中断的频率也可以保持在比较低的水平。

线路控制寄存器用于控制接口电路(UART)的一些参数,如数据位数、停止位长度、校验方式等。同时线路控制寄存器的最高位用于确定 +0 和 +1 两个寄存器是数据寄存器与中断允许寄存器还是用于设定通信速率。

线路状态寄存器的 8 个位可以表示 UART 多方面的状态,如前一个字节是否发送完毕、是否

接收到新数据、是否产生错误等。如果接口软件采用查询方式,就是查询状态寄存器,以断定是否需要进行的事件发生。

6.3 并行接口

并行接口一般用于连接打印机。这是并行接口最初唯一的用途,但是随着 PC 的发展,并行接口也逐步发展为多用途的、相对高速的设备间接口。目前,并行接口的传输速率与 USB 1.1 接口相当。IEEE 1284 标准的最终版本于 1994 年 3 月被批准。该标准定义了并行接口的物理特性,包括数据传输模式和物理及电气规范。

6.3.1 并行接口简介

目前,在 PC 机上的并行接口是以 Centronics 接口为基础发展起来的。Centronics 接口是一种三线信号交互的 8 位并行接口,这种接口不支持外部设备选址,因此在输出端只能接一个设备。由于并行接口主要用来连接打印机,所以并行接口的信号线定义都是为了连接打印机的方便。基本的并行接口的信号线定义如表 6-5 所示。

表 6-5 并行接口连接打印机时的信号

PC 机 DB25 并行接口引脚	信号线 功能定义	信号方向 (主机→打印机)	对应的 Centronics 36 连接器引脚
1	选通(STB)	→	1
2~9	数据 D0~D7	→	2~9
10	应答信号(ACK)	←	10
11	忙信号(BUSY)	←	11
12	无纸信号(PE)	←	12
13	联机信号(SLCT)	←	13
14	自动进纸(AUTO LF)	→	14
15	错误信号(ERROR)	←	32
16	初始化(INIT)	→	31
17	选择输入(SLCT IN)	→	36
18~25	地(GND)		14,16

STB:低电平有效,用于主机对打印机的数据选通。

AUTO LF:低电平有效,打印完后自动进纸换行,有些打印机通过 DIP 开关设置。

INIT:低电平有效,使打印机的控制器初始化信号,同时清除打印缓冲区。

SLCT IN:低电平有效,使打印机处于联机状态。

ACK:低电平有效,表示打印机准备好,可以接收数据。

BUSY:高电平有效,表示打印机处于忙状态,包括正在输入数据、正在打印、脱机状态、打印机就绪状态。

PE:高电平有效,表示打印机缺纸。

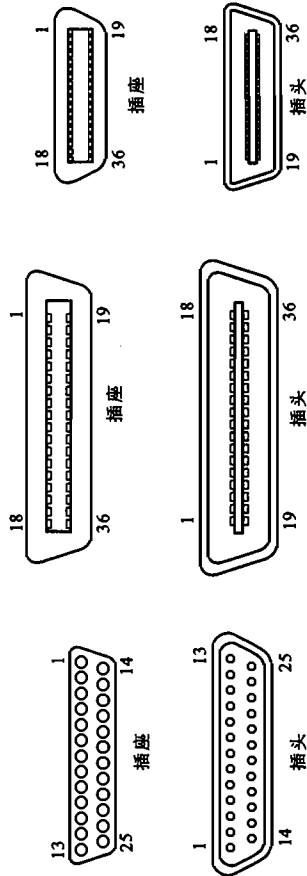
SLCT:高电平有效,表示打印机为联机状态。

ERROR:低电平有效,表示打印机出错,包括无纸、脱机以及其他错误状态。

IEEE 1284 规范所衍生的一个条约则定义了软件接口。已经成立的 IEEE 1284 委员会将负责制定一种软件标准,与 IEEE 1284 规范兼容的硬件都将使用该标准。该标准的研究主要是为了规范并行接口芯片厂商之间的差异。

IEEE 1284 支持在计算机与打印机之间或两台计算机之间以更高的吞吐量进行连接,所使用的打印机电缆不再是普通的打印机电缆。IEEE 1284 规范中的打印机电缆使用了双绞线技术,这样可以提高传输的可靠性。

IEEE 1284 标准还定义了并行接口连接器,其中包括两种现有的类型(称为 A 型和 B 型)和一种被称为 C 型的小型连接器。A 型指的是标准的 DB25 连接器,它用于大多数 PC 系统并行接口的连接;而 B 型指的是大多数打印机中标准 36 引脚 Centronics 模式的连接器;C 型是一种新的小型高密度 36 引脚的连接器。这 3 种连接器如图 6-9 所示。



(a) IEEE 1284 A 型连接器

(b) IEEE 1284 B 型连接器

(c) IEEE 1284 C 型连接器

图 6-9 IEEE 1284 的 3 种连接器

并行接口最基本的操作过程是,当打印机就绪时,BUSY 信号是低电平,然后计算机把数据放在数据线上,并把选通脉冲送到 STB 线上,这时打印机将 BUSY 忙信号变为高电平,并读出锁存的数据,把数据放到打印队列中,同时输出一个 ACK 响应脉冲。在 ACK 脉冲之后,BUSY 信号变为低电平。如果打印机发现有错误,并处于检查状态,打印机就启动 ERROR 线。

1. 正常信号交互

在正常信号交互时,计算机向打印机输出数据线上输出有效信息。在选通脉冲变为逻辑负值之前,至少 1μs 时间数据必须是有效的,并且在选通脉冲变为低电平后,数据必须至少保持

1 μs 是有效的, 选通脉冲为低电平的时间大约为 $1 \mu\text{s} \sim 500 \mu\text{s}$ 。

选通脉冲的下降沿使得打印机送出 ACK 响应信号。选通脉冲变为低电平到 ACK 信号之间的时间延迟大约为 $2 \mu\text{s} \sim 10 \mu\text{s}$ 。以上时间是正常数据传输情况下的情况下。

2. 忙状态信号交互

当打印机的打印缓冲区中有打印数据, 或者当垂直进纸、换行、删除、报警、选择状态时, 打印机就处于忙状态。当打印机收到这些控制某种机械操作的特殊字符时, 它需要用几微秒以上的时间进行忙操作, 这时信号交互的时序变成 BUSY 状态来对此做出响应。

也有些打印机不使用 BUSY 线, 因为不论是正常的还是较忙的信号交互, 它们的结束都是一样的, 这样就可以用两线信号交互代替三线。另外一些打印机用一个开关来补充 BUSY 线, 不论是两线还是三线都可以使用这种开关。

6.3.2 并行接口的工作模式

IEEE 1284 并行接口标准定义了五种不同的端口操作模式, 其中包括了传输速率比较高的 EPP 模式和 ECP 模式, 5 种操作模式如表 6-6 所示。

表 6-6 IEEE 1284 并行接口操作模式

并行接口模式	数据传输方向	数据传输速率
半字节(4位)	输入	50 Kbps
字节(8位)	输入	150 Kbps
兼容	输出	150 Kbps
EPP	输入/输出	500 Kbps ~ 2 Mbps
ECP	输入/输出	500 Kbps ~ 2 Mbps

输入和输出可以采用不同的操作模式, 将这 5 种模式进行组合, 可以得到 4 种并行接口类型, 如表 6-7 所示。

表 6-7 IEEE 1284 的 4 种并行接口类型

并行口类型	输入模式	输出模式	注释
SPP(标准并行接口)	半字节	兼容	8 位输出/4 位输入
双向并行接口(PS/2)	字节	兼容	8 位输入/输出
EPP(增强型并行接口)	EPP	EPP	8 位输入/输出
ECP(增强性能端口)	ECP	ECP	8 位输入/输出, 使用 DMA

1. 标准并行接口 SPP

最早的 PC 机的并行接口的功能几乎是唯一的, 就是连接打印机, 所以当时的并行接口只是能输出的。当时的 PC 机没有快速的输入接口, 当出现快速的输入/输出设备时, 人们希望并行接口是双向的。一种简单的改造方案就是将 4 根信号线作为 4 位输入的连接。因此, 这些端口就能实现 8 位输出(称为兼容模式)和 4 位输入(称为半字节模式)。这种方法在台式计算机系统

统中很常见。而 1993 年以后生产的计算机系统则可能都带有功能更强大的并行接口, 例如双向并行接口、EPP 或 ECP。

标准并行接口的有效传输速率为: 输出 150 Kbps, 输入 50 Kbps。

2. 双向 8 位并行接口

随着在 1987 年 PS/2 系列计算机的问世, IBM 推出了双向并行接口。这种并行接口仍然能够在现在的 PC 兼容系统找到, 并且被称为“双向”或“PS/2 类型”接口。这种接口在设计上为在计算机与外部设备之间利用并行接口进行高速双向通信开辟了道路, 这种接口是通过定义一些在并行接口连接器中以前未使用的引脚以及定义一个标明通道中信息流动方向的状态位来实现的。它支持真正的 8 位(称为字节模式)输入。

这种并行接口通过使用标准的 8 位数据线可以实现 8 位输入和输出, 而且在外部设备使用这种接口时比使用 4 位的接口要快得多。双向并行接口在进行输入和输出时都能达到 150 Kbps 的传输速率, 有些较新的系统将此作为其标准模式。

3. 增强型并行接口 EPP

(增强型并行接口(Enhanced Parallel Port, EPP)是一种更新的规范, 有时也被称为高速模式并行接口)。EPP 是由 Intel 等几家公司合作开发的, 于 1991 年 10 月发布。最早支持 EPP 的产品有 Zenith Data Systems 公司的便携式计算机、Xircom 公司的袖珍局域网适配器以及 Intel 的 82360 SL I/O 芯片。当前, 几乎所有的包含有多模式并行接口的计算机系统都支持 EPP 模式, 这些系统中的超级 I/O 芯片都被集成在主机板上。

EPP 基本上与 ISA 总线相等的速率进行传输操作, 并且其吞吐量比常规并行接口提高了 10 倍。EPP 是专门为并行接口外部设备所设计的, 这些外部设备包括局域网适配器、磁盘驱动器和磁带机。EPP 已经被包括在新的 IEEE 1284 并行接口标准中。使用 EPP 有可能使传输速率达到 2 Mbps。

当今主机板上使用的所有超级 I/O 芯片以及集成了 I/O 功能的南桥芯片都支持 EPP。因为 IEEE 1284 标准中定义了 EPP, 所以它也得到了软件和驱动程序的支持, 其中 Windows NT 就支持 EPP。

4. 增强性能端口 ECP

另一种类型的高速并行接口为增强性能端口(Enhanced Capability Port, ECP), 是由 Microsoft 和 HP 公司联合开发的, 它于 1992 年正式发布。与 EPP 一样, ECP 改善了并行接口的性能, 并且要求专门的硬件逻辑电路。

自从最初被发布后, 与 EPP 一样, ECP 也被包括在 IEEE 1284 中。然而与 EPP 不同的是, ECP 没有设计为可以支持便携式 PC 的并行接口外部设备, ECP 的设计目标是性能更高的打印机和扫描仪提供廉价的连接方式。此外, ECP 模式还支持 DMA 传输, 而 EPP 没有定义对 DMA 的使用。

大多数新的提供了 ECP 的计算机系统都支持高吞吐率通信。在许多情况下, 可以通过 BIOS 设置将 ECP 变成 EPP 或标准并行接口。然而, 如果要得到最佳的吞吐量, 那么建议使用 ECP 模式。

6.3.3 并行接口的使用

PC 机可配有两个并行接口,即 LPT1 和 LPT2。二者结构是相同的,都可以作为打印机接口使用。由于有的机器不配置 LPT2,因此常将 LPT1 作为打印机接口。标准并行接口有 3 个端口:数据端口、状态端口和控制端口,分别与数据寄存器、状态寄存器和控制寄存器相对应。表 6-8 所示为两个并行接口的端口地址。

表 6-8 并行接口的端口地址

并行接口	数据端口	状态端口	控制端口
LPT1	378H	379H	37AH
LPT2	278H	279H	27AH

CPU 向打印机传送打印数据或控制命令时,分别通过数据端口和控制端口,对数据寄存器和控制寄存器进行写操作,而 CPU 要获得打印机状态时,则通过状态端口对状态寄存器进行读操作。

IBM PC 最初的设计者们设想并行接口将只被用于与打印机的通信。在后来的发展中,能够使用并行接口的设备的数量迅猛增加。包括磁带机、局域网适配器、CD-ROM 驱动器和扫描仪等,这些设备都可以通过并行接口连接。这些设备中的多数也可以通过 USB 接口与系统连接,有的设备支持这两种连接方式,因而具有极大的灵活性。

也许双向并行接口最常见的用途是在一台计算机与另一台计算机之间传送数据。如果传送数据的两台计算机都使用了 EPP/ECP,那么实际上就能够以高达 2 MBps 的速率进行通信,这比串行接口或红外接口的数据传输速率要快得多。许多商用程序也支持并行接口文件传输,如 Laplink 与 PC Anywhere 等,MS-DOS 6、Windows 也内嵌了对并行接口文件传输的支持。

并行接口也能被用来将 SCSI 外部设备连接到 PC 上。通过并行接口 SCSI 转换器,可以将任何类型的 SCSI 设备,如硬盘、CD-ROM 驱动器、磁带驱动器或者扫描仪连接到 PC 上。如果需要支持两个以上的设备,需要购买一个 SCSI 宿主适配器。但需要注意的是,即使是 EPP/ECP,其速率也不及最低速的 SCSI 接口。

6.4 USB 接口

通用串行总线(Universal Serial Bus, USB)接口是一种全新的外部设备接口。从 1998 年开始,PC 机主板开始支持 USB 接口。近几年,随着越来越多的 USB 接口外部设备的出现,USB 接口已成为 PC 机主板的标准配置。从发展趋势上看,USB 将取代 PC 机的大部分标准和非标准接口。

USB 是外设总线标准,是由在 PC 和电信产业中的大型公司,包括 Compaq、DEC、IBM、Intel、Microsoft、NEC 和 Northern Telecom 共同开发的。

6.4.1 新型的串行传输技术简介

最新的高性能外部总线设计趋势是使用串行结构,这种结构通过 1~2 根信号线一次发送 1 位数据。而并行结构则同时使用 8 根、16 根或更多的信号线来发送数据。在相同的时钟频率下,并行总线要快得多。但是,提高串行连接的时钟频率却要比提高并行连接的时钟频率要容易得多。

并行连接将面临多个问题,其中最大的问题是信号的不同步和振动。不同步和振动是將高速并行接口的连接电缆长度限制为 3 m 或更短的原因。出现不同步和振动的原因是:虽然 8 位或 16 位数据被发送器同时发送,但是当它们到达接收端时,传输延迟已经使得一些数据位先于其他数据位到达。电缆越长,在电缆的另一端第一个和最后一个数据位到达的时间间隔将越长。信号的不同步,使得不能以更高的传输速率或更长的电缆来进行通信。振动指的是信号到达接收端时,电压在两个逻辑电平之间的变化过程中的上下波动。

使用串行总线可以一次发送 1 位数据。由于无需担心每一位数据的到达时间,因而可以大幅度地提高时钟频率。例如,EPP/ECP 的最大传输速率为 2 MBps,而 USB 2.0 支持的传输速率可以为 480 Mbps(约 60 MBps),是高速并行接口的 30 倍。

在高时钟频率下,并行信号将相互干扰。但串行信号只有 1 根或 2 根信号线,所以串行传输又具有一个优势,那就是可以忽略电缆导线之间的串扰和干扰。

一般说来,并行电缆比串行电缆要贵得多。在并行电缆中,除了需要许多额外的导线来传送多个数据位外,所用的电缆也需要特制,以尽量减少相邻数据线之间的串扰和干扰。这就是为什么 SCSI 电缆比较昂贵的一个原因。相对而言,串行电缆要廉价得多。主要原因是,它使用的导线要少得多。其次,即使在高速下,它的屏蔽要求也要简单得多。正因为如此,在更长的距离之间也可以比较容易地保证串行数据的可靠性,这就是建议的并行接口电缆长度要比串行接口的短的原因。

基于上述原因,并为了满足新的即插即用外部设备接口和减少便携式计算机上拥挤的物理端口的需要,出现了像 USB 这样的高性能串行总线。USB 是现在所有 PC 的标准功能,并且它可以当通用外部接口使用。对于某些应用领域(例如连接视频设备)来说,IEEE 1394 也可以作为一种选择,它正在向高带宽方向发展,如高分辨率扫描仪和外部硬盘驱动器等。

6.4.2 USB 简介

USB 是一种外部设备总线标准,其标志如图 6-10 所示。它的设计使 PC 的外部设备实现了即插即用。USB 的出现不再需要专用的接口,也减少了专用 I/O 卡的使用,从而也减少了因添加新卡而重新配置系统的需要,大大节省了像 IRQ(中断请求)这样重要的系统资源,所有连接到 USB 接口上的设备只需要 1 个 IRQ 即可。带有 USB 的 PC 可以支持对外部设备的自动识别与设置,只需将外部设备在物理上连接到 PC,操作系统可以检测到 USB 外部设备,为外部设备安装驱动程序。USB 在同一台计算机上最多可以同时支持 127 台设备的运行。

Intel 公司是 USB 最早的支持者,从 1996 年以后所有的 PC 芯片组就已经将 USB 作为标准包含在内。其他芯片组生产厂商也将 USB 视为与串行、并行接口一样的台式机和笔记本电脑



图 6-10 USB 标志

的标准特性。

Intel 及其 USB 合作开发厂商一起创建了 USB 实现者论坛(USB Implementers Forum, USB-IF)来开发、支持和推进 USB 体系结构的发展。USB-IF 于 1996 年 1 月发布了 USB 1.0 版,于 1998 年 9 月发布了 USB 1.1 版。其中 1.1 版主要是对关于集线器和该规范其他领域中一些问题的说明。大多数设备和集线器应该都是遵循 USB 1.1 标准的,即使它们是在 USB1.1 规范发布之前生产的。之后,USB-IF 发布了 USB 2.0 标准,其速度比最初的 USB 快 40 倍,而仍然具有完全的向后兼容性。通过插入一块 PCI 卡(台式计算机)或 PC 卡(兼容 Cardbus 的笔记本电脑),就可以在设有内置 USB 连接器的老式计算机上使用 USB 端口。

6.4.3 USB 的技术特点

USB 1.1 版被定义为通过简单的 4 线连接以 12 Mbps(1.5 MBps)的速率运行,USB 2.0 的速率达到了 480 Mbps。USB 2.0 总线支持多达 127 台设备,并且使用了级联-星状的拓扑结构,该结构由 PC 内的扩展集线器、所有的 USB 外部设备、甚至是独立的集线器盒所构成。注意,虽然 USB 标准允许连接多达 127 台设备,但是所有的设备只能共享固定的带宽,这就意味着每增加一台活动的设备,总线就将慢一些。

1. USB 传输

USB 采用平衡传输方式,就是同一个信号使用两条线传送,接收端判断两条线的电位差,以断定传送的逻辑数据。平衡传输能有效地提高抗干扰能力,也就能采用更高的传送速率。USB 接口只有 4 根线,其中 V_{cc} 、GND 两根用于向设备供电, -Data 和 +Data 是信号线,如图 6-11 所示。

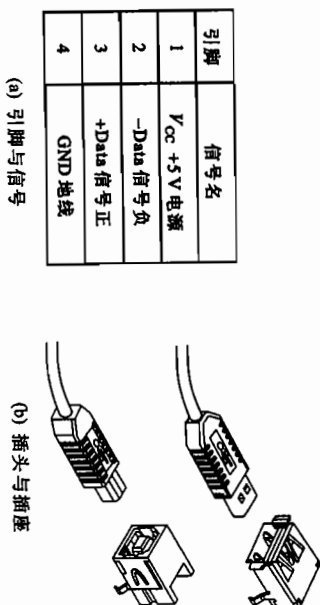


图 6-11 USB 信号、插头和插座

USB 采用半双工传输方式,可以双向传输但不能同时进行。USB 1.1 标准的集线器支持 12 Mbps 的全速和 1.5 Mbps 的低速外部设备。USB 2.0 标准增加了 480 Mbps 的高速模式,但同时支持 USB 1.1 的全速和低速模式。

2. NRZI 编码

USB 使用的是被称为不归零反向码(Non Return to Zero Invert, NRZI)的数据编码方法。

NRZI 是一种对串行数据进行编码的方法。在该方法中,0 和 1 都由相对并相互作用的电平和高电平来表示,在编码位之间不会返回到零(或基准)电平,如图 6-12 所示。在 NRZI 编码中,用 1 表示信号电平没有变化,用 0 表示信号电平有变化。一个逻辑 0 的连续序列将使得 NRZI 的数据在每个位时间进行切换。一个逻辑 1 的连续序列将导致数据长期没有变化,为了避免这种情况,需要在连续 6 个 1 之后插入一个 0,接收端再按照这样的规律将插入的 0 删除。NRZI 是一种高效率的编码方法,因为它不需要额外的时钟脉冲,不会浪费时间和带宽。

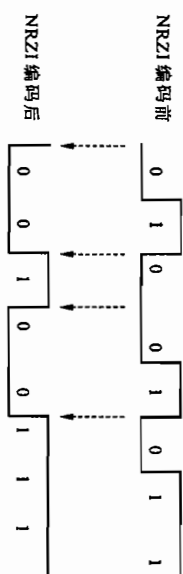


图 6-12 NRZI 编码

3. USB 集线器

USB 设备被分为集线器和功能设备,或两者的结合。集线器提供了对 USB 的额外连接点,支持连接外加的集线器或者功能设备。功能设备是连接到 USB 上的单独设备,例如键盘、鼠标、数码相机、打印机及 U 盘等。在 PC 设备系统中的初始接口被称为根集线器,它们是 USB 的起点。

大多数主板带有 2~8 个 USB 接口,每个接口都能连接功能设备或者附加的集线器。有些计算机的前面板上设计了一两个 USB 接口,为偶尔使用的设备提供了方便,如数码相机、移动硬盘、闪存卡读卡器等。

集线器在本质上就是连线的集中器。在星状拓扑结构中,集线器可以连接多台设备,如图 6-13 所示。每个连接点被称为一个端口。大多数集线器有 4 个或 8 个端口,也可能更多。为了保证更大的扩充性,可以将附加的集线器连接到现有集线器的端口上。一个集线器都被分配一个唯一的地址。集线器最多可级联 5 层。

4. USB 供电

除了为连接 USB 外部设备提供附加的插座,集线器还为连接的所有外部设备提供电源。集线器可以识别外部设备的动态连接,并且在外部设备的初始化期间为每个外部设备提供至少 0.5 W 的电源。在主机 PC 驱动程序软件的控制下,集线器可以为外部设备提供更多的电源,最多可达 2.5 W。USB 的即插即用特性支持系统询问所连接外部设备的电源需求,如果超过了系统实际的供电能力,将发出一个警告。

5. USB 中断

USB 接口的一个优点是只需要一个 PC 中断。这意味着很多个外部设备共用一个中断通道,不需要为每一个外部设备单独分配中断。中断通道在系统中是有限的资源,设有 USB 接口的系统可能会因为连接的外部设备太多而产生中断通道资源紧张的状况。

3. 中断传输

支持像游戏手柄、鼠标和键盘等输入设备,这些设备与主机之间的数据传输量小,无周期性,但对响应时间敏感,要求马上响应。

4. 数据块传输

适用于打印机、扫描仪、外接移动盘等外部设备,这些外部设备与主机之间传输的数据量大,USB在满足带宽的情况下才进行该类型的数据传输。

USB采用分块带宽分配方案,若外部设备超过当前带宽分配的要求,则不能进入该设备。USB为等时传输和中断传输类型的外部设备保留带宽,并保证数据按一定的速率或延时传送。数据块和控制外部设备按可用的最佳带宽来传输数据。

6.5 IEEE 1394

USB: 成环状、通周传输
1394: 点对点、双向传输、全双工

IEEE 1394 是由 IEEE 标准委员会于 1995 年下半年发布的。IEEE 是(美国)电气电子工程师学会(Institute of Electrical and Electronic Engineers)的缩写,该组织专门从事有关标准的制定。1394 是 IEEE 标准的序号。它是今天音频和视频多媒体设备的大容量数据传输移动需要的产物。IEEE 1394 的主要优点是速度快。

6.5.1 IEEE 1394 标准

IEEE 1394 也被称为火线(FIREWIRE),其最初的版本实际上被称为 IEEE 1394a,由于它于 2000 年被采纳为标准,有时也称为 IEEE 1394a. 2000。IEEE 1394a 标准在最初的 IEEE 1394 标准的基础上解决互操作和兼容问题,它使用了与最初的 IEEE 1394 标准相同的连接器,并支持同样的传输速率。

IEEE 1394a 标准定义了两种总线模式:底板(Backplane)模式和电缆(Cable)模式。其中底板模式支持 12.5 Mbps、25 Mbps、50 Mbps 的传输速率;电缆模式支持 100 Mbps、200 Mbps、400 Mbps 的传输速率。

IEEE 1394b 是为下一代 PC 所制定的标准,它将 IEEE 1394a 的传输速率直接提高到 800 Mbps 和 1 600 Mbps,如果使用光纤的话,最高传输速率可提高到 3.2 Gbps。

IEEE 1394 是横跨 PC 及家电产品平台的一种通用接口,适用于大多数需要高速数据传输的产品,如高速内置式硬盘、CD-ROM、DVD-ROM、扫描仪、打印机、数码相机、摄像机等。IEEE 1394 接口分为有供电功能的 6 引脚 A 型接口(如图 6-14 所示)和无供电功能的 4 引脚 B 型接口。A 型接口可以通过转接线兼容 B 型接口,但是 B 型转换成 A 型后则没有供电能力。6 根导线中的 4 根导线用于传送数据,2 根导线供电。传送数据的 4 根线以平衡方式传送数据和时钟脉冲。和 USB 相比,IEEE 1394 多了时钟信号,可以进行同步串行传输。6 引脚的 A 型接口在苹果计算机和周边设备上使用很广,而在消费类电子产品以及 PC 上多半都是采用的简化过的 4 引脚 B 型接口,需要配备单独的电源适配器。IEEE 1394 接口可以直接作为网卡使用,也可以通过过集线器扩展出更多的接口。没有 IEEE 1394 接口的主板也可以通过插接 IEEE 1394 扩展卡的

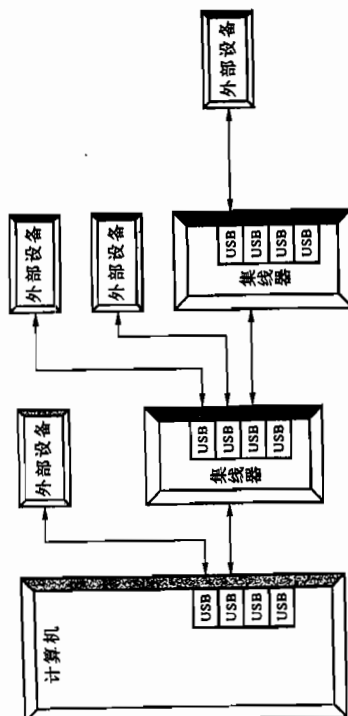


图 6-13 通过集线器扩展 USB 接口

6. 即插即用和热拔插

USB 规范的另一个优点是自动识别外部设备,这个特性大大简化了安装,因为完全不用为一个外部设备设置唯一的 ID 或标识符——一切均由 USB 自动处理。另外,USB 设备可以进行热拔插,这就是说每次连接或断开一个外部设备时,不必关机或重新启动计算机。

7. USB 的差错控制

USB 接口是高级接口,接口自身带有检错/纠错机制,这一特点给程序设计者带来极大的便利,可以不考虑数据传输错误的问题。

6.4.4 USB 传输模式

USB 作为 PC 机现在和未来的通用接口,必须满足各种各样的外部设备的传输要求。在这方面,USB 的设计是做了周全的考虑,USB 支持 4 种基本的数据传输模式:控制传输、等时传输、中断传输及数据块传输。当一个 USB 外设的某一种类型信息的传输特点确定后,就可以在上述 4 种传输模式中选择一种最合适的模式。4 种传输模式就是为了满足 4 种不同的传输需求。

1. 控制传输

支持外设与主机之间的控制、状态、配置等信息的传输,为外部设备与主机之间提供一个控制通道。所有外部设备都必须支持控制传输模式,这样主机与外部设备之间就可以传送配置和命令/状态信息。

2. 等时传输

支持有周期性且有限的时延以及带宽和数据传输速率不变的外部设备与主机间的数据传输。该模式无差错校验,故不能保证数据传输的正确性,适用于像计算机-电话集成(CTI)系统和音频系统与主机间的数据传输。

方式获得此功能。



图 6-14 IEEE 1394A 型插座和电缆的结构

大多数 PC 适配卡可以支持 200 Mbps 的速率。通过采用菊花链或树状连接方式，一个单独的 IEEE 1394 适配卡上最多可以连接 63 台设备。与 USB 设备不同，IEEE 1394 设备可以不使用集线器进行菊花链连接。不过对于可以进行热插拔的设备来说，一般推荐使用集线器。

与 USB 一样，IEEE 1394 支持完全的即插即用，并且具有热插拔功能。IEEE 1394 不需要复杂的终端，接口上连接的设备可以从接口获取多达 1.5A 的电能。与复杂的 SCSI 接口相比，IEEE 1394 接口提供了相同或更好的性能，而且费用要低得多，连接也更简单。

IEEE 1394 是构建在菊花链和树状拓扑结构上的，它支持多达 63 个结点，每个结点上可以支持多达 16 台设备的菊花链。如果还不够用的话，那么该标准还支持最多 1 023 条桥接的总线，这样就可以互连 64 000 个结点！另外，与 SCSI 一样，IEEE 1394 能够在同一条总线上支持不同传输速率的设备。

6.5.2 IEEE 1394a 和 USB 的比较

由于 USB 和 IEEE 1394 在形式和功能上的相似性，使人们容易对这两种技术产生混淆。表 6-9 总结了这两种技术的区别。

表 6-9 IEEE 1394a 和 USB 的对比

	IEEE 1394	USB 1.1	USB 2.0
是否必须有主机	不	是	是
最大设备数	63	127	127
最大电缆长度(m)	4.5	5	5
传输速率(Mbps)	400	12	480
未来的传输速率(Mbps)	400 ~ 3 200	无	无
典型应用	DV 便携式摄像机、HDTV、机顶盒、视频播放设备	键盘、鼠标、游戏杆、Modem、打印机	DV 便携式摄像机、HDTV、机顶盒、视频播放设备

1. 应用场合

目前，市场上出售的所有 PC 机都包含了 USB 接口，但对 IEEE 1394 的支持还是没有达

到这样的水平，只有笔记本电脑上普及率较高。即使 USB 的普及占据了压倒优势，但是 IEEE 1394 仍然有一定的市场。IEEE 1394 生存在普及的高速 USB 2.0 接口阴影下的主要原因也许是 USB 是以 PC 为中心的，而 IEEE 1394 却不是。IEEE 1394 的一个重要的好处就是不要求连接 PC 主机。IEEE 1394 可以被用来直接将数字视频 (DV) 便携式摄像机连接到视频处理系统，进行复制或编辑。这就是 IEEE 1394 能够在数字视频领域仍然通用，甚至超过 USB 2.0 的原因。

一般说来，IEEE 1394 为当前和将来的 PC 用户提供了空前的多媒体性能。现在的外部设备，特别是数字视频设备，仍旧相当昂贵。但是与所有新兴的技术一样，将来价格一定会降下来并且对居家和办公中使用的新 PC 敞开大门，使多数用户能够享受其进行高级视频编辑的性能。将来如果 PC 有更多的多媒体需要，那么 IEEE 1394 可能也会随之普及，而 USB 2.0 则主要为传统外部设备提供连接。

2. 成本

在成本方面，USB 2.0 较占优势。因为目前的主机板芯片组中都内置了 USB 主控制器，并且目前大多数外部设备都以 USB 接口为标准配置。因此用户基本上不需要再投入其他费用，就可以享受 USB 所带来的便利。与 USB 相比，IEEE 1394 控制器的结构比较复杂，要想将它集成到主机板芯片组中，无论在技术上还是在成本上都有一定难度，所以目前市面上几乎很少有集成 IEEE 1394 控制器的芯片组。要想实现 IEEE 1394 功能，除了主机板以集成附加芯片的形式提供外，一般只能通过插接 IEEE 1394 扩展卡来实现，这样做的直接结果就是使用成本上升了。

由于 IEEE 1394 最初的定位是在多媒体应用方面，与 USB 的大众化路线不一样，所以 IEEE 1394 的设备相对于 USB 设备会贵很多，加上 IEEE 1394 还要收高额的专利费，造成了使用成本居高不下。这是 IEEE 1394 的最大弱点。但随着未来芯片组整合 IEEE 1394 控制器，相信这个问题将会有所缓解。

3. 易用性

在易用性方面，IEEE 1394 很早就得到了 Windows 的全面支持，而 USB 2.0 直到 Windows XP SP1 才得到支持，而且 IEEE 1394 支持点对点的连接，如果两台计算机通过 IEEE 1394 相连，不必对计算机进行 IP 或任何其他设置就可以直接使用。此外，USB 2.0 只提供了 5 V 的直流电压和 0.5 A 的电流，虽然对于一般的设备来说已经够用了，但对于外置刻录机、MO 驱动器和打印机等耗电比较大的设备，就必须外接电源才能使用；而 IEEE 1394 提供了 8 V ~ 40 V 的电压及 5 A 的电流，理论上最大可以提供 200 W 的功率，远远高于 USB 2.0。

4. 传输速率

速率总是一个不断变化的参数。IEEE 1394 提供的数据传输速率比 USB 1.1 的传输速率快约 15 倍以上。虽然 USB 2.0 的出现改变了这种局面，但随着 IEEE 1394 更高速率版本（如建议的 IEEE 1394b 的基本速率达 1.6 Gbps）的出现，使其又重新夺回了传输速率的优势。USB 1.1 明显是为低速外部设备（例如键盘、鼠标、调制解调器和打印机）设计的，而 USB 2.0 可以被用来连接最快速的外部设备。IEEE 1394 主要用来连接高性能数字视频的电子产品和宽带计算机。

虽然 USB 2.0 可以提供 480 Mbps 的传输速率，略高于 IEEE 1394a 的 400 Mbps，那么是否意

意味着 USB 2.0 在传输速率方面更具优势呢? 答案是否定的。在一般情况下, USB 2.0 的实际传输速率只有 USB 1.1 的 2~13 倍, 远远达不到其理论值, 而且如果几台设备共用一个 USB 接口, 主控制芯片会对每台设备可以支配的带宽进行分配、控制, 这时的传输速率就更低了。而目前主流的 IEEE 1394a 则很少存在这种情况。IEEE 1394a 在突发传输速率、平均读取速率/写速率、工作性能、文件复制速率等方面都要远优于 USB 2.0, 而 IEEE 1394b 的优势将更为明显。不过, IEEE 1394 有一个缺点, 就是 IEEE 1394 总线需要占用大量的资源, 因此要让其达到最佳传输速率需要高速 CPU 来配合。

6.6 IDE 接口

集成设备电路 (Integrated Device Electronics, IDE) 属于内部接口, 是 PC 机用于连接硬盘、光盘驱动器的通用接口。一般 PC 机主机板上有两个 IDE 接口。IDE 接口的正式名称为 AT 附件 (AT Attachment, ATA), 是一个 ANSI 标准。“AT 附件”名字中的“AT”, 取自首次推出 16 位 ISA 总线的原始 IBM AT 计算机。

6.6.1 IDE 接口简介

IDE 接口用于连接硬盘 (如图 6-15 所示) 或是光盘驱动器, 使用带状电缆连接位于主机板上的 IDE 插座和硬盘/光驱 (IDE 设备) 的 IDE 插座, 主机板上的一个 IDE 接口可以同时连接两个 IDE 设备, 这两个 IDE 设备一个是主设备 (Master), 另一个是辅设备 (Slave)。IDE 设备都可以设置为主设备或辅设备。

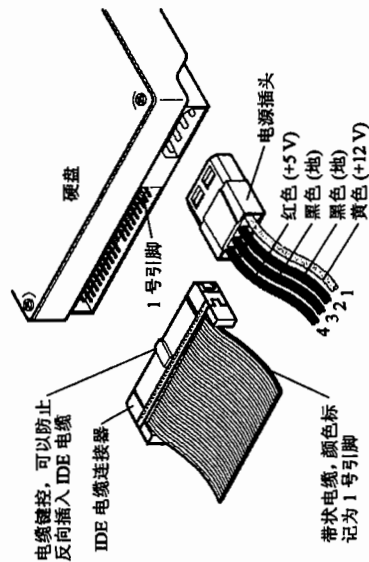


图 6-15 IDE 硬盘的连接

IDE 电缆是 40 线的带状电缆, 电缆插头带有键控, 而且 20 号引脚被阻塞, 如图 6-16 所示, 这样做是为了防止将带状电缆插反了。

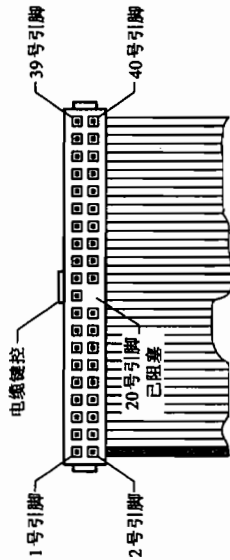


图 6-16 IDE 电缆连接头

40 线带状电缆专门用于承载总线适配器电路和驱动器之间的信号。为了尽量确保信号完整并消除部分同步和噪声问题, 电缆的长度不能超过 18 英寸。支持高速传输模式的 ATA 驱动器特别容易受到电缆故障和过长电缆的影响。如果电缆过长, 就会产生数据中断和其他错误, 这些错误在读/写驱动器时都会遇到。

使用 66 MBps 或 100 MBps 的传输速率的驱动器必须使用一种特殊的低质量 80 线电缆, 这种电缆增加了 40 根线, 但仍然使用 40 引脚的连接, 多余的引线用于接地, 以减少噪声。如果系统没有检测到 80 线电缆, 根据设计, 驱动器和主机适配器将禁用高速的 66 MBps 或 100 MBps 模式。如果驱动器处于 33 MBps 或更低速率的模式, 也建议使用这种电缆, 因为这只会好处而绝对没有坏处。图 6-17 所示是典型 ATA 电缆的尺寸和外观。

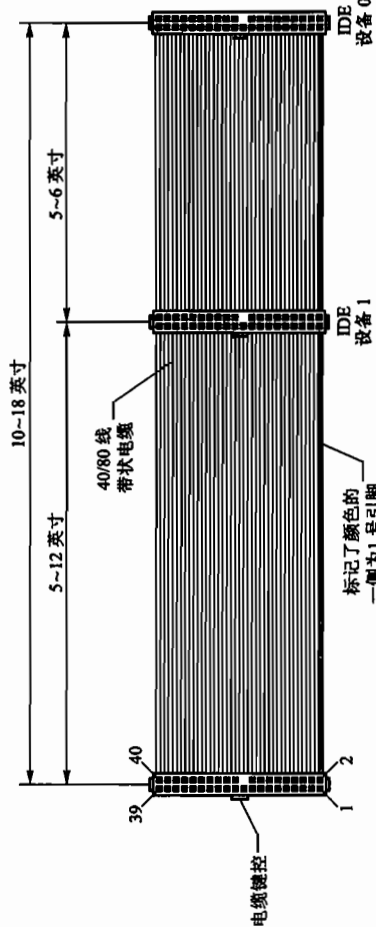


图 6-17 IDE 电缆

基本的 IDE 信号见表 6-10。从信号名称中可以看到, IDE 是 16 位并行接口, 中断和 DMA 通道分别使用 IRQ14 和 DRQ3。大多数信号直接来自 ISA 总线。

表 6-10 ATA 连接器

引脚号	信号名	引脚号	信号名
1	总清零	21	DRQ3
2	地	22	地
3	数据位 7	23	-IOW
4	数据位 8	24	地
5	数据位 6	25	-IOR
6	数据位 9	26	地
7	数据位 5	27	I/O CH RDY
8	数据位 10	28	SPSTYNC: CSEL
9	数据位 4	29	-DACK3
10	数据位 11	30	地
11	数据位 3	31	IRQ14
12	数据位 12	32	-IOCS16
13	数据位 2	33	地址位 1
14	数据位 13	34	-PDIAG
15	数据位 1	35	地址位 0
16	数据位 14	36	地址位 2
17	数据位 0	37	-CS1FX
18	数据位 15	38	CS3FX
19	地	39	-DA/SP
20	键脚(缺脚)	40	地

6.6.2 IDE 接口的版本和发展过程

IDE 接口是微型计算机最早使用的外存储器接口,经历了一个很长时期的发展,版本比较多,如表 6-11 所示。直到出现 S-ATA 接口,IDE 才走完了其发展历程。

表 6-11 IDE 接口发展简表

标准	年代	PIO 模式	DMA 模式	UDMA 模式	传输速率(Mbps)	说明
ATA-1	1986	0-2	0		8.33	
ATA-2	1996	0-4	0-2		16.67	又称为 Fast-ATA 或 EIDE
ATA-3	1997	0-4	0-2		16.67	
ATA-4	1998	0-4	0-2	0-2	33.33	UDMA/33 或 ATA-33
ATA-5	1999	0-4	0-2	0-4	66.67	UDMA/66 或 ATA-66
ATA-6	2001	0-4	0-2	0-5	100	UDMA/100 或 ATA-100

1. ATA-1

符合 ATA-1 标准的接口技术于 1986 年开始使用,但 ATA-1 标准于 1994 年才正式通过。ATA-1 定义了原始的 AT 附件接口,这是一个基于 ISA 总线的、在磁盘驱动器和主系统之间的集成总线接口。在 ATA-1 规范引入和记载的主要特性有:

- 40/44 引脚连接器和电缆连接。
- 主/从电缆选择驱动器配置选项。
- 支持基本程控输入/输出(Programmed Input/Output,PIO)和 DMA 模式的信号定时。
- 支持柱面 - 磁头 - 扇区(Cylinder-Head-Sector,CHS)和逻辑块地址(Logical Block Address,LBA)驱动器参数变换。

标准的 IDE 连接器总共有 44 个引脚,大多数的驱动器使用了前面 40 个引脚。增加的 4 个引脚(41~44),主要用在笔记本和膝上型计算机的小型 2.5/1.8 英寸驱动器上。因空间有限,这些驱动器没有单独的电源插座,因此额外的引脚主要用于为驱动器供电。

2. ATA-2

ATA-2 是对原始 ATA-1 标准的一个主要升级,于 1996 年获得通过,也许它是对 ATA 最大的改动。ATA-2 的升级对主系统与存储设备之间的接口做了通用的定义,而不仅仅是用于磁盘驱动器。和原 ATA-1 标准相比,ATA-2 中增加的功能主要包括:

- 快速 PIO 和 DMA 传输模式。
- 支持电源管理。
- 支持可移动设备。
- 支持 PCMCIA(笔记本计算机的标准扩展总线)设备。
- 定义驱动器支持高达 137.4 GB。
- 对容量高达 8.4 GB 的驱动器定义了标准 CHS/LBA 变换方法。

3. ATA-3

ATA-3 是比 ATA-2 标准小一些的版本,首次发表于 1997 年。其中包括一个明确的总规范以及很少的说明和修订。最主要的改动包括:

- 删除单字节(8 位)DMA 传输协议。
- 增加自监视、分析和报告技术(Self-Monitoring, Analysis and Reporting Technology,S.M.A.R.T.),支持对系统性能下降的预测。
- 加入安全模式,允许用通行字保护设备的访问。

ATA-2 和 ATA-3 是原 ATA-1 标准的扩展。最重要的扩展是性能增强,如快速 PIO 和 DMA 方式。ATA-3 还增加了一个简单口令的安全方案、更复杂的电源管理、S.M.A.R.T。这样使驱动器在发生故障时保留下问题的痕迹。

4. ATA/ATAPI-4

ATA-4 于 1998 年首次发布,包括对标准的几个重要增补。其中包含所谓 AT 附件包接口(AT Attachment Interface,ATAPI)的包命令特性,可以使一些设备,如 CD-ROM 驱动器、LS-120 超级软盘驱动器、磁带驱动器和其他种类的存储设备,通过一条公共的接口连接。在 ATA-4 出现以前,ATAPI 是一个分散颁布的标准。

- 高达 33 MBps 的 ultra - DMA (UDMA) 传输方式 (称为 UDMA/33 或 ultra - ATA/33)。
- 全面的 ATAPI 支持。
- 支持高级电源管理。
- 定义一根可选的 80 线、40 引脚的电缆,用以改善抗噪性能。

5. ATA/ATAPI - 5

这个版本的 ATA 标准仍然是建立在 ATA - 4 的基础上。标准草案里增加的主要内容包包括:

- 高达 66 MBps 的 Ultra - DMA (UDMA) 传输方式 (称为 UDMA/66 或 ultra - ATA/66)。
- UDMA/66 操作需要的 80 线电缆。
- 增加了自动检测是 40 线还是 80 线电缆的功能。
- 只有当检测到 80 线电缆时,快于 UDMA/33 的 UDMA 方式才有效。

从最初的 ATA - 1 到 ATA - 5,每一种标准都是相互兼容的,也就是说,可以将一个 ATA - 5 驱动器插入到 ATA - 1 系统中,或将 ATA - 1 驱动器插入到 ATA - 5 系统中。在这样的情况下,驱动器和系统在性能和功能上以其最低共同点进行操作。

6. ATA - 6

ATA - 6 是目前最新的 IDE 接口的标准。与 ATA - 5 硬盘接口技术相比,ATA - 6 可以使控制器与硬盘之间以 100 MBps 的传输速率进行数据交换 (所以也叫做 ATA - 100),它不仅能提高硬盘数据传输速率,同时还能够减轻硬盘数据缓存的负担。新的 ATA - 100 接口硬盘传输线由 80 根导线组成,但使用传统的 40 引脚连接线,以向下兼容老式 ATA - 33、ATA - 66 接口的设备。此外,为了保护数据的完整性,ATA - 100 同时也采用循环冗余校验 (Cyclic Redundancy Check, CRC) 技术来检查数据的传输错误。其原理是在每个数据块中加入一个帧检查序列 (FCS),FCS 包含了数据块的详细信息,专门用于测试发送及接收的数据的正确性,如果发现问题可立即纠正。由于它具有占用 CPU 资源少、检测准确等优点,因而为越来越多的厂商所采用。

6.7 SATA 接口

被广泛使用的 IDE 的最高版本是 ATA - 100,虽然也有传输速率为 133 MBps 的 ATA - 133,但并没有得到芯片组的普遍支持。并行 ATA 使用带状的并行电缆,虽然数据传输速率可达 100 MBps,但也产生了信号定时、电磁兼容以及电缆完整性等各种问题。这些问题在一种称为串行 ATA (Serial ATA, SATA) 的新型接口中得到了很好的解决 SATA 接口的标志如图 6 - 18 所示。



图 6 - 18 SATA 图标

6.7.1 从 ATA 到 SATA

SATA 标准对于 ATA 标准保持向后兼容,之所以这样说,是因为它与现有的软件保持兼容,这些软件可以不加修改地运行在这种新的体系结构上。换句话说,现有的 BIOS、操作系统以及在并行 ATA 上运行的各种工具都可以在 SATA 上很好地运行;这意

味着 SATA 支持所有 ATA 和 ATAPI 设备,包括 CD - ROM、CD - RW 驱动器、DVD 驱动器、磁带设备以及并行 ATA 支持的任何存储设备。

当然,两者在物理上是全然不同的,即不可能将 ATA 设备插入到 SATA 宿主适配器中,反之亦然。SATA 在物理上的改进有很大好处,这是因为 SATA 使用的电缆非常细,连接器只有 7 个引脚,这样使连接变得很简单,而且可以很容易地插入到更小的电缆连接器中。接口芯片的引脚数也减少了,使用的电压也降低了。这些改进都是针对在设计并行 ATA 时遇到的问题而进行的。

虽然 SATA 不会在一夜之间完全集成到 PC 机中去,但它最终会取代 ATA 作为 PC 机外存储设备接口的工业标准。从 ATA 到 SATA 的转变是一个逐步的过程,在这个转变过程中,ATA 仍然可以使用。估计在未来的 10 年内,并行 ATA 仍然可以继续使用,而越来越多的 PC 系统也开始使用 SATA。

SATA 标准的开发始于 2000 年 2 月。当时 Intel 开发者论坛宣布成立一个 SATA 工作组,最初的成员有 APT Technologies、Dell、IBM、Intel、Maxtor、Quantum 以及 Seagate。第一个 SATA 规范即 1.0 版本于 2000 年 11 月发布。

6.7.2 SATA 简介

SATA 性能优越,是取代 ATA 的理想接口。到目前为止,它有 3 个版本,它们都使用同样的电缆和连接器,只是传输速率不同。表 6 - 12 列出了 SATA 规范的 3 个版本。

表 6 - 12 SATA 规范

SATA 类型	总线宽度 (b)	总线频率 (MHz)	数据周期/时钟周期	带宽 (MBps)
SATA - 150	1	1 500	1	150
SATA - 300	1	3 000	1	300
SATA - 600	1	6 000	1	600

从上表中可以看出,SATA 一次只传输 1 位数据。但它的传输速率很高,这是采用串行传输的优点,可以大幅度地提高传输速率。

数据线缆只有 7 根线,而且非常细,两端的键控式连接器只有 14 mm 宽,如图 6 - 19 所示。每一根电缆只在两端有连接器,可以直接将设备连接到宿主适配器 (通常在主机板上) 上。因为每条电缆只支持一个单独的设备,所以也没有 ATA 接口的串/并设置问题。电缆两端是可以互换的,即主机板上的连接器与设备的连接器是相同的,电缆的两端也是完全相同的。SATA 电缆的最大长度是 1 m,比 ATA 的最大长度 0.45 m 要长很多。尽管这种电缆又细又长又便宜,但其传输速率却很高,最初为 150 MBps,之后发展到 300 MBps 和 600 MBps。

除了可以通过工业标准的 4 引脚设备电源连接器支持 5 V 和 12 V 电压以外,SATA 还提供了一种能够支持 3.3 V 的 15 引脚电源电缆和连接器。尽管这种新的电源连接器有 15 引脚,但其宽度只有 24 mm;其中 3 组电源分别是 3.3 V、5 V 和 12 V (见表 6 - 13),每一级电压都可以提供 4.5 A 的电流。这对于最耗电的驱动器来说都是足够的。为了与现有电源设备兼容,SATA 驱动器既可以做成最初标准的 4 引脚设备电源连接器,也可以做成新的 15 引脚 SATA 电源连接器,或者两者兼有。

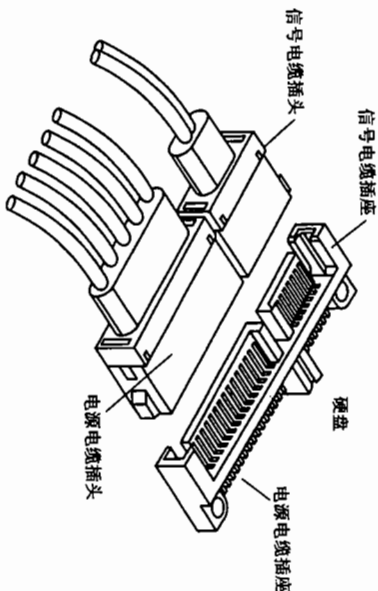


图 6-19 SATA 硬盘信号和电源连接器

表 6-13 SATA 的电源连接器引脚

引脚	信号	说明
P1	+3.3 V	+3.3 V 直流电源
P2	+3.3 V	
P3	+3.3 V	
P4	GND	
P5	GND	地线
P6	GND	
P7	+5 V	
P8	+5 V	
P9	+5 V	+5 V 直流电源
P10	GND	
P11	GND	
P12	GND	
P13	+12 V	+12 V 直流电源
P14	+12 V	
P15	+12 V	

SATA 对在电缆上传输的数据使用了一种称为 8b/10b 的特殊编码/解码方法。8b/10b 传输码最初由 IBM 在 20 世纪 80 年代发明,主要用于高速数据通信。现在,这种编码方法也用于多种高速数据传输标准中。8b/10b 编码方法的主要目的是要保证不能传输连续 4 位 0 (或 1),这是一种称为“RLL 0,4”的行程长度受限 (Run Length Limited) 编码方式,其中 0 代表编码字符中连续 0 的最小个数,而 4 代表连续 0 的最大个数。

8b/10b 编码方式也保证了在一个 10 位编码字符中,不会有超过 6 个 0 (或 1) 或者不会少于 4 个 1 (或 0)。由于 1 和 0 是以电压变化的方式在线路上传输,因此这种编码方式保证了电压变化的间隔非常均匀,因而脉冲波形也很规整平稳。这就为电路提供了一个稳定的负载,从而提高了可靠性。在从 8 位数据到 10 位编码字符的转变中,还有几个 10 位编码没有使用,这些多余的字符编码用于提供流量控制、数据包定界、差错检验以及其他特殊用途。

SATA 使用一种称为差分 NRZ (Non-Return to Zero, 不归零制) 的物理传输方式。这种方式使用一对线路,按平衡方式传输,电压为 +0.25 V 或 -0.25 V。如果一根线路为 +0.25 V 电压,另一根则是 -0.25 V 电压,两根线路的电压差始终保持为 0.5 V。这意味着如果给定一根线路的电压波形,则另一根的电压波形将与此相反。这种差分传输方式最大限度地减少了电磁辐射,而且有很强的抗干扰能力。

6.8 SCSI 接口

小型计算机系统接口 (Small Computer Standard Interface, SCSI) 是一种能连接多种类型的设备连接到计算机的通用接口。在 PC 领域,SCSI 是将高速磁盘驱动器连接到诸如工作站或网络服务器这样的高档 PC 机的常用接口。

6.8.1 SCSI 概述

SCSI 非常灵活,它不仅是磁盘接口,还是个系统级的接口,允许连接许多不同种类的设备。SCSI 是一条总线,可以支持总共 7 个或 15 个设备,如图 6-20 所示。

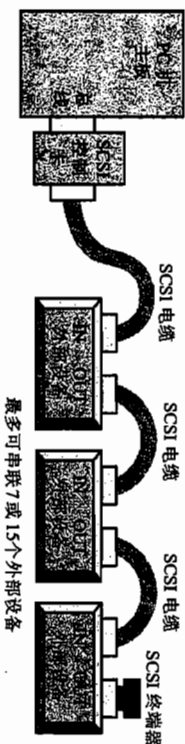


图 6-20 PC 机扩展 SCSI 连接

SCSI 控制器称为主适配器,其作用就像 SCSI 总线和 PC 系统总线之间的网关。总线上的每

一个设备有一个内置的控制器。SCSI 总线不直接与如硬盘之类的设备通信,而与驱动器中内置的控制器通信。

一条 SCSI 总线可以支持多达 8 或 16 个物理单元,每一个单元都需要一个标识符(ID),其中主适配器占用一个,其余 7 个或 15 个分配给其他的外部设备,可将 SCSI 接口的硬盘、磁带驱动器、CD-ROM 驱动器、图形扫描仪或其他设备连接到一个 SCSI 主适配器上。大多数系统可支持 4 个主适配器,如果每一个最多连接 15 个设备,总共可连接 60 个设备!

SCSI 接口的外部连接采用的是菊花链方式。SCSI 设备都有 IN 和 OUT 两个连接器插座,其中 IN 用于连接主机端电缆,OUT 用于连接下一个 SCSI 外部设备。如果没有下一个外部设备,应在 OUT 端口插上 SCSI 终结器。

如果在 PC 机上使用 SCSI 设备,由于芯片组一般不支持 SCSI,主板上没有 SCSI 接口,所以需要总线扩展卡的方式扩展 SCSI 接口。

SCSI 外部设备通常包括设备本身、SCSI 控制器和 SCSI 适配器,只有这样,设备才是完整的,可以直接插到 SCSI 总线上。这种类型的设备通常称为内嵌式 SCSI 设备(SCSI 接口是内置的)。例如,大多数 SCSI 硬盘驱动器除了将 SCSI 总线适配器电路(通常是单个芯片)附加到控制板上外,在技术上与 IDE 驱动器是一样。人们并不需要知道在 SCSI 驱动器内部是哪一种控制器,因为系统并不直接与控制器通信,虽然看起来好像与标准 IDE 驱动器一样插在系统总线上,但实际上是与装在系统总线上的 SCSI 主适配器进行通信。只有在 SCSI 协议下才能访问 SCSI 驱动器。

6.8.2 SCSI 电缆和连接器

SCSI 有两种电气上不同的版本,单端和差分,这两种版本在电气上不兼容,不能互连。在 PC 领域很少有差分 SCSI 的应用,所以很少会遇到这种问题。对每一种电气类型,有两种基本的 SCSI 电缆类型:A 型电缆(8 位标准 SCSI)和 P 型电缆(16 位宽 SCSI)。50 引脚 A 型电缆用于大多数 SCSI-1 和 SCSI-2 的安装,是常用的电缆。宽 SCSI 应用中采用具有 68 引脚的 P 型电缆。通过特殊适配器可以互连 A 型和 P 型电缆,在一根总线上可以混合连接 8 位标准和 16 位宽 SCSI 设备。

SCSI 电缆是特别屏蔽的,电缆中心层传送最重要的高速信号,不太重要的或慢速的信号放在外面两层上。这种特殊的构造使得 SCSI 电缆很昂贵,并且比其他类型的电缆要粗。不过要注意,这种特殊结构只是用于外部 SCSI 电缆。用于连接在屏蔽罩(如机箱)内部设备的电缆可以采用非常廉价的扁平电缆。

6.8.3 SCSI 的版本

由于发展的历史比较长,SCSI 有多个版本。SCSI 是美国国家标准学会(American National Standards Institute, ANSI)制定的标准。目前 SCSI 的最高传输速率是 320 MBps。表 6-14 简要列出了 SCSI 的各个版本的基本参数。

表 6-14 SCSI 的版本

SCSI 标准	模式	时钟频率 (MHz)	传输宽度 (b)	传输速率 (MBps)	外部设备 数目	电缆类型
SCSI-1	异步	5	8	4	7	A 型
	Fast-5	5	8	5	7	A 型
	Fast-10	10	16	10	15	P 型
SCSI-2	宽	10	8	10	7	A 型
	Fast-10	10	16	20	15	P 型
	Ultra	20	8	20	7	A 型
SCSI-3	Fast-20	20	16	40	15	P 型
	Ultra2	40	8	40	7	A 型
	Ultra2/宽	40	16	80	15	P 型
	Fast-80DT	40	16	160	15	P 型
	Fast-160DT	80	16	320	15	P 型

1. SCSI-1

SCSI-1 是第一个对 SCSI 的实现,其主要特性如下:

- 8 位并行总线。
- 5 MHz 异步或同步操作。
- 4 MBps(异步)或 5 MBps(同步)的传输速率。
- 带有低密度内部连接器和中央化外部连接器的 50 引脚电缆。
- 单端(Single End, SE)信号非平衡传输。
- 无源终端。
- 可选的总线奇偶校验。

SCSI-1 现在已不再使用。该标准已被 ANSI 作废,并由 SCSI-2 取而代之。

2. SCSI-2

SCSI-2 规范是 SCSI-1 的改良版本,对 SCSI-1 规范中的某些部分进行了压缩,并添加了一些新特性和可选项。SCSI-1 和 SCSI-2 的设备是兼容的,但 SCSI-1 设备不具有 SCSI-2 设备中的附加特性。

SCSI-2 增加了若干新的特性:

- 必选的总线奇偶校验。
- 可使用快速 SCSI(10 MHz)。
- 可使用宽的 SCSI(16 位传输)。
- 可采用高密度 50 引脚电缆连接器。
- 用于改良的单端传输的活动终端。

SCSI 与 IDE 比较: 硬盘连接设备
但送介

- 用于延长总线长度的高压差分(High Voltage Differential, HVD)信号传输。

宽 SCSI 总线上的并行数据传输位宽为 16 位。较宽的连接需要新型电缆。标准的 50 引脚 8 位电缆称为 A 型电缆, P 型电缆指的是具有 68 引脚连接器的 SCSI 电缆。如果需要连接宽 SCSI 设备并工作在 16 位模式下, 那么需要使用 P 型电缆。

3. SCSI-3

术语 SCSI-3 用来描述当前仍在发展中的一组标准。简而言之, 它是未来产品要遵循的下一代标准。与 SCSI-1 和 SCSI-2 不同的是, SCSI-3 不是一个包含 SCSI 所有的层和接口的文档, 而是一个包含着主要命令、特殊命令集和电气接口及协议的文档集合。其命令集包括硬盘接口命令、磁带驱动器的命令、RAID(廉价冗余磁盘阵列)的控制命令及其他命令。其中还包括一个物理和电气接口的总体 SCSI 体系结构模型以及一个用于控制最常用的 SCSI 形式的 SCSI 并行接口(SCSI Parallel Interface, SPI)。

SCSI-3 附加的主要内容包括:

- Ultra2(Fast-40)SCSI。
- Ultra3(Fast-80DT)SCSI。
- Ultra4(Fast-160DT)SCSI。
- 新的低压差分(Low Voltage Differential, LVD)信号。
- 去掉了高压差分(HVD)信号。

SCSI-3 的主要修改包括新的 Fast-40(Ultra 2)和 Fast-80DT(Ultra 3)高速驱动器及适配器, 这些特性令 SCSI 性能提高至 160 MBps。新增内容还有低压差分(LVD)电气标准, 可延长电缆长度。老式的高压差分(HVD)信号已从标准中删除。

关于 SCSI-3 传输速率的度量要同时考虑时钟频率以及传输速率。8 位传输指每次传输一个字节(8 位), 因此, 如果时钟频率是 40 MHz(Fast-40 或 Ultra2 SCSI), 那么传输速率就是 40 MBps。如果使用 16 位传输而时钟频率仍保持在 40 MHz, 那么传输速率可达到 80 MBps。对于 Fast-80DT(DT: Double Transition), 总线时钟频率仍维持在 40 MHz, 但每个时钟周期传输两次, 速率可达 160 MBps。同样地对于 Ultra4 SCSI, 总线时钟频率为 80 MHz, 一次传输两个字节, 而每个时钟周期传输两次, 传输速率为 320 MBps。Ultra4 也称为 Ultra320, 这是目前最快的并行 SCSI。

6.8.4 SCSI 到 SAS

2001 年底, Compaq、IBM、LSI Logic、Maxtor 和 Seagate 宣布成立 Serial Attached SCSI 工作组, 目标是将并行 SCSI 与 SATA 的优点相结合, 定义一个新串行点对点的企业级存储设备接口。这个接口标准就是 SAS, 其标志如图 6-21 所示, 也可以理解为 SAS 是 SCSI 发展的串行化, 与 ATA 到 SATA 的发展道路是一样的。



图 6-21 SAS 标志

SCSI 的优点在于完善的命令集, SATA 的长处则是串行的物理层, Serial Attached SCSI 正是将二者结合在一起。同样的物理互连部分让串行连接 SCSI(Serial Attached SCSI, SAS)支持 SATA

设备成为可能(反之则不行)。

SAS 的域、设备及接口完全继承了 SCSI 和 ATA 的相应概念。虽然接口和线缆的电气规格取自 SATA, 但点对点连接的距离延长至 10 m, 而连接对象也不再局限于启动设备(可理解为适配器)或目标设备(可理解为硬盘或光驱等), 还可以是扩展设备。扩展设备可以理解为路由器, 它具有至少两个外部接口。表 6-15 是 SATA、SCSI 和 SAS 的比较。

表 6-15 SATA、SCSI 和 SAS 的比较

参数	SATA	SCSI	SAS
设备寻址数量	1	16	128
传输距离(m)	1	25	10
双端口支持	否	否	是
拓扑结构	点对点	共享总线	点对点/扩展器
传输速率(Mbps)	150/300/600	320/640	150/300/600
通信方式	半双工	半双工	全双工
协议	ATA	SCSI	SCSI

6.9 红外线接口

实现无线传输, 一直是人们的梦想, 红外技术首先将人们的这一梦想变成现实。IrDA 是一种利用红外线进行点对点通信的技术, 在当前的所有类似的短距离无线应用中, IrDA 是最广泛的技术。

红外数据协会(Infrared Data Association, IrDA)成立于 1993 年, 它是一个独立的组织, 其宗旨是为可互操作、低费用的红外数据通信制定标准。最开始采用 IrDA 标准的无线设备仅能在 1 m 范围内以最高 115.2 Kbps(SIR)的速率传输数据, 不久 IrDA 的传输速率发展到 4 Mbps(FIR)以及 16 Mbps(VFIR)。

6.9.1 红外线传输的特点

红外通信属于无线通信, 无需电路连接。和其他的无线通信方式相比, 红外通信无需专门申请特定频率的使用许可, 具有移动通信设备所必需的体积小和功率低的特点, 在点到点的连接下保密性强, 速率最高可以达到 16 Mbps, 使得对于要求传输速率高、使用次数少、移动范围小、价格比较低的设备, 如打印机、扫描仪、数码相机等, IrDA 技术成为一种选择。

然而, 通信距离短、通信过程中不能移动、遇障碍物时通信中断、功能单一和扩展性等缺点, 使红外技术明显受到蓝牙和 Wi-Fi 等无线通信技术的挑战。IrDA 并没有停滞不前, 标准已经从原来的 FIR 提高到最新的 VFIR, 传输速率由 4 Mbps 提高到 16 Mbps, 接收角度也由传统的 30°扩展到 120°。这样在台式计算机上采用低功耗、小体积、移动余地较大的含有 IrDA 接口的键

盘、鼠标,就有了基本的技术保障。

不是所有的计算机都有 IrDA 接口, IrDA 只是在笔记本电脑和掌上计算机上比较普及,在台式计算机上很少配置。如果要在台式计算机上扩展 IrDA,简单的方法是通过 USB 转换器实现,如图 6-22 所示。



图 6-22 USB 接口 IrDA 适配器

6.9.2 IrDA 简述

红外数据协会的第一版本, IrDA 1.0, 提供了 115.2 Kbps 的数据传输速率。IrDA 1.1 把数据传输速率提高到 4 Mbps, 同时保持与版本 1.0 产品的兼容。IrDA 标准的关键是制定红外通信协议。红外连接本身并不可靠, 不同于有线连接以及半永久的插接式连接, 红外连接终端可能自由移动, 甚至会在传输的过程中因被遮挡而中断连接。红外数据协会定义了一套协议, 或者说是一个协议栈, 提供了建立和保持连接的标准, 以尽可能减少通信错误。

协议的最底层——物理层, 定义了设备通过红外线进行连接时的距离、相对角度要求。协议规定红外线传输距离为 0 m ~ 1 m, 相对角度为 $0^{\circ} \sim 15^{\circ}$, 为保证在各种品牌和类型的设备之间的物理互连, 规定了包括红外线光谱范围、调制方式、发射端的发光强度、接收端的接收灵敏度及动态范围、干扰光强度等。

红外数据协会规定的数据传输速率的范围为 2.4 Mbps ~ 4 Mbps。在此数据传输速率下, 对于传送器的光强度和接收器的敏感性做了规定, 保证 0 m ~ 1 m 的连接能够正常工作。选择接收器的敏感性是为了在 1 m 远处微弱的光线能正确接收, 而在距离最近处最强的光线也不会使接收器饱和。

实际使用的通信速率是设备双方通过信号联络商定的。连接开始时, IrDA 一般要求以 9.6 Kbps 的速率开始, 但是如果两端终端支持的话, 随后的传输速率可以提高。数据传输速率不达到 4 Mbps 时使用归零反向码 (Return to Zero Inverted, RZI) 调制, 而 4 Mbps 的数据传输速率使用脉冲定位调制 (Pulse Position Modulation, PPM)。

1. SIR

图 6-23 为数据传输速率最高为 115.2 Kbps 的 IrDA 1.0 版的物理层的结构示意图。串行红外线 (Serial InfraRed, SIR) 是建立在异步串行收发器 UART 的基础上的, 就是在标准的异步串行接口上外接一个红外收发器, 它是串行接口的直接扩展。数据在作为红外线脉冲传送前首先被编码。因为异步收发器使用的是不归零码, 而红外传送使用归零码, IrDA 要求能够反向归零的调制脉冲来点亮发光二极管 (LED), 这样的方式可以减小传送功率。SIR 标准要求的脉冲宽度是 3/16 个位宽周期。对于 115.2 Kbps 的传输速率, 最小脉冲宽度允许 1.41 μs , 如图 6-24 所示。

2. FIR

IrDA 1.1 版本把数据传输速率提高到 4 Mbps, 也被称为快速红外线 (Fast InfraRed, FIR), 并且与 IrDA 1.0 的产品保持兼容。因为任何一次连接都被要求必须从 9.6 Kbps 开始, 然后把传输速率提高到两边的接收终端都支持的最高传输速率。这样的话, 一个支持 FIR 的产品也必须能够工作在 SIR 的 9.6 Kbps 的速率下。IrDA 仅仅要求连接时传输速率能够达到 9.6 Kbps, 更高的速率是可选择的。因此, 一个速率 4 Mbps 的产品也能够和速率为 9.6 Kbps 的产品进行通信。

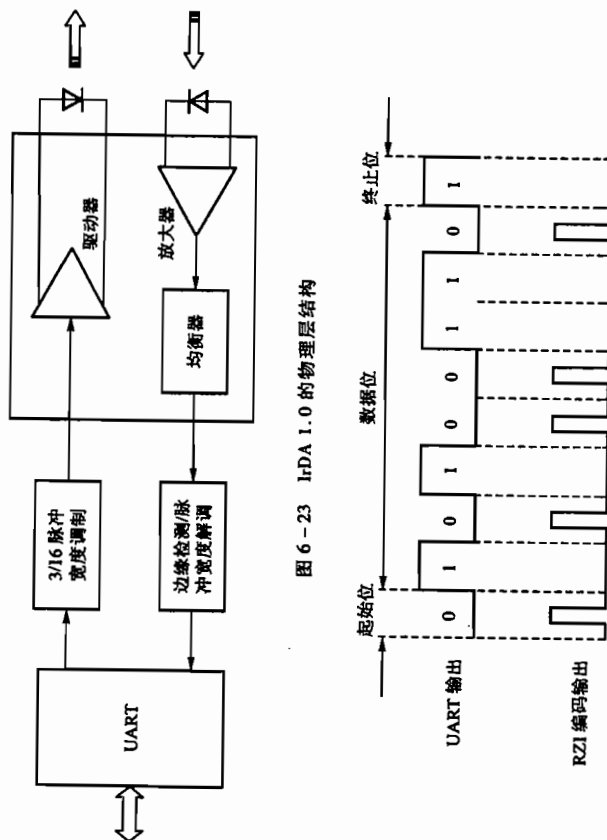


图 6-23 IrDA 1.0 的物理层结构

图 6-24 SIR 的 RZI 编码

一个 4 Mbps 的红外线连接使用四脉冲位置调制 (4Pulse Position Modulation, 4PPM) 电路来代替 SIR 使用的 3/16 调制电路。所谓 4PPM 就是在在一个时隙内通过脉冲的位置来传送信息。如图 6-25 所示, 两个数据位结合起来构成了一个 500 ns 的“数据位对”。这个数据位对被分成 4 个 125 ns 的时隙, 两个位能够编码成 4 个状态中的一个, 如 00、01、10、11。这取决于这些状态中哪个是当前状态, 单个的脉冲能够放在第一个、第二个、第三个或第四个 125 ns 时隙的任何一个中。这样一个解调器能够在输入的比特流相位同步之后, 在 500 ns 时隙内通过脉冲的位置来检测数据模式。

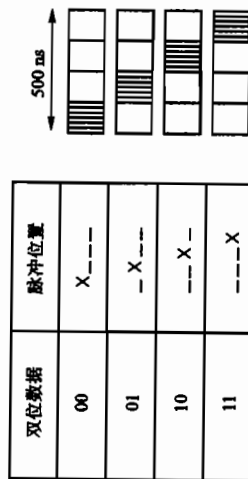


图 6-25 4PPM

6.10 无线接口

红外接口的优点是造价低廉,技术简单,保密性好,速度快等。但红外接口也有很严重的缺陷,主要它是不支持移动,这是因为红外通信有方向性,通信双方必须在相互的发送方向的有效角度之内;另一方面,红外通信的距离比较短,不能穿透墙壁,只能在可视空间内通信。再有就是IrDA只支持点对点和对多点连接,不支持多点连接。这些缺陷主要来源于红外通信的固有特点。无线通信方式可以克服这些缺陷。

6.10.1 蓝牙

Ericsson 公司 1994 年开始研究一种能使手机与其附件(如耳机)之间互相通信的无线模块。4年后, Ericsson、Nokia、IBM 等公司共同推出了蓝牙(Bluetooth)技术,主要用于通信和信息设备的无线连接。

蓝牙技术利用短距离、低成本、低成本的无线连接替代了电缆连接,从而为现存的数据网络和小型的外围设备接口提供了统一的连接。它具有许多优越的技术性能,以下介绍它的一些主要技术特点。

1. 射频特性

蓝牙设备的工作频段选在全世界范围内都可以自由使用的 2.4 GHz 的 ISM(工业、科学、医学)频段,这样用户不必经过申请便可以在 2.4 GHz ~ 2.5 GHz 范围内选用适当的蓝牙无线电收发器频段。蓝牙采用无线电跳频通信,频段采用 23 个或 79 个频点,间隔均为 1 MHz,采用双工方式。蓝牙的无线电发射机采用调频(FM)方式,从而能降低设备的复杂性。最大发射功率分为 3 个等级,即 100 mW(20 dBm)、2.5 mW(4 dBm)和 1 mW(0 dBm),设备之间的有效通信距离大约为 10 m ~ 100 m。

2. TDMA 结构

蓝牙的数据传输速率为 1 Mbps,采用数据包的形式按时隙传送,每时隙 0.625 μ s。蓝牙系统支持实时的同步定向连接和非实时的异步不定向连接,蓝牙技术支持一个异步数据通道、3 个并发的同步语音通道或 1 个同时传送异步数据和同步语音的通道。每一个语音通道支持 64 Kbps 的同步语音,异步通道支持的最高速率为 721 Kbps,反向应答速率为 57.6 Kbps 的非对称连接,或者是速率为 432.6 Kbps 的对称连接。

3. 使用跳频技术

跳频是蓝牙使用的关键技术之一。对应单时隙包,蓝牙的跳频速率为 1 600 跳/秒;对于多时隙包,跳频速率有所降低;但在建立连接时则提高为 3 200 跳/秒。使用这样高的跳频速率,使得蓝牙系统具有足够高的抗干扰能力。

作为一种电缆替代技术,蓝牙具有低成本、高速率的特点,它可把内嵌有蓝牙芯片的计算机、手机和多种便携通信终端互连起来,为其提供语音和数字接入服务,实现信息的自动交换和处理,并且蓝牙的使用和维护成本要低于其他任何一种无线技术。目前蓝牙技术开发重点是多点连接,即

一台设备同时与多台其他设备互连。今后,市场上不同厂商的蓝牙产品将能够互连互通。

蓝牙技术的应用主要有以下 3 类:

• 语音/数据接入:是指将一台计算机通过安全的无线链路连接到通信设备上,实现与广域网的连接。

• 外部设备互连:是指将各种设备通过蓝牙链路连接到主机上。

• 个人局域网(Personal Area Network, PAN):主要用于个人网络与信息的共享与交换。蓝牙产品涉及 PC、笔记本计算机、移动电话等信息设备和 AV 设备、汽车电子、家用电器和工业设备领域。蓝牙在个人局域网中获得了很大的成功,应用包括无绳电话、PDA 与计算机的互连、笔记本计算机与手机的互连以及无线 RS-232C、RS-485 接口等。

采用蓝牙技术的设备使用方便,可自由移动。与无线局域网相比,蓝牙无线系统更小、更轻薄,成本及功耗更低,信号的抗干扰能力更强。

6.10.2 Wi-Fi

Wi-Fi(Wireless Fidelity,无线高保真)也是一种无线通信技术,正式名称是 IEEE 802.11b,与蓝牙一样,同属于短距离无线通信技术。Wi-Fi 速率最高可达 11 Mbps。虽然在数据安全方面比蓝牙技术要差一些,但在电波的覆盖范围方面却略胜一筹,可达 100 m 左右,不用设家庭、办公室,就是小一点的整栋大楼也可使用。

1990 年 IEEE 802 标准化委员会成立 IEEE 802.11 无线局域网标准工作组。该标准定义了物理层和媒体访问控制(Media Access Control, MAC)规范。1999 年,IEEE 802.11a 标准制定完成,该标准规定无线局域网工作频段在 5.15 GHz ~ 5.825 GHz,数据传输速率达到 54 Mbps,传输距离控制在 10 m ~ 100 m。IEEE 802.11a 采用正交频分复用(Orthogonal Frequency Division Multiplexing, OFDM)的扩频技术,可提供 25 Mbps 的无线 ATM(Asynchronous Transfer Mode,异步传输模式)接口、10 Mbps 的以太网无线帧结构接口以及 TDD(Time Division Duplex,时分双工)/TDMA(Time Division Multiple Access,时分多址)空中接口,支持语音、数据及图像传输业务。

1999 年 9 月,IEEE 802.11b 被正式批准,该标准规定无线局域网工作频段为 2.4 GHz ~ 2.4835 GHz,数据传输速率达到 11 Mbps。该标准是对最初的 IEEE 802.11 标准的一个补充,采用点对点和基本模式两种运作模式,其数据传输速率可以根据实现情况在 11 Mbps、5.5 Mbps、2 Mbps 与 1 Mbps 之间自动切换,而且在传输速率为 2 Mbps 与 1 Mbps 时与 IEEE 802.11 兼容。IEEE 802.11b 使用直接序列扩频(Direct Sequence Spread Spectrum, DSSS)作为协议。IEEE 802.11b 和工作在 5 GHz 频率上的 IEEE 802.11a 标准不兼容。由于价格低廉,IEEE 802.11b 产品已经被广泛地投入市场,并在许多实际工作场所运行。

Wi-Fi 技术的最具诱惑力的方面在于将 Wi-Fi 与基于 XML 或 Java 的 Web 服务融合起来之后,可以大幅度减少企业的 IT 成本。例如,许多企业选择在每一层楼或每一个部门配备 802.11b 的接入点,而不是采用电缆线把整幢建筑物连接起来。这样一来,可以节省铺设大量电缆所需的费用。

WLAN 未来最具潜力的应用将主要是在家办公(Small Office Home Office, SOHO)、家庭无线网络以及不便安装电缆的建筑物或场所。目前这一技术的用户主要来自机场、酒店、商场等公共热点场所。

实验

1. 通过 Windows 2000/XP 的设备管理器查询一台计算机的串行接口和并行接口的参数;实现接口的南桥芯片(或 ICH)型号、接口参数和接口使用的系统硬件资源。
2. 通过 Windows 2000/XP 的设备管理器查询一台计算机的外存储器接口 (IDE 或 SATA) 参数;实现接口的南桥芯片(或 ICH)型号、IDE 或 SATA 接口各个通道使用的传输模式和接口使用的系统硬件资源。
3. 通过 HARDINFO 软件检测一台计算机的资源分配情况;中断通道和 DMA 通道的分配信息。

184 第6章 接口

Microsoft 的桌面操作系统 Windows XP 和嵌入式操作系统 Windows CE, 都包含了对 Wi-Fi 的支持。其中, Windows CE 同时还包含对 Wi-Fi 的竞争对手蓝牙等其他无线通信技术的支持。

由于投资 IEEE 802.11b 产品的费用较低,许多厂商介入这一领域。Intel 推出了集成 WLAN 技术的笔记本电脑芯片组,不用外接无线网卡,就可实现无线上网。国内的联想、清华同方、方正等公司都推出无线网卡等无线网络解决方案。

最新的 IEEE 802.11g 标准是对流行的 IEEE 802.11b(即 Wi-Fi 标准)的提速(传输速率从 11 Mbps 提高到 54 Mbps)。IEEE 802.11g 接入点支持 IEEE 802.11b 和 IEEE 802.11g 客户设备。同样,采用 IEEE 802.11g 网卡的笔记本电脑也能访问现有的 IEEE 802.11b 接入点和新的 IEEE 802.11g 接入点。不过,基于 IEEE 802.11g 标准的产品目前还不多见。IEEE 802.11a 的主要缺点是不能与 IEEE 802.11b 设备连通,而且与 IEEE 802.11b 相比,IEEE 802.11a 网卡比较昂贵,接入费用也比较高。

习题

1. 名词解释

DMA、DTE、DCE、FIFO、MODEM、SPP、EPP、ECP、BIOS、IRQ、NRZI、PnP、IDE、ATA、PIO、LBA、CHS、S.M.A.、R.T、SCSI、SAS、JrDA、PPM、SIR、FIR、Wi-Fi、WLAN

2. 说明接口、接口电路、端口 3 个名词的基本概念,它们之间有什么关系?
3. 与查询控制方式相比,中断控制方式有什么特点?简述中断控制方式的程序实现的基本方法。
4. 接口的平衡和非平衡信号方式有什么不同?RS-232C、并行接口、USB、IEEE 1394 分别采用什么信号方式?

5. RS-232C 中定义的 DTE 和 DCE 是什么?两者有什么不同?

6. 计算机中的 UART 是什么?UART 中的 FIFO 的作用是什么?

7. 从性能和特点两个方面比较并行接口的 EPP 和 ECP 两种模式。

8. USB 接口给出的通信速率是信号线的传输速率,实际的数据传输速率要低一些。哪些因素降低了实际数据传输速率?

9. 简述 USB 电源管理。

10. USB 提供的 4 种传输模式是什么?USB 接口的移动硬盘、音频设备、鼠标分别适合哪一种传输模式?

11. 简述 USB 采用的 NRZI 编码。编码后的数据位数有没有改变?

12. 在提高接口的传输可靠性方面,USB 采用了哪些措施?

13. 在外部设备的连接方式上,USB 和 IEEE 1394 有什么不同?

14. 简述 USB 和 IEEE 1394 在信号传输上的同步方法。

15. 简述 ATA-100 的性能和技术特点。

16. 与 ATA 相比,SATA 的主要优势是什么?

17. 简述 SCSI-3 的不同版本。

18. 什么是 SAS?SAS 采用的关键技术是什么?

19. 简述 iDA 的基本特点,说明 SIR、FIR 和 VFIR 的技术指标。

20. 客观的分析和比较 iDA 和蓝牙在实际应用中的限制因素。



第7章

外部存储器

计算机存储系统分为内部存储器和外部存储器两部分。与内部存储器相比,外部存储器的特点是非易失性、大容量,有些外部存储器的存储体甚至可更换。本章将分别介绍 PC 机中使用的硬盘、软盘、光盘以及移动存储系统。

7.1 硬盘

硬盘为计算机提供大容量的、可靠的、高速的外部存储手段,与其他外部存储系统相比,其速度是最快的,容量也是最大的。本节从硬盘的发展、结构、性能和使用几个方面介绍硬盘存储系统。

硬盘是计算机系统中最重要的部件之一。硬盘是一个密封单元,在 PC 中用作最主要的外部存储器。在这里,非易失性或永久性存储器,意味着即使没有给计算机供电,存储器仍能保持数据。因为硬盘能一直保持数据,直到用户故意擦除为止,所以 PC 将其用来存储程序和数据。正因为如此,当硬盘发生故障时,后果通常是非常严重的。为了保证计算机正常、可靠地工作,了解硬盘的工作原理、正确地使用和维护硬盘是很重要的。

7.1.1 硬盘的结构与工作原理

硬盘的基本物理构造是由旋转的盘片、磁头和电路部分组成,磁头在盘片上沿着径向方向移动并在磁道和扇区中读取或写入数据。磁头在称为磁道的同心圆上读/写数据,扇区是磁道的分段,一个扇区的存储容量是固定的。

1. 硬盘的组成

硬盘通常有多个磁盘盘片,按一定间隔同轴堆叠,盘片的两个面均用来存储数据。大多数硬盘至少有 2~3 个盘片,即有 4~6 个盘面。通常,硬盘驱动器每一盘面有一个磁头,所有的磁头安装在一个同步的传动机构上。磁头一致地沿盘面径向移动,因为它们安装在同一个支架上,所以某个磁头不能单独地移动。硬盘的内部结构如图 7-1 所示。

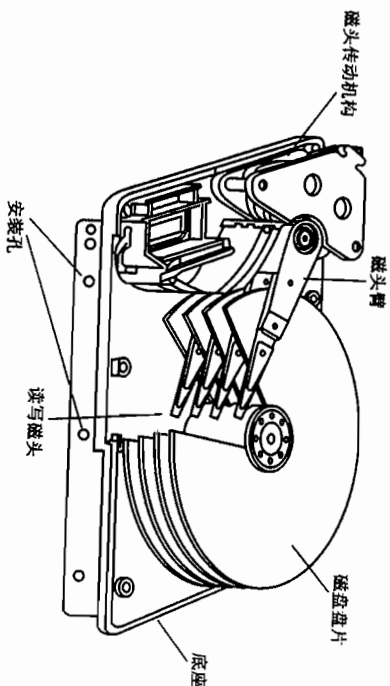


图 7-1 硬盘的内部结构

在正常工作情况下,大多数硬盘驱动器的磁头不接触盘面。当驱动器正在高速旋转时,非常薄的空气垫层使每个磁头悬浮在盘片之上或之下一个很小的距离。如果空气垫层受到灰尘颗粒或震动的干扰,磁头可能会触及高速旋转的盘片。这种接触事件称为磁头碰撞。磁头碰撞可能造成数据读写错误或磁盘表面的损伤。

盘片组件是密封的,不能拆卸,硬盘内部能够达到一生不染,加上高精度的磁头定位,盘片上磁道密度能做得非常高。现在,硬盘驱动器的磁道密度达几万 TPI(每英寸磁道数)。在绝对洁净的生产车间,包含盘片的磁头盘组被组装和密封。因为只有很少几家公司能修理磁头盘组,所以修理或更换其内部部件的费用十分昂贵。

2. 磁道、扇区和柱面

磁道是磁盘一个面上的单个数据存储圆圈。如果将磁道作为一个存储单元,从数据管理效率的角度来看容量太大。许多盘片的一个磁道能存储 100 000 字节甚至更多的数据,用于文件存储时的效率就太低了。因此,磁道被划分成若干个编号上号的圆弧段,称为扇区;整个磁盘的各个盘面的同一编号的磁道的集合叫做柱面,如图 7-2 所示。

不同类型的磁盘驱动器依据磁道密度的不同,将磁道划分成不同数量的扇区。例如,软盘格式化使用每磁道 8~36 扇区,而硬盘通常以更高的密度存储数据,每磁道可以划分为 17~100 个或更多的扇区。PC 系统中,通过标准格式化程序产生的扇区容量为 512 字节,这个数字在 PC 整个历史过程中一直没有变化过。

磁道上的扇区编号从 1 开始,不像磁头或柱面编号从 0 开始。例如,1.44 MB 软盘包含编号 0~79 的 80 个柱面和编号 0~1 的 2 个磁头,而每个柱面上的每个磁道有 18 个扇区,编号为 1~18。

所示。

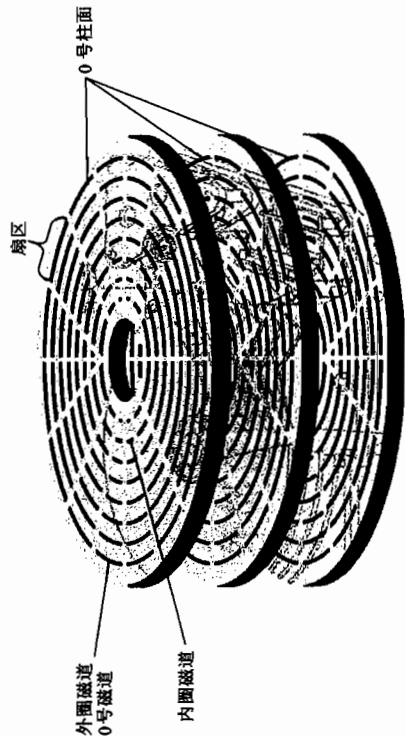


图 7-2 磁道、扇区和柱面

当磁盘被格式化后,格式化程序在每个扇区的数据前后创建一个标识区域,供磁盘控制器用来为扇区编号和辨认每个扇区的开始和结束。这些区域作为每个扇区的数据区域前缀和后缀,占用了磁盘总存储容量的一部分。这就是磁盘的未格式化容量和格式化容量不同的主要原因。注意,市场上销售的有些硬盘已经预格式化过了,制造商给出的标称值也是格式化后的容量。另一个有意义的发展是 IBM 等公司已制造出无记录驱动器,即不再在每个扇区的前后加上标记。这将使更多的磁盘空间用于实际的数据存储。

3. 记录格式

磁盘上的每个扇区有一个前缀部分,或叫扇区头,用以标志扇区的开始并含有扇区编号;还有一个后缀部分,或叫扇区尾,其中包括一个校验和,用于确保数据内容的完整性。每个扇区包含 512 字节的用户数据。低级格式化过程通常对数据字节填充一些特定的值。

扇区头和尾独立于操作系统、文件系统和存储在驱动器上的文件。除了扇区头和尾外,在扇区内部,同一磁道的扇区之间以及不同的磁道之间都还有间隙存在,但这些间隙没有可用的数据空间。这就类似于录音盒带上歌曲之间的无声音间隙。前缀、后缀和间隙计入磁盘未格式化和格式化后的容量损失中。例如,一个 4 MB 软盘 (3.5 英寸) 格式化后的容量是 2.88 MB, 一个 2MB 软盘格式化后的容量是 1.44 MB。

严格地说,前面所说的每个磁盘扇区的大小都是 512 字节,在技术上并不正确。每个扇区确实允许存储 512 字节的数据,但数据区域仅是扇区的一部分。磁盘上的每个扇区通常占据 571 字节,其中的 512 字节用来存储用户数据。扇区头和尾要求的实际数量是随驱动器的不同而变化的。目前,少数现代驱动器使用无 ID 记录方案,实际上免去了存储扇区头信息的额外开销。在无 ID 记录方案中,磁道上所有的空间都用于存储数据。

可以将每个磁盘扇区看作一本书中的一页,每一页有正文,但不是整页都充满了正文,至少每页的顶、底、左和右都留有边缘空白。一些像章节标题(磁道和柱面编号)和页号(扇区编号)

信息放置在边缘空白内。一个扇区的“边缘区域”是在低级格式化过程中产生和写入的。格式化过程也同时对每个扇区的数据区域写入模拟数据值。当在磁盘上执行高级格式化后,PC 的文件系统能对每个扇区的数据区域进行写入,但是扇区头尾信息在正常的写操作中不会被修改,除非再次对磁盘进行低级格式化。

表 7-1 列出了一个典型的硬盘上扇区的格式,此硬盘每个磁道有 17 个扇区,这是比较早期的硬盘,但其基本扇区和磁道结构与新式驱动器基本一样,只是后者每磁道中有更多的扇区。现代硬盘在每个磁道中安置的扇区多达 150 个或更多,这些扇区的具体格式可能与这个例子稍有不同。

表 7-1 典型磁盘磁道格式

大小(B)	数据段名称	说明
16	POST INDEX GAP	全为 4EH,在磁道开始处的索引标志之后
13	ID VFOLOCK	全为 00H,使 VFO 与扇区 ID 同步
1	SYNC BYTE	A1H,通知控制器后面是数据
1	ADDRESS MARK	FEH,定义后面跟随的是 ID 数据
2	CYLINDER NUMBER	定义柱面编号
1	HEAD NUMBER	定义特定磁头的编号
1	SECTOR NUMBER	定义扇区的编号
2	CRC	ID 数据的循环冗余校验
3	WRITE TURN-ON GAP	00H,格式化时写入,隔离 ID 数据和用户数据
13	DATA SYNC VFO LOCK	全为 00H,使 VFO 与 DATA 同步
1	SYNC BYTE	A1H,通知控制器后面是数据
1	ADDRESS MARK	F8H,定义后面跟随的是用户数据区
512	DADA	用户数据区
2	CRC	用户数据的循环冗余校验
3	WRIETURN-OFF GAP	00H,由数据更新形成,用以隔离用户数据
15	INTER-RECORD GAP	全为 00H,作为主轴转速变化的缓冲
693	PRE-INDEX GAP	全为 4EH,位于磁道末尾,索引标志之前

注:每扇区总字节数=571B;每扇区数据字节数=512B;

每磁道总字节数=571×17+16+693=10 416B;每磁道数据字节数=512×17=8 704B。

由上可见,每个磁道上数据的可用空间(例如为 8 704B)比未格式化之前的容量(例如为 10 416B)少了约 16%。这对大多数硬盘来说都是正确的,只是百分比可能稍有不同。

POST INDEX GAP:提供了一个磁头切换恢复阶段,所以从一个磁道切换到另一磁道时,磁头不需要等待磁盘做额外旋转就能连续地读扇区。因为磁盘是连续旋转的,磁头要用一小段时间沿半径方向从一个磁道移动到另一磁道,不可能在两个不同磁道间立刻读出连续的扇区。当磁头移到新磁道时,下一个扇区的头部已经转过去了。扇区之间留下的间隙为磁头移动到另一磁道

提供了时间。

有些驱动器中,此间隙不能提供足够的时间供磁头移动,出现这种情况时,可通过偏移不同磁道上的扇区为驱动器获得一段额外的时间,这样一来下一个扇区的到达时间被延迟了。换句话说,低级格式化过程偏移扇区编号,使各个磁道上相同编号的扇区不相互邻接,一个磁道上的9扇区可以和下一个磁道上的8扇区邻接,这个8扇区再邻接下一个磁道上的7扇区,如此等等。最佳偏移值取决于磁盘旋转速度与磁头径向移动速度。

磁头偏移曾经是一个做低级格式化时由用户自己设置的参数。现今的IDE和SCSI驱动器已在工厂做低级格式化并设置了最佳偏移值。扇区的ID数据包括柱面编号、磁头编号、扇区编号字段以及用以校验数据的循环冗余校验(Cyclical Redundance Checking, CRC)字段。大多数磁盘控制器在低级格式化或表面分析时,用磁头编号字段的第7位标注扇区损坏。然而,此约定不是绝对的。有些磁盘控制器用别的方法标注坏扇区,但通常都标注在某个ID字段中。

WRITE TURN-ON GAP:紧接着ID区域的CRC字段,它提供了填充位,以保证后面的用户数据域正确记录。

用户数据区包含存储在扇区中的全部512字节数据,后面紧跟一个CRC字段来校验数据。虽然许多控制器使用两个字节的CRC,但是控制器若要实现一个更长的纠错码(Error Correcting Code, ECC),就有多于两个字节的內容要存储。存储的ECC数据使检测出用户数据错误并纠正错误成为可能。纠错/检错能力取决于驱动器使用的ECC编码及其控制器的实现。

WRITE TURN-OFF GAP:是一个缓冲,使ECC(CRC)字节得以完全恢复。

INTER-RECORD GAP:提供了一种适应驱动器主轴速度变化的手段。当磁道被格式化时磁盘的转速可能正巧稍慢于正常速度,当磁道被写入数据时可能磁盘的转速稍快于正常速度。在这种情况下,此间隙可防止将任何信息意外地覆写到下一个扇区中。这个间隙的实际长度要依据磁道格式化和每次数据字段被更新时的磁盘转速而定。

PRE - INDEX GAP:考虑到整个磁道的速度容限,此间隙的长度取决于格式化时磁盘的转速和写频率容限。

扇区的前缀信息是极其重要的,因为它包含了柱面、磁头和扇区的编号信息。所以,这些内容中只能是在低级格式化时写入。

7.1.2 硬盘的性能指标

硬盘在PC系统中是必不可少的,同时也随着PC的发展而发展。下面介绍硬磁盘的主要指标以及这些指标在20多年的时间里的发展与变化。

1. 容量

容量永远是硬盘最重要的指标。硬盘最大存储容量从1982年时5.25英寸全高驱动器的10 MB,发展到现今的3.5英寸半高或更小驱动器的近千GB,笔记本系统中2.5/1.8英寸驱动器容量也达几百GB。容量小于10 GB的硬盘驱动器在当今的个人计算机中已经很少见了。硬盘容量的不断增大的主要因素是磁盘存储信息的面密度的增加。

面密度是硬盘驱动器工业界用于衡量硬盘容量增长率的标志。面密度被定义为沿磁道长度方向上每英寸位数(位密度)与沿径向方向上每英寸磁道数量(磁道密度)的乘积,单位用每平方

英寸兆位或吉位(Mb/in^2 或 Gb/in^2)来表示,作为驱动器记录技术的综合密度。

面密度 = 位密度 \times 磁道密度 (Mb/in^2 或 Gb/in^2)

面密度(或驱动器容量)约每2~3年要提高一倍。许多公司也正在开发新的记录介质和磁头技术来支持更高的面密度,例如,陶瓷或玻璃盘片、磁致电阻磁头、伪接触记录等。要获得更高的面密度,主要挑战是制造相互能靠得更近工作的驱动器磁头和磁盘。为了在给定尺寸的盘片上存储更多的数据,磁道之间必须更近一些,因此必须提高磁头定位磁道的精度。这意味着随着硬盘容量增加,必须减小磁头到盘面的悬浮距离。

2. 传输速率

磁盘的传输速率对整个计算机系统的性能有着重要的影响,但它也是一个容易让人误解的概念。问题源于目前的驱动器可以指定几个传输速率。多数驱动器生产商都可以报告最多5个传输速率。

接口传输速率:取决于接口的类型,如新的ATA-100接口为100 MBps。

媒体传输速率:从盘片读取数据流的速率,分为原始最大媒体传输速率、原始最小媒体传输速率。

格式化传输速率:从盘片上读取用户数据的速率,分为格式化最大媒体传输速率、格式化最小媒体传输速率。

媒体传输速率远比接口传输速率重要,这是因为媒体传输速率是真正的从磁盘上读取数据的速率,它反映的是从磁盘盘片上读取数据的快慢程度,是任何持续的传输希望达到的最大速率。

磁盘制造商都报告一个最小和一个最大媒体传输速率,因为现在的驱动器都有记录每个磁道不同扇区数目的分区记录方式,可能一个驱动器沿径向被分为16个区,最里侧分区的扇区数目约是最外侧分区扇区数目的一半,由于驱动器以固定速率旋转,从外侧的柱面(每个磁道有更多的扇区)读取数据,要比从里侧的柱面读取数据快得多。

另一个问题是原始传输速率与格式化传输速率。原始传输速率是指从媒体上读取比特流的快慢程度。由于并不是所有的比特都是数据,而且在磁头从一个磁道移向另一个磁道时可能会错过一些比特的数据,因而格式化传输速率表示的是从磁盘上读/写的真正的数据速率。

注意,一些厂商可能只报告一个原始传输速率,但格式化传输速率通常是原始传输速率的2/3,因而可以计算出格式化传输速率。同样,有些厂商可能只报告最大传输速率,由于最小传输速率是最大传输速率的一半,平均传输速率是最大传输速率的3/4,因而也可以计算出最小传输速率和平均传输速率。

影响传输速率指标的因素主要有两个:旋转速度和线性记录密度。如果两个驱动器的每个磁道具有相同的扇区数,旋转快的那个驱动器传输数据也快;同样,如果两个驱动器具有相同的旋转速度,具有较高记录密度的驱动器传输数据更快些。所以,传输速率和两个参数的乘积成正比。

3. 转速

最初,大多数硬盘以3600 rpm的速度旋转,约比软驱快10倍。这个速度在很长的一段时间里几乎没有变化过。然而,现在不少的硬盘驱动器的转速已经提高了。有的驱动器以

5 400 rpm、5 600 rpm、6 400 rpm、7 200 rpm、10 000 rpm、甚至 15 000 rpm 速度旋转。提高转速的最直接的目的就是提高数据传输速率。

4. 高速缓存程序和高速缓存控制器

在软件级别上,磁盘高速缓存程序如 SMARTDRV (DOS) 或 VCACHE (Windows) 会对磁盘驱动器性能产生重要的影响。这些程序进行 BIOS 硬盘驱动器中断,拦截应用程序和设备驱动程序对磁盘 BIOS 的读/写调用。当应用程序想从硬盘驱动器读取数据时,缓存程序拦截读请求,将读请求按通常的方式发送给硬盘控制器,将从硬盘上读出的数据保存在自身的高速缓存中,然后再把数据传送给应用程序。

根据高速缓存的大小不同,扇区上的大量数据可以被读入并保存在缓存中。当应用程序想读取更多数据时,缓存程序再次拦截请求并检查缓存以验证被请求的数据是否还在缓存里。如果在的话,缓存程序就将数据从缓存传给应用程序,而不需另一次的硬盘驱动器操作。由于缓存的数据存储在内存里,这种方法可以极大加速访问,对磁盘驱动器性能产生很大影响。

大多数驱动器现在都有某种形式的内置硬件缓冲区或高速缓存,它们使驱动器可以在硬件一级缓存数据;这对于通常的性能测量软件是不可见的。许多 ATA 和 SCSI 驱动器的缓存都直接位于驱动器电路板上控制器里。多数新的 ATA 驱动器都有 512 KB 到 2 MB 内置缓存,而许多 SCSI 驱动器的内置缓存则已达到了 16 MB。

5. 平均寻道时间

平均寻道时间通常以毫秒(ms)为单位度量,是指将磁头从一个柱面随机移到另一个柱面所需的平均时间。测量这个参数的一种方法是运行很多次随机寻道操作,然后将花费的时间除以执行的寻道次数,得到单次寻道的平均时间。

平均寻道时间只与驱动器本身有关,接口或控制器的类型对这个参数没有什么影响。它是磁头驱动机构能力的一个评判。

另一个参数是平均访问时间,它是平均寻道时间与另一个参数——延迟的和。延迟是从磁头到达一条目标磁道开始,到扇区可以开始访问所需的平均时间。一般情况下,这个值是磁盘旋转半圈所用的时间。所以,平均访问时间是指导驱动器可以访问到一个随机请求的扇区之前所需的平均时间。

延迟是关系到磁盘读/写性能的一个参数。减小延迟会提高对数据或文件访问的速度,而且只能通过更快地旋转驱动器盘片来完成。

6. S.M.A.R.T

S.M.A.R.T (Self-Monitoring, Analysis, and Reporting Technology, 自监测、分析和报告技术) 是一种提供磁盘驱动器故障预测的工业标准。当对给定驱动器启用了 S.M.A.R.T 后,驱动器监视一些对驱动器性能降低敏感的或起指示作用的预先确定的属性。基于被监测属性的变化,可以进行故障预测。如果认为可能发生故障, S.M.A.R.T 会生成一个状态报告以便系统 BIOS 或驱动程序向用户通知即将发生的故障,这样用户有可能在实际问题发生之前备份驱动器数据。

具有 S.M.A.R.T 能力的驱动器监测大量随制造厂家不同而不同的属性。设备制造商根据各种属性在预测特定的驱动器性能下降或错误条件时所起的作用大小来选择属性。大多数驱动

器制造商都把使用的一组特定属性和这些属性的标识视为与开发商相关的和专有的。

有的驱动器监测磁头在磁介质之上的悬浮高度,如果这个高度偏离适当的数值,驱动器会出错。其他的驱动器可以监测不同的属性,如纠错(ECC)电路(用于监测在读/写数据时是否发生错误)。在不同驱动器上监测的一些属性包括:

- 磁头悬浮高度。
- 数据吞吐率。
- 旋转加速时间。
- 重分配(空闲的)扇区数。
- 寻找错误的速率。
- 寻道时间。
- 驱动器旋转加速重试次数。
- 驱动器校准重试次数。

每个属性都有一个阈值,用来判断是否存在性能下降或错误条件。这些阈值由驱动器厂家设置,不能改变。

当被监测属性的变化足以激活一个 S.M.A.R.T 警告时,驱动器会通过 IDE/ATA 或 SCSI 命令向系统 BIOS 里的硬盘驱动程序发送一条警告消息,驱动程序把这条消息转发给操作系统。

用户收到这样一条警告后,首先要做的是对它加以注意并备份驱动器上的所有数据。将数据备份到新介质上而不要覆盖可能已存在的任何以前的正常备份,以免在备份操作完成之前驱动器就发生了故障。

7.1.3 分区和格式化

关于硬盘格式化,有两种格式化:物理的格式化(低级格式化)和逻辑的格式化(高级格式化)。硬盘需要两次独立的格式化操作,而且硬盘在这两次格式化过程之间还需要另一个步骤:将分区信息写到磁盘上。

硬盘被设计为可用于多个操作系统,所以分区是必需的。通过把物理格式化成一个个对于任何操作系统和高级格式化(这对不同操作系统是不同的)总是相同的过程。分区使一个硬盘驱动器可运行多种操作系统,或者允许一个操作系统将磁盘用作多个卷标或逻辑驱动器。因此,硬盘驱动器存储数据之前必须依次完成 3 个步骤:低级格式化、分区和高级格式化。

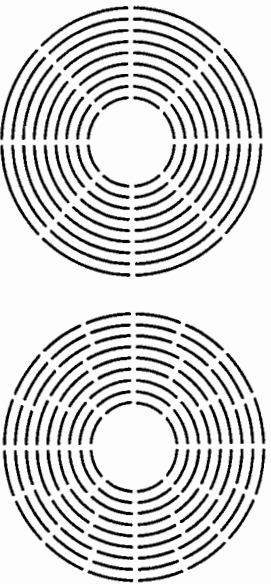
1. 低级格式化

在低级格式化过程中,格式化程序将磁盘的磁道划分成特定数目的扇区,创建扇区之间和磁道之间的间隔,并记录扇区头部和尾部信息。它还用一个空字节值或一个测试值来填充每个扇区的数据域。对于软盘,每条磁道上记录的扇区数与磁盘的类型和驱动器有关;对于硬盘,每条磁道的扇区数与驱动器和控制器接口有关。

最初,PC 硬盘驱动器使用一个单独的控制器,其形式为扩展卡或集成到主板上。因为控制器可用于各种磁盘驱动器,甚至可由不同的厂商制造,在控制器和驱动器之间的通信必须达成一致。由于这个原因,写入一条磁道的扇区数趋向于相对一致。

当前 PC 的 ATA/IDE 和 SCSI 驱动器里每条磁道上的扇区数可以是 17 ~ 900,甚至更多。

所有的 IDE 和 SCSI 驱动器实际上都使用一种称为分区记录的技术,在每条磁道上写可变数目的扇区。如果不利用分区记录,每条磁道上的扇区数以及数据位都是固定的,这意味着每英寸长度上的实际位数是可变的,在内磁道上每英寸的数据位更多,而外磁道上则少一些。数据的存取速率和旋转速度始终是不变的,每条磁道上的位数也不变。这就是标准记录格式。标准记录格式浪费了外磁道的空间,因为外磁道比内磁道长,但却拥有和内磁道同样的数据量,如图 7-3(a)所示。在低级格式化时,增加硬盘驱动器容量的一种方法是在磁盘的外柱面上建立比内柱面上更多的扇区。外柱面的周长更大,因此可以存储更多的数据。



(a) 标准记录方式

(b) 分区记录方式

图 7-3 标准的和分区记录方式

使用分区记录记录的驱动器将全部柱面分成若干组,称为区域。从磁盘的中心向外,各个连续的区域在每条磁道上的扇区数逐渐增多,如图 7-3(b)所示。一个特定区域里的所有柱面在每条磁道上的扇区数都相同。区域的个数随磁盘驱动器的不同而不同,大多数磁盘有 10 个以上的区域。

分区记录方式的另一个特点是传输速率随磁头所在区域的不同而变化。一个分区记录记录的驱动器仍以固定速率旋转,但是由于在外部区域里每条磁道有更多的扇区,因此这里的数据传输是最快的。相应地,在读/写内部区域的数据传输是最慢的。这就是为什么目前所有驱动器都报告最小和最大持续传输速率的原因,它们与用户在磁盘上读写数据的位置有关。

作为一个例子,表 7-2 列出了 IBM Travelstar 320GB 2.5 英寸笔记本计算机驱动器所定义的区域、各个区域里每条磁道上的扇区数以及数据传输率。

表 7-2 IBM Travelstar 320GB 2.5 英寸硬盘驱动器的分区记录信息

区域	每磁道扇区数	传输速率 (Mbps)	每磁道字节数 (B)	每区域扇区数
0	617	28.49	315 904	835 418
1	598	27.60	306 005	809 241
2	578	26.70	296 107	783 063
3	559	25.81	286 208	756 886
4	540	24.92	276 309	730 709

续表

区域	每磁道扇区数	传输速率 (Mbps)	每磁道字节数 (B)	每区域扇区数
5	520	24.03	266 411	704 531
6	501	23.13	256 512	678 354
7	482	22.24	246 613	652 177
8	462	21.35	236 715	625 999
9	443	20.46	226 816	599 822
10	424	19.56	216 917	573 645
11	404	18.67	207 019	547 467
12	385	17.78	197 120	521 290
13	366	16.88	187 221	495 113
14	346	15.99	177 323	468 935
15	327	15.10	167 424	442 758

这个驱动器在每一面磁盘上总共有 21 664 条磁道,这些磁道被分成 16 个区,每个区有 1 354 条磁道。第 0 区包含最外面的 1 354 条磁道,它是最长的,也含有最多扇区:617 个。每个扇区大小为 512 个字节,因此这个区的每条磁道可以存储 315 904 个字节的用户数据。而第 15 区的磁道含有 327 个扇区,只有 167 424 个字节。

因此,利用分区记录,这个磁盘驱动器的每个盘面包含 10 225 408 个扇区,存储容量为 5 235 MB。若不使用分区记录,则整个盘面每条磁道上的扇区都被限制为 327 个,总共只有 7 084 128 (1 354 × 16 × 327) 个扇区,存储容量为 3 627 MB。分区记录把这个驱动器的存储容量提高了 44%。

另外各个区域的数据传输速率不同。最外部区域(第 0 区)磁道的传输速率为 28.49 Mbps,这比最内部区域(第 15 区)的 15.10 Mbps 高 89%。

2. 分区

在硬盘上创建分区可以使硬盘支持多个独立的文件系统,每个文件系统都位于自己的分区里,从而每个系统可以使用自己的方式分配文件空间。每个硬盘驱动器必须至少有 1 个分区,最多有 4 个分区,各个分区可以支持相同的或不同类型的文件系统。今天,PC 操作系统通常使用 3 种文件系统。

(1) FAT

FAT 代表 File Allocation Table,即文件分配表。DOS 和 Windows 各个版本支持的标准文件系统。FAT 分区在 DOS 下支持最长 11 个字符的文件名,在 Windows 9x/2000(以及以后的版本)下支持最长为 255 个字符的文件名。标准 FAT 文件系统使用 12(软磁盘)或 16 位数字来标识簇。一个簇包含多个扇区,簇是最小的文件分配单位。FAT 的最大簇容量为 32 KB,所以其最大分区容量是 2 GB。

通过 FDISK 程序,用户只能创建两个物理 FAT 分区——主分区和扩展分区,但可将扩展分区进一步划分为最多 25 个逻辑分区。其他的分区程序,如 Partition Magic,可以创建最多 4 个主分区或 3 个主分区加 1 个扩展分区。

(2) FAT32

FAT32 即 32 位 FAT。微软的操作系统从 Windows 95(第二版)开始支持。FAT32 使用 32 位数据标识簇,这样最大分区容量可达到 2 048 GB。

(3) NTFS

NTFS 代表 Windows NT File System,即 NT 文件系统。这是 Windows NT/2000/XP 操作系统特有的文件系统。NTFS 提供了 FAT 文件系统没有的扩展属性和安全特征。

在这 3 种文件系统中,FAT 文件系统的兼容性最好,可以被几乎所有操作系统访问。FAT32 和 NTFS 提供了更先进的性能,但不能被所有的操作系统所访问。

分区是通过用户操作系统里的 FDISK 程序完成的。FDISK 允许用户指定每个分区的大小,最小为 1 MB 或驱动器容量的 1%,最大为驱动器的整个容量或特定文件系统所允许的最大容量。

FDISK 不能用于改变分区的大小,而只能进行分区的删除和创建。删除分区会删除分区上的数据,创建分区会删除位于该分区的部分数据;如果不想破坏数据,可以使用第三方工具软件,如 Partition Magic 或 Partition Commander。驱动器一旦被分区之后,必须通过使用特定的操作系统执行高级格式化。

3. 高级格式化

在高级格式化过程中,操作系统创建管理磁盘上文件和数据所必需的结构。FAT 分区有一个磁盘卷引导扇区以及文件分配表(FAT)的两个副本,在每个格式化后的逻辑驱动器上还有一个根目录。这些数据结构使操作系统能够管理磁盘上的空间,记录文件踪迹,甚至管理有缺陷的区域,使得它们不引起故障。

高级格式化不是驱动器的实际物理格式化,它是为磁盘创建一张内容目录;而在低级格式化即实际的物理格式化过程中,磁道和扇区被写到磁盘上。如前所述,DOS 的 FORMAT 命令可以执行磁盘的低级和高级格式化操作,但对硬盘只执行高级格式化操作。IDE 和 SCSI 硬盘驱动器的低级格式化由制造商执行,用户不应该也不需要执行低级格式化操作。

光盘系统

从 20 世纪 70 年代发展起来的光学信息存储新技术,早期的开发主要集中在激光数字音响 CD(Compact Disc)盘和激光视频录像 LD(Laser Disc)盘。后来 CD 技术被引入到计算机系统。由于光存储技术有很多优点,随着计算机应用的普及和对信息存储与检索需求的快速增长,推动了光存储技术的迅速发展。

光盘系统属于可更换存储媒体的外存储系统,不像硬盘那样将盘片固定在驱动器内。和硬盘相比,光盘不仅存储潜力大,还具有更多的优点。主要表现在下列几点。

- 存储密度高:光盘的线密度一般高于 1 000 b/mm,道密度是每毫米 600 ~ 700 道,所以光盘的面密度一般可达 10 Mb/cm² ~ 100 Mb/cm²。

- 数据保存时间长:光盘的记录介质是封在两层保护膜中的,并且激光存、取过程是非接触式的,无磨损,抗污染,因此寿命很长,数据保存时间一般在 10 年以上。

- 价格低:由于光盘介质非常便宜,每片盘成本低于 1 元人民币。

目前,光存储系统的光盘主要是两类,即 CD 和 DVD。同时,根据光盘的写特性的不同,两类光盘都有只读型的、一次可写型的以及可擦写型的,如表 7-3 所示。

表 7-3 根据写特性对光盘的分类

光盘写类型	CD	DVD
只读	CD-ROM	DVD-ROM
一次可写	CD-R	DVD-R、DVD+R
可擦写	CD-RW	DVD-RW、DVD+RW、DVD-RAM

7.2.1 光盘分类

标准光盘的直径为 12 cm(5.25 英寸),如图 7-4 所示,还有一种直径为 8 cm 的小光盘。

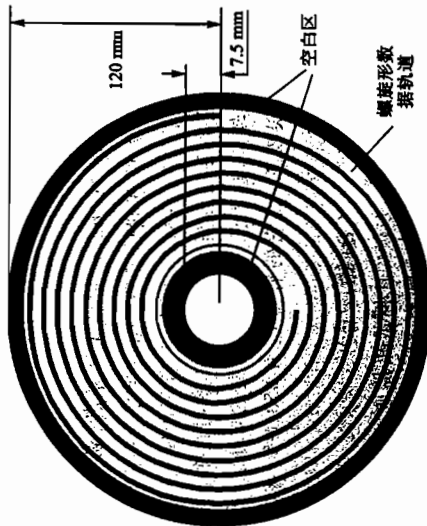


图 7-4 120 mm 标准光盘

光盘的另一种分类是根据具体的应用分类,相当于是光盘记录信息的标准。1982 年飞利浦和索尼公司共同推出了激光唱盘,通称 CD(Compact Disc)。自那时以来,CD 技术已经取得了巨大的发展和成功。它们相继制定了许多标准,如表 7-4 所示,以适应多媒体体的各种应用。记载这些标准的文件分别用不同的颜色包装以示区别,所以称它们为“彩皮书”。

表 7-4 CD 类光盘的应用标准

标准	说明
红皮书 CD - DA (CD - Digital Audio)	数字式高保真音频存储标准,最早的光盘标准。
黄皮书 CD - ROM	用于计算机的只读外存储系统,容量 650 MB。
绿皮书 CD - I (CD - Interactive)	可配合专用播放机播放的交互式光盘。
CD - ROM XA	将 CD - I 的音频和视频引入到 PC。
橙皮书 CD - R/CD - RW	定义可写 CD 的标准。
Photo - CD	基于 CD - ROM XA 和 CD - R 的数字照片存储标准。
白皮书 VCD (Video - CD)	基于 CD - ROM XA 和 CD - I 技术, MPEG - 1 标准压缩的数字视频光盘。
蓝皮书 CD - Extra	在音频光盘嵌入视频、注释及其他信息。
紫皮书 DDCD	双倍密度的 CD 标准。

1. CD - DA (Compact Disc - Digital Audio) 激光唱片(红皮书)

是为了存储数字式高保真音乐而制定的标准。其中规定了 CD 的尺寸、特性、编码、错误校正等,是之后的 CD 发展的基础。

2. CD - ROM (CD - Read Only Memory) 只读 CD 光盘(黄皮书)

CD - ROM 光盘是以 CD 光盘为信息载体的电子出版物。一张 CD - ROM 光盘可以存储 600 MB 以上的数据。它还具有易于分发、保管、方便携带、有利于保密等优点。

由于 CD - DA 的成功,加上它的存储容量大,人们就考虑用在个人计算机上进行其他数据的储存。但 CD - DA 的纠错能力较低,达不到计算机存储器的要求。所以,在 CD - DA 的标准基础上又增加了错误检测和错误校正,而它的容量定为 650 MB。这一标准称为 CD - ROM 标准。

3. CD - I (CD - Interactive) 交互式 CD (绿皮书)

CD - I 是一种配合家用电视机使用的专供 CD 播放机播放的光盘。通过一个遥控器对节目内容进行交互式控制。CD - I 盘可存放 7 000 幅图像或有数字立体声伴音的 72 分钟全屏动画。CD - I 是在 CD - ROM 标准的基础上增加了音频、视频、文字、数据的人机交互格式以及多媒体等技术的规范。通过家用电视机、音响设备以及计算机显示器可以交互播放 CD - I 光盘上的多媒体节目。

4. CD - ROM XA

CD - ROM XA 是 CD - ROM 的扩展,在 CD - ROM 的黄皮书的基础上增加了 3 方面的特性。首先改进了扇区数据格式,目的是将计算机数据信息、音频和视频信息更好地混合存储在同一轨

道内,改进了原来的数据和音频、视频的数据格式不能混合存储的缺点;其次,CD - ROM XA 标准增加一种交错存取技术,将数据、音频、视频交织地存放在轨道上,可以较好地保证在播放时数据、音频和视频的同步;最后,CD - ROM XA 采用 ADPCM 音频压缩编码方式,和原来黄皮书规定的 PCM 编码方式相比较,可以大幅度地减少音频信息占用的空间。

5. Photo CD

Photo CD 是 CD - ROM 技术在摄影领域的应用。一张光盘可存放 100 张彩色照片,还可以配上音乐、语言解说和文字说明。

Photo CD 兼容的播放机是一套家庭 A/V 娱乐系统的组成部分,用来播放 Photo CD 和音频光盘。

光盘上图像的压缩和存储采用了 Kodak 公司独有的 Photo YCC 编码格式,每一幅图像包括高达 6 种不同的分辨率。Kodak 公司定义了几种不同类型的 Photo CD 光盘,以满足不同类型的用户需要。Photo CD Master 光盘是用户使用的标准格式,能存储 100 张为 $2\,048 \times 3\,072$ 像素分辨率的照片。PRO Photo CD Master 光盘是为专业摄影师设计的,使用大的胶片格式,例如 70mm、120mm 或 4×5 英寸。除 Photo CD Master 光盘上已有的分辨率外,这类光盘增加了一个更高的分辨率 ($4\,096 \times 6\,144$ 像素)。由于增加了这个高分辨率的图像,这类光盘可以容纳大约 25 ~ 100 幅图像,具体数量取决于胶片的格式。

6. VCD (Video CD) 电影光盘(白皮书)

通常简称 VCD 电影光盘。这是一种数字电视光盘,存有数字化的电视图像和伴音,是激光视盘 (LD) 的替代产品。一张 VCD 光盘可存储 74 分钟的电视节目,图像质量和声音质量分别达到家用录像机 VHS 和激光唱片 CD - DA 的水平。

VCD 是在基于 CD - I 和 CD - ROM XA 的基础上开发的标准, VCD 盘上的视频和音频信号分别采用国际标准 MPEG - 1 和 ADPCM 进行压缩编码,播放时要进行实时解压处理。

7. CD - R (CD - Recordable) 和 CD - RW (CD - Rewritable) (橙皮书)

橙皮书于 1989 年发布,包括三部分,第一部分是关于 CD - MO (CD - Magneto Optical, 磁光盘) 的标准,但由于技术过时,在产品问世之前被撤销了;第二部分是关于 CD - R 的标准;第三部分描述了 CD - RW 的标准。

CD - R (曾经被称为 CD - WO, CD - Write Once) 是可写入式 CD 光盘,但写入的信息是永久性的,不可改写。数据可以分多次写入 (称为 Multisession Recording)。CD - R 盘片的几何尺寸、数据记录的格式和 CD - ROM 一样,因此可以在 CD - ROM 驱动器上读出。

CD - RW 为可擦写的 CD 光盘,橙皮书标准的第三部分给出了定义,很快成了 CD - R 的替代产品。CD - RW 使用了一种不同类型的介质,在这种介质上光驱至少能重写 1 000 次。尽管比 CD - R 介质略贵一些,但 CD - RW 仍然远比其他可更换介质便宜。这使得 CD - RW 这种适用技术广泛应用于系统备份、文件归档以及其他的数据存储任务。

CD - RW 的缺点是光盘的低反射率限制了盘上数据的可读性。许多老式的 CD - ROM 和 CD - R 光驱并不能读 CD - RW 光盘,但在所有 24X 及以上的 CD - ROM 驱动器有一种多读 (MultiRead) 能力,也就是说它们能读 CD - RW 光盘。另外,CD - RW 光驱也具有刻写 CD - R 光盘和读标准 CD - ROM 光盘的能力。

7.2.2 光盘的数据存储方式

1. 光盘的扇区格式

最初的黄皮书中定义了两种 CD 扇区数据格式:模式 1 和模式 2。模式 1 中包含纠错码 (ECC) 和检错码 (EDC), 而模式 2 却没有, 且两种模式不能混合记录到一条轨道中。经过 CD-ROM XA 的扩展, 扩展了新的数据格式, 经过 CD-ROM XA 扩展后的数据格式如图 7-5 所示。

模式 1	同步	头	子头	用户数据	EDC	空	ECC
	12	4		2 048	4	8	276
模式 2 格式 1	同步	头	子头	用户数据	EDC		ECC
	12	4	8	2 048	4		276
模式 2 格式 2	同步	头	子头	用户数据			EDC
	12	4	8	2 324			4

图 7-5 CD 光盘扇区格式

经过 CD-ROM XA 扩展的模式 2 增加了一个子头数据域, 用来表示数据的类型, 可以是音频、视频或其他数据。同时, 模式 2 有两种格式, 它们的区别是, 模式 2 的格式 1 有纠错码 (ECC), 可以纠正正在光盘读/写过程中的错误, 计算机的数据文件应该采用这一格式。模式 2 格式 2 没有纠错码, 将格式 1 的纠错码的 276 个字节分配给了用户数据, 增加了轨道的利用率, 这一格式适合存储对错误不敏感的音频和视频数据。

2. CD-ROM 文件系统

CD-ROM 的黄皮书规范只规定了数据扇区如何在光盘上存储, 并没有定义文件系统, 也没有提及这些数据如何在文件中存放以及不同的 PC 机操作系统如何格式化这些数据。很显然, 这种情况阻碍了计算机行业中 CD-ROM 应用程序的兼容和发展。

CD-ROM 文件系统标准的制定是从 1985 年开始的, 第一个标准 High Sierra, 是几家公司合作在 High Sierra 宾馆制定的文件格式规范, 由此而得名。High Sierra 光盘文系统的制定打开了一条大规模生产和发行 CD-ROM 软件的道路。采纳这个标准也使得光盘发行者为自己的软件提供跨平台的支持, 从而很容易地制造 DOS、UNIX 及其他操作系统格式的光盘。

High Sierra 格式被国际标准化组织 (International Organization for Standardization, ISO) 采纳, 1988 年, 经过几次改进, 它以 ISO 9660 标准重新发布。ISO 9660 与 High Sierra 不完全相同, 但支持 High Sierra 格式的驱动程序很快得到更新, 既能支持 ISO 9660, 也能支持以前的 High Sierra 格式。

微软于 1988 年编写了 MSCDEX.EXE (Microsoft CD-ROM Extension) 驱动程序, 并将其发布

给 CD-ROM 硬件和软件生产商, 让他们附到其产品一起出售。微软在 Windows 95 及其后续版本中直接嵌入了对 ISO 9660 和 Joliet 文件系统的支持, 而不必再另外使用驱动程序了。

Joliet (一个城镇的名字) 是由微软为 Windows 95 及其后续版本开发的 ISO 9660 扩展标准, 它允许记录 CD 时的文件名长度达 64 个字符, 可以含有空格以及国际统一字符集 (Unicode International Character Set) 中的任何字符; Joliet 对于那些不能使用长文件名的程序也保留了 8.3 结构的别名。同时 Joliet 支持多会话记录。

另一种文件系统是 UDF (Universal Disk Format), 是一个相对较新的文件系统, 是用于像 CD-ROM 和 DVD 等光介质的工业标准格式。与标准 CD-ROM 使用的 ISO 9660 相比, UDF 具有一些优点, 但最突出的是它采用了数据包写 (Packet Writing) 技术, 该技术是将数据一小部分一小部分地写到光盘中, 将其视为标准的磁驱动器。UDF 文件系统允许使用长达 255 个字符的文件名。标准 CD-ROM 光驱、驱动程序或者像 DOS 一样的操作系统不能识别 UDF 格式的光盘。可记录光驱可以识别该格式, 普通的 CD-ROM 光驱必须符合 MultiRead 规范才能读取 UDF 格式光盘。

以上介绍的 4 种文件系统对于操作系统的支持见表 7-5。

表 7-5 CD 文件系统格式

CD 文件系统	DOS/Windows 3.x	Windows 9x/Me	Windows NT/2000/XP	Mac OS
High Sierra	是	是	是	是
ISO 9660	是	是	是	是
Joliet	是 ^①	是	是	是 ^①
UDF	否	是 ^②	是 ^②	是 ^②

① 显示短文件名 (8.3 格式)。

② 仅当安装了 UDF 驱动程序时才支持。

7.2.3 DVD

DVD 代表 Digital Versatile Disc, 即通用数字光盘, 其驱动器和 CD 兼容, 任何 DVD-ROM 光驱同时也是 CD-ROM 光驱。DVD 除了密度较高以外, 其他技术与 CD 完全相同。DVD 标准极大地提高了 CD-ROM 尺寸光盘的存储容量, 因而为一些应用提供了更大的空间。

DVD 光盘可以双面存储信息, 而且每一面可以存储双层。DVD 的盘面盘就可以存储 4.7 GB (单层) 到 8.5 GB (双层) 容量的数据, 是 CD 容量的 11.5 倍。目前在市场上的 DVD 电影光盘大多是单面单层 (通常称为 DVD5) 和单面双层 (通常称为 DVD9) 标准。双面 DVD 容量是上述值的两倍, 双面单层和双面双层 DVD 的容量分别为 9.4 GB 和 17 GB。不过在目前的光驱上要读取另一面, 还得手工将其翻转过来。

DVD 光盘最多可以存储两层信息。单面单层盘的初始容量为 4.7 GB, 这种盘与目前的 CD-ROM 直径、厚度完全相同。利用 MPEG-2 格式进行压缩后, 它足以存储 133 分钟的视频信息, 这对于全长度、全屏幕、全动作的电影来说也是足够的, 可以包括 3 个 CD 质量的音频通道和 4 个字幕通道。单面双层 DVD 可以存储 240 分钟以上的视频数据。各种 DVD 光盘的容量如表 7-6 所示。

202 第7章 外部存储器

表 7-6 DVD 光盘的不同特性的类型

格式	类型	容量 (GB)	兼容性
DVD - ROM	只读	单面最大 8.5	最基本的只读光盘
DVD - R	可刻录	单面最大 4.7	兼容 DVD - ROM
DVD - RAM	可重写	单面最大 4.7	与 DVD - ROM 不兼容
DVD - RW	可重写	单面 4.7	兼容 DVD - ROM
DVD + RW	可重写	单面 4.7	兼容 DVD - ROM

在 DVD 规格得到统一之后,1997 年 DVD 论坛由会员投票通过了第一个 DVD 刻录规范——DVD - RAM。另一方面索尼与飞利浦坚持独立开发自己的 DVD 刻录规范,并联合惠普共同研制了另一个 DVD 刻录规范,这就是 DVD + RW 与 DVD + R 规范。与此同时,先锋公司也在 1997 年拿出了它的刻录式 DVD 技术——DVD - R 以及之后的 DVD - RW,由于面向主流 DVD 视频市场并且是对 DVD - RAM 弱点的重要补充,所以迅速通过了 DVD 论坛的认证,成为 DVD 官方指定的非数据存储应用的刻录式 DVD 标准。这样,DVD 刻录技术就出现了三大类、五种规范(DVD - RAM、DVD - R/RW、DVD + R/RW)。为了能让这些刻录格式得到统一,两大阵营仍在继续努力,这其中就包括 DVD 论坛推出的 DVD - Multi 规范和一些 DVD + RW 联盟会员公司私下开发的 DVD \pm RW 规范(又称 DVD - Dual)。但它们并不是新的刻录技术,而是将现有的五种规范进行相应的结合。其中 DVD - Multi 支持 DVD 论坛两大刻录技术、三种刻录规范,而 DVD \pm RW 则是 DVD + RW 与 DVD - RW 相结合的产物。

目前看来,DVD + RW 是业界最有前途的标准,因为它最便宜、最容易使用,并且与现有的各种格式的兼容性也很好。

1. DVD - RAM

DVD - RAM 是一种由松下、日立以及东芝公司联合推出的可写 DVD 规范,如表 7-7 所示,它使用类似于 CD - RW 中的相位变化技术。但是,由于在介质反射率和数据格式上的差异,多数标准的 DVD - ROM 光盘都不能读 DVD - RAM 盘。可以读取 DVD - RAM 盘的光驱必须符合 MultiRead 2 规范。符合 MultiRead 2 标准的 DVD - ROM 和 DVD 视频播放器都可以读取 DVD - RAM 盘。

表 7-7 DVD - RAM 规范

存储容量	单面盘 2.6 GB,双面盘 5.2 GB
光盘直径	80/120 mm
光盘厚度	12 mm(0.6 mm \times 2,黏合结构)
刻录方法	相位变化
激光波长	650 nm
数据位长度	0.41 ~ 0.43 μ m
刻录磁道间隔	0.74 μ m
轨道格式	摆摆式平地 and 凹槽

最早的 DVD - RAM 盘容量为 2.6 GB(单面)或 5.2 GB(双面)。第二代的 DVD - RAM 容量为 4.7 GB(单面)或 9.4 GB(双面)。

一般来说,DVD - RAM 驱动器或播放器可以读取 DVD - Video、DVD - ROM 和 CD 介质盘,目前使用的大部分 DVD - ROM 和 DVD 视频播放器不能读取 DVD - RAM 介质盘。

2. DVD - R

DVD - R 是一种类似于 CD - R 的一次写介质,因此它对于记录存档数据或发行光盘是理想的介质,DVD - R 盘可以在标准的 DVD - ROM 驱动器上读取。

DVD - R 的单面容量为 3.95 GB,大约是 CD - R 容量的 6 倍;双面盘的容量加倍。DVD - R 规范如表 7-8 所示。

表 7-8 DVD - R 规范

存储容量	单面盘 3.95 GB,双面盘 7.9 GB
光盘直径	80/120 mm
光盘厚度	12 mm(0.6 mm \times 2,黏合结构)
刻录方法	有机染料层刻录方法
激光波长	635 nm(刻录);635/650 nm(播放)
数据位长度	0.293 μ m
刻录磁道间隔	0.80 μ m
轨道格式	摆摆式凹槽

3. DVD - RW

DVD 论坛于 1998 年推出了 DVD - RW(Rewritable)标准,该标准主要由先锋公司制定,它使用了相位变化技术,比 DVD - RAM 更兼容于标准的 DVD - ROM 驱动器。基于这种技术的驱动器在 1999 年末开始推向市场。尽管新类型的 DVD 驱动器已经与 CD - R/RW 标准更加兼容,但协调各种可写的 DVD 标准仍然是该标准目前存在的一个主要问题。

4. DVD + RW

DVD + RW,也称为 DVD 相位变化可重写规范(DVD Phase Change Rewritable)。该规范注定要成为最重要的 DVD 刻录标准,因为它是最便宜的、最容易使用的、与现有格式兼容得最好的标准。DVD + RW 规范(如表 7-9 所示)由飞利浦、索尼、惠普等公司联合开发和支持,这些公司都是一个称为 DVD + RW 联盟的工业组织的成员。

表 7-9 DVD + RW 规范

容量	单面盘 4.7 GB,双面盘 9.4 GB
最长视频刻录时间	单面盘 4 小时,双面盘 8 小时
光盘直径	80/120 mm
光盘厚度	12 mm(0.6 mm \times 2,黏合结构)

续表

激光波长	650 nm(与 DVD 视频相同)
数据位长度	0.293 μm
刻录磁道间隔	0.80 μm
轨道格式	摆动的螺旋凹槽

DVD + RW 是唯一一个与现有的 DVD 视频播放器和 DVD - ROM 驱动器在 PC 机和娱乐应用的实时视频刻录和随机数据刻录方面完全兼容的可写格式。DVD + RW 不仅仅可以用于 PC 的数据存储,也可以直接以 DVD 视频的格式刻录视频信息;这是可写 DVD 工业期待已久的一次重大突破,因此,DVD + RW 肯定会取代消费级的家用录像机。

随着娱乐及 PC 应用的继续发展,DVD + RW 可能成为将 DVD 视频娱乐设备和 PC 的 DVD - ROM 驱动器无缝集成到一起的唯一的可写式 DVD 标准。DVD + RW 技术与 CD - RW 非常类似,DVD + RW 光驱也可以读取 DVD - ROM 光盘以及所有的 CD 盘,包括 CD - R 和 CD - RW。

根据 DVD + RW 标准,刻录过程可以随时挂起和继续执行而不会浪费空间,并能将多个刻录会话链接起来,这在随机刻录和视频应用中非常有用。这种“无损失的链接”也可以使用一个新数据块代替任何一个 32 KB 的数据块(最小的刻录单元),定位精度的间隔为 1 μm 。在将数据刻录到轨道上时,为了达到这样高的定位精度,预先印制的凹槽应该以较高的频率摆摆。从凹槽中读取的定时和地址信息可以非常准确。

DVD + RW 的具备快速格式化特性,就是将一张 DVD + RW 空白盘放入到驱动器以后,几乎立刻就可以直接在上面刻写数据,实际的格式化是在真正有数据写入之前以后台方式进行的。

在已经发布的这三种可写式 DVD 规范中,DVD + RW 是最有希望将可写式 DVD 广泛应用于家庭娱乐产品和集成到新式的 PC 机中的标准。

7.2.4 CD 和 DVD 技术

CD 和 DVD 采用类似的技术,两种光盘采用同样的尺寸(直径为 120 mm,厚度为 1.2 mm,中心孔洞直径为 15 mm),盘上都有以聚碳酸酯基印制的凹陷和平地。CD 光盘只有单面单层盘,而 DVD 与 CD 不同的是,DVD 除了可有双面结构外,每面还可以有两层用来刻录数据,每一层单独压制,然后结合到一起最终形成 1.2 mm 厚的光盘。制作过程也很相似,只是 DVD 每一面的每一层是由一块单独的聚碳酸酯塑料压制而成的,然后合并到一起形成完整的光盘。CD 和 DVD 的主要不同在于 DVD 的密度更高,激光波长更短,从而可以存储更多数据。

CD 和 DVD 一样,每一层都是以一个单一的螺旋形物理轨道的形式印制,轨道从光盘的最里端开始向外环绕;光盘以逆时针方向旋转(从下端看),每个螺旋轨道都包含凹陷和平地。DVD 的每一层都覆盖一层反射激光的金属膜;外层的金属膜较薄,以便激光穿过它读取里层的数据。如果是单面盘,上面部分会有一个标签;如果是双面盘,靠近中心处有一个小环用做标签。

光盘的轨道伸展开来是一条很长的、连续的轨道,为了便于数据的管理,将轨道按照等长度切割成小段,每一个小段为一个扇区,每一个扇区具有固定的存储容量。CD 和 DVD 的扇区的有

效用户数据容量都是 2 048 字节。

如果比较 CD 和 DVD 就可以发现,DVD 标准之所以可以提供比 CD 大得多的容量,就是在数据密度上的增加,轨道密度、数据的线密度都增加了很多,如图 7-6 所示。

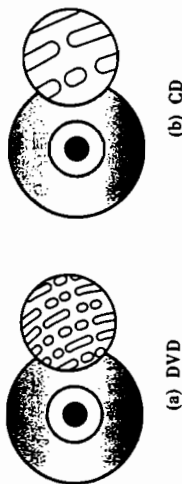


图 7-6 CD 和 DVD 的轨道和线密度

表 7-10 列出了 CD、单层 DVD 和双层 DVD 的轨道参数。可以看到,从 CD 发展到 DVD 伴随着多个和密度相关的参数的提高。另外,表中的 CD 的参数是 650 MB 容量的光盘参数,除此之外,还有一种容量为 700 MB 的 CD 光盘,后者是在前者的基础上略微减小轨道间距,使得容量增加了约 50 MB。

表 7-10 CD 和 DVD 的轨道数据

参数	CD	单层 DVD	双层 DVD
凹陷深度 (nm)	0.125	0.105	0.105
凹陷宽度 (nm)	0.6	0.4	0.4
凹陷长度 (nm)	0.9 ~ 3.3	0.4 ~ 1.87	0.44 ~ 2.05
轨道间距 (nm)	1.6	0.74	0.74
轨道密度 (圈/mm)	625	1 351	1 351
轨道总圈数	22 188	49 324	49 324
轨道总长度 (m)	5 772	11 836	11 836
扇区总数	333 000	(2 292 897)	(2 083 909)
扇区长度 (mm)	17.33	5.16	5.68
扇区容量 (B)	3 234	(4 836)	4 836
扇区用户数据 (B)	2 048	2 048	2 048
有效用户数据容量 (B)	681 984 000	4 695 853 056	8 535 691 264

7.2.5 下一代光存储技术

目前出现的下一代光存储标准主要是“蓝光”和“HD - DVD”(高清晰 DVD),这两种标准都是在 2005 年发布的。两者都采用了波长 405 nm 的蓝紫色激光作为读写光盘的光源,取代目前通用的波长 650 nm 的红色激光,这样可以大幅度提高光盘数据存储容量。一般的 DVD 光盘的存

储容量为4.7 GB(单面单层),而“蓝光”DVD的存储容量高达25 GB,“HD-DVD”光盘的容量也达到15 GB,足以存储高清晰度数字影视节目和海量数据。

但是这两种标准却无法互相兼容。两种标准都有不同的企业联盟的支持,形成了两个阵营,“蓝光”阵营的厂商包括索尼、松下、三星和飞利浦等,“HD-DVD”阵营厂商包括东芝、NEC、三洋等。根据两种标准,有厂商推出不同的DVD光盘、播放机、刻录机等。

7.2.6 光盘驱动器

光盘驱动器是读/写光盘的设备,在目前的计算机上几乎属于必不可少的外部设备。光盘驱动器的类型也从最初的CD-ROM驱动器,发展到CD-RW、DVD-ROM驱动器,目前功能比较强的是DVD-RW驱动器。

由于新型的光盘类型的不断出现,使得光盘的种类数量不断增加,新型的光盘驱动器一般都要兼容老式的光盘类型。另外,光盘驱动器的写盘能力不同,也就使得光盘驱动器的类型很多。

1. 光盘驱动器的构成

从光盘上读信息是一个将低能的激光束从光盘上各层的反射层反射回来的过程。发光二极管发出一束激光,经过反光镜和聚焦透镜照射到光盘上后再反射回来,反射回来的激光束经过反光镜、光束分离器和聚焦透镜,最后到达光电检测器。如果激光遇到的是轨道上的平地(平坦部分),它就会被反射回来;如果激光遇到的是凹陷,就没有激光返回。

随着光盘旋转,激光不间断地发出,照射到凹陷和平地上,而接收器则感应是否有激光返回。每当激光通过凹陷的边缘时,光敏电阻就会感受到有反射光(或者反过来)的状态变化。这样,穿过凹陷边缘引起的每一次反射状态的变化就会被解释为比特1,没有反射状态的变化被解释为比特0,然后再将比特流转换为实际的数据。光盘驱动器的内部结构如图7-7所示。

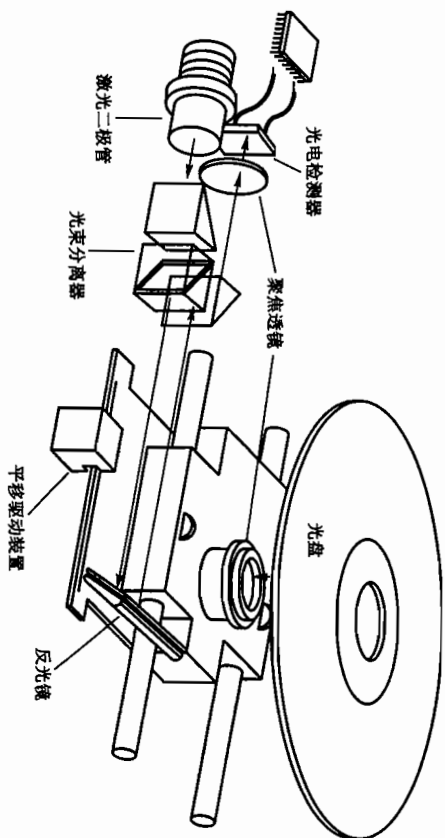


图7-7 光盘驱动器的内部结构图

2. 光盘驱动器的分类

最早的光驱当然是CD-ROM驱动器。当出现CD-R和CD-RW光盘后,随之出现了CD-R/RW驱动器。DVD类驱动器最早的也是DVD-ROM驱动器,在这一时期还有一种被称为“康宝”(Combo)的驱动器。Combo驱动器只能读DVD盘,但可以写CD-R/RW盘。Combo驱动器的出现主要是由于价格的原因,当时DVD刻录机很昂贵而且没有普及,但DVD-ROM光盘已经非常普遍,Combo驱动器满足了大部分人的应用需求而且价格不高。随着DVD刻录设备价格的逐步下降,DVD刻录机最终将取代Combo驱动器。

由于目前DVD刻录设备的标准并不统一,比较大的疑惑是DVD-R/RW和DVD+R/RW两者的比较和选择。DVD+R/RW的特点是,借助于出色的物理设计,再配合强大的逻辑功能,DVD+R/RW的功能已经超越DVD-R/RW。也就是说,DVD-R/RW能做到的,DVD+R/RW都可以,反之却不然;而且在速度的进步上一直领先于DVD-R/RW。客观地说,两者的性能差别不大,对于普通的PC用户来说,选择DVD-R/RW或DVD+R/RW都是没有问题的;DVD+R/RW的优势在消费电子领域里才会更加明显(DVD+R/RW随机寻址与无链接能力提供了更好的可编辑性),对于日常的数据存储与影片的保存,二者都可以胜任。

表7-11 驱动器与光盘的兼容性

驱动器类型	光盘类型					
	CD-ROM	CD-R/RW	DVD-Video	DVD-ROM	DVD-RAM	DVD-R/RW
CD-ROM	R	R				
CD-R/RW	R	R/W				
DVD-ROM	R	R	R	R		R
Combo	R	R/W	R	R		R
DVD-RAM	R	R	R	R	R/W	R
DVD-RW	R	R/W	R	R		R/W
DVD+RW	R	R/W	R	R	R	R/W

注:表中R表示只读,R/W表示可读、可写。

3. 光盘驱动器的性能指标

(1) 数据传输速率

数据传输速率表明了给定的时间里驱动器可以从光盘读取并传输到主机的数据量。通常,传输速率表示驱动器读取大量顺序数据流的能力。

传输速率有两种度量方式。CD/DVD光驱标注的最常见的形式是“×”倍速,定义为一个特殊的基准速率的倍数。根据最初的标准,CD-ROM驱动器的基本传输速率为153.6 KBps,传输速率为该值2倍的驱动器就标注为2×,传输速率为该值40倍的标注为40×,依此类推。DVD驱动器的基准速率为1385 KBps,速率为该值20倍的驱动器标注为20×。注意,×倍速通常指的是读取光盘最外边数据(最末端)时所达到的最大值。光盘开始部分的传输速率可能只有该

值的一半,当然,平均传输速率会在二者之间。

对于可刻录 CD 驱动器,传输速率以不同的方式来标注。CD - R 驱动器一般列出两个速率(一个写速率,一个读速率);CD - RW 驱动器则列出 3 个速率,传输速率一般以 A/B/C 的形式标注,其中 A 表示写 CD - R 盘的速率,B 表示写 CD - RW 盘的速率,C 表示读盘的速率。市场上首先问世的第一个 CD - RW 驱动器标注为 2/2/6,目前已经有 20/10/40 的版本出现了。

(2) 访问时间

访问时间是驱动器接收到读命令到实际开始读一位数据之间的时间间隔。这个时间以毫秒(ms)为单位,其典型值是 95 ms。访问时间通常是一个平均访问时间,实际的访问时间则完全依赖于数据在光盘上的位置。许多制造商给出的访问时间是计算对光盘一组随机读时间的平均值。

很显然,更快的平均访问速率是大家所希望的,特别是当依赖于驱动器快速定位和取数据的时候。CD/DVD 驱动器的访问时间在不断地改善,但些些平均访问时间仍比硬盘慢得多,从 200 ms 到低于 100 ms 不等,而典型的硬盘驱动器的访问时间只有 8 ms。两者速度的差别主要在于驱动器本身的构造不同:硬盘在更小的介质的表面区域有多个读磁头,而 CD/DVD 则只有一个读激光束,它必须要访问整个光盘。另外,CD 上的数据都组织在一条螺旋形的长轨道上,当驱动器定位它的激光束来读轨道上的数据时,必须首先估计到光盘的距离并向前或向后移动到螺旋轨道上合适的位置。从外部边缘上读数据要比从内部读取数据需要更长的访问时间,除非是恒定角速度驱动器,因为这种驱动器可以以固定角速度旋转,因此对外部磁道的访问时间等于对内部磁道的访问时间。

DVD 驱动器通常标注两个访问时间,一个是读 DVD 盘的访问时间,另一个是读 CD 盘的访问时间。读 DVD 盘的访问时间通常比读 CD 盘的访问时间长 10 ms ~ 20 ms。

(3) 缓存容量

大多数驱动器都带有内部缓冲器。这些缓冲存储器是实际的存储芯片,安装在驱动器的电路板上,使它在发送数据给 PC 之前可以准备或存储更大的数据段。CD/DVD 典型的缓冲器大小为 128 KB。可刻录 CD 或 DVD 驱动器一般具有 2 MB ~ 4 MB 以上的大容量缓冲器,可以平滑写操作。一般来说,驱动器越快,就有容量更大的缓冲器,以处理更高的传输速率。

CD/DVD 驱动器带有缓冲器具有很多好处。缓冲器可以保证 PC 以固定速率接收数据。当一个应用程序从驱动器请求数据时,数据可能分散在光盘上不同区域的文件之中。因为驱动器的访问速度相对较慢,在读取不同区域的数据时会使得驱动器间歇性地向 PC 发送数据。在访问较小的文件时,感觉上不明显,当通过速率较低又没有数据缓冲器的驱动器访问大文件时,这种现象就很明显了。建议的最小缓冲器容量为 128 KB,为了获取更好的性能,可以使用具有 256 KB 以上缓存的驱动器。

4. 光盘驱动器的转速控制

在主机电机的转速控制方面,CD/DVD 驱动器要比硬盘复杂得多。由于光盘的扇区是等长度的,光盘的转速和数据传输速率存在明确的关系。目前的主轴电机控制主要有以下几种方式。

(1) 恒定角速度 (Constant Angular Velocity, CAV)

从图 7-8 可以看出,CAV 的特点是转速恒定,传输率持续攀升,这种方式也是目前硬盘所采用的。当读写光盘的外围时,由于单位转动时间内光头的记录区域越来越大,所以数据传输率也就越来越高。CAV 的优点就是主轴电机无需变速,容易控制。但在刻录时,由于单位时间内划分的记录区域长度越来越大,而要在相同的时间记录更多的数据,必须实时地控制激光功率的提高以满足更高倍速刻录的要求,因此对激光功率控制精度的要求比较高。

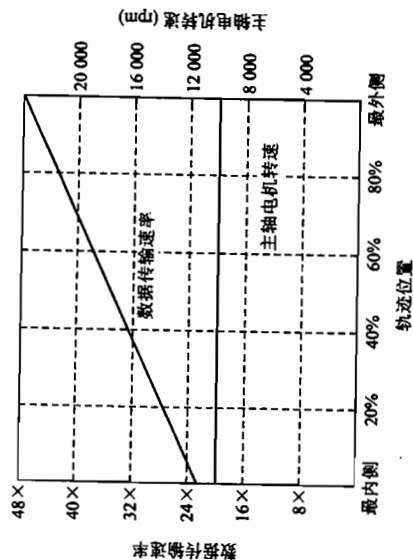


图 7-8 CAV 控制方式曲线

(2) 恒定线速度 (Constant Linear Velocity, CLV)

CLV 方式的特点是数据传输率恒定(单位时间内光头相对于光盘移动的距离恒定),但主轴电机的转速逐渐降低。由于在单位时间内光头走过的距离相同,所以数据传输率固定,但电机控制电路要求有较高的控制精度。

对于刻录来说,CLV 可以说是最为理想的刻录方式,因为它的刻录效率最高,起始与终止速率是一样的,不像 CAV 那样有明显的落差。然而,CLV 的线速度恒定也意味着在周长较短的光盘内圈要达到指定的线速度,电机的转速就要很高。若以 CLV 方式进行 16x DVD 刻录,那么电机的转速就要高到 21 000 rpm;若以 CLV 方式进行 48x CD 刻录,转速则将近 23 000 rpm。要知道盘径小得多的硬盘,目前最高的转速也只有 15 000 rpm,所以,CLV 只能用于低速的刻录。

(3) 局部恒定角速度 (Partial - Constant Angular Velocity, P - CAV)

P - CAV 刻录可以看成是 CAV 与 CLV 的一种组合。一开始采用 CAV 方式,转速不变,数据传输率持续提高,到指定的切换点时,则变为 CLV 方式,转速下降,数据传输率保持恒定。从某种角度上说,P - CAV 是为 CLV 服务的,克服了 CAV 方式在读写内圈轨道时主轴电机速度过快的问题,而到后期随着读写半径的加大,CLV 就有发挥的空间了,因此最后切换回 CLV 以保证最佳的效率。显然,如果条件允许,越早切换至 CLV 方式,刻录速度就会越快。

(4) 区域恒定线速度(Zone - Constant Linear Velocity, Z - CLV) 顾名思义,区域恒定线速度就是将读写区域划分成不同的小区域,在每个区域内采用不同速率的 CLV 方式,而在区域结合点进行速率与转速的变化。

Z - CLV 技术的出现就是为了避开光盘内圈对 CLV 的限制。先以低速起步,分阶段地进行提速。能提速的原理已经在 P - CAV 的介绍中说明了。显然,最高速率的平台越长越好,最低速率的平台越短越好。从理论上讲,台阶越为密集,Z - CLV 的效率就会更高,因为它更有效地利用了光盘周长逐渐增加的变化,但是切换速率时要对数据处理和电机控制有较大的变动,因此从实际应用效果上说,台阶越多可能控制起来更复杂。

7.3 移动存储器

从结构上考虑,移动存储器可以分为两大类。第一类是存储介质可更换的,如磁盘、光盘以及大容量的可更换盘片的磁盘系统等,也包括各种存储卡。第二类是外接的存储模块,通过标准的接口和计算机连接,如 U 盘、移动硬盘等。移动存储系统之所以种类繁多,主要是因为出现在不同的时代,有一些品种已经逐步被新的品种替代。目前使用的移动存储器主要是光盘、U 盘、移动硬盘和存储卡。

7.3.1 移动盘

移动盘主要包括两类,第一类是由半导体存储器为存储核心,这一类主要是 U 盘。另外一类是以硬磁盘为存储核心,通常称这一类为移动硬盘。两类移动存储器都使用计算机的通用接口,目前使用的接口主要是 USB 接口,但移动硬盘也有使用其他的接口的,如并行接口和 IEEE - 1394 接口。

1. U 盘

U 盘(USB 接口移动盘),通常也被称为闪存盘。U 盘是一个通用 USB 接口的无需机械驱动器的微型高容量移动存储产品,它采用的存储介质为闪存存储器(Flash Memory)。U 盘不需要额外的驱动器,将驱动器及存储介质合二为一,只要接上计算机上的 USB 接口就可独立地读写数据。U 盘可用于存储任何格式数据文件和在计算机间方便地交换数据。

U 盘作为新一代的移动存储器,彻底替代了软盘。U 盘的体积很小、重量极轻、约为 15 g,特别适合随身携带,如图 7 - 9 所示。盘中无任何机械式装置,抗震性能极强。另外,U 盘属于固态存储器,具有防潮防磁,耐高低温等特性,安全可靠性能很好。

从容量上看,U 盘的容量在不断增大,从早期的 16 MB 到目前的数 GB。从读写速率上看,U 盘采用 USB 接口,读写速率较软盘大大提高。



图 7 - 9 USB 接口移动盘

U 盘的存储核心是闪存存储器(简称闪存)。根据逻辑结构的不同,闪存主要分为 NAND 型和 NOR 型两种。前者可提供更大的容量,但不支持代码本地执行,读取速度也较慢(但写入速度较快);而 NOR 型闪存支持代码本地运行,读取速度也稍快(但写入速度稍慢),但主要缺点在于很难实现较高的存储密度。不同的特性让这两者分别属于不同的应用:NAND 广泛用于数据存储相关的领域,如移动存储产品、各种类型的存储卡、MP3 音乐播放器等,而 NOR 型闪存主要用于手机、掌上电脑等需要直接运行代码的场合。

U 盘的指标主要是容量和接口传输速率,容量是随着集成度的发展逐步提高的,从较早的 16 MB,发展到现在的几 GB。接口传输速率取决于 USB 接口的版本,有 1.1 和 2.0 两种版本,前者的传输速率是 12 Mbps,后者的传输速率是 480 Mbps。由于闪存的写入速率通常为 350 Kbps,而读取方面稍微快一点,可以达到 700 Kbps,所以 U 盘的读写速率远远达不到 USB 2.0 接口的传输速率,而是更接近 USB 1.1 的传输速率。希望 USB 2.0 接口的 U 盘能达到 480 Mbps 的持续传输速率是不现实的。

2. 移动硬盘

移动硬盘实际上就是采用了 USB 接口或 IEEE - 1394 接口的硬盘,如图 7 - 10 所示。较早的移动硬盘都是由 2.5 英寸的笔记本电脑硬盘构成的。随着小型和微型硬盘的发展,新型的移动硬盘也逐步采用 1.8 英寸、1 英寸的硬盘,使得移动硬盘的体积大大减小。

移动硬盘的多数技术指标取决于硬盘的指标,如容量、寻道时间和缓存容量等。它的传输速率与硬盘的介质传输速率和接口速率都有关系,但由于硬盘的介质传输速率远远高于 USB 1.1 接口的传输速率,如果是采用 USB 1.1 接口的移动硬盘,接口的传输速率是移动硬盘的传输瓶颈。USB 2.0 和 IEEE - 1394 接口的传输速率和硬盘的介质传输速率比较匹配,如果要求传输速率比较高,应该选用 USB 2.0 或 IEEE - 1394 接口的产品。



图 7 - 10 移动硬盘

7.3.2 存储卡

存储卡是新兴的存储技术,使用它作为笔记本电脑的主存或辅存已经有数年时间了。随着数码相机和 MP3 播放器的兴起,这种曾经不起眼的小产品已经变成了必不可少的配件了。

存储卡内部的记忆体是一种非易失性存储器,与按字节存储数据的内存不一样,它是以分块方式存储数据的。在向其中一块写入新数据之前,必须将该块中的数据全部擦除。

目前,闪存的存取速率、单纯的存储性要求和很小的体积,使其成为一些手提设备(如笔记本电脑和数码相机)的最好搭档。由于存储卡首先是在数码相机里使用的,它也被称为“数字胶卷”,但与普通胶卷不同的是,它可以重复使用。

现在常用的存储卡有好几种,它们几乎都是在数码相机里使用过的。主要有以下几种:

- CF(CompactFlash)卡

- SM(SmartMedia)卡
- MMC(MultiMediaCards)卡
- SD(Secure Digital Card)卡
- 记忆棒(Memory Stick)

1. CF(CompactFlash)卡

CF卡(如图7-11所示)是由美国SanDisk公司于1994年研制成功的,它是早些时候应用最为广泛的存储卡,可永久保存数据,并且无需电源,速度快,价格低于其他类型的存储卡。由于把存储模块与控制器集成在一起,这样CF卡的外部设备就可以做得比较简单而没有兼容性问题,特别是升级换代时也可以保证与旧设备的兼容性,保护了用户的投资,而且几乎所有的操作系统都支持这种卡。在笔记本电脑数码相机等数码产品中,CF卡得到了非常广泛的应用。



图 7-11 CF卡

CF卡的大小为43 mm × 36 mm × 3.3 mm。CF卡同时支持3.3 V和5 V的电压,可以在这两种电压下工作。它有两种接口标准:CF Type II和CF Type I。从物理结构上看,CF Type II卡和CF Type I卡的每个插孔的间隔大小一样,但是CF Type II卡比CF Type I卡厚一些,为5 mm。CF Type I卡看起来只有Type II厚度的一半。另外,Type II比Type I插槽要宽,所以CF Type II卡不能在Type I卡插槽上使用。目前已经推出容量高达数GB的CF卡。CF卡主要在佳能、柯达、尼康等数码相机上使用。

CompactFlash 学会(<http://www.compactflash.org>)是一个发展该标准的国际组织。

2. SM(Smart Media Card)卡

SM卡(如图7-12所示)是日本东芝公司研制开发的小型存储卡,SM卡的尺寸为45 mm × 37 mm × 0.76 mm,且很轻、很薄,重只有约1.8 g。与大部分存储卡不同的是,SM卡由塑胶制成。SM存储卡上只有闪烁存储器模块和接口电路,本身没有控制芯片,所以其兼容性相对较差,没有得到广泛推广,仅在富士、奥林巴斯和三星的老款数码相机中使用,新推出的数码相机已经基本不再采用SM卡的了。

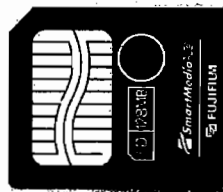


图 7-12 SM卡

固态软盘专题组(<http://www.ssfdc.org.jp/english>)是一个发展Smart Media Card标准的国际组织。

3. MMC(MultiMediaCards)卡

MMC(多媒体卡)是由美国SanDisk公司和德国西门子公司于1997年共同开发的多功能存储卡,如图7-13(a)所示,可用于手机、数码相机、数码摄像机、MP3等多种数码产品。它具有体积小、重量轻、高兼容性的特点,外形尺寸只有32 mm × 24 mm × 1.4 mm,重量在2 g以下,并且耐冲击,可反复读写30万次以上,驱动电压为2.7 V ~ 3.6 V。为更好地捕捉市场变化,满足当今的移动数码时代体积更小、重量更轻、容量更大的需要,还推出了更为小巧的迷你型RS-MMC卡,如图7-13(b)所示,体积为24 mm ×

18 mm × 1.4 mm,只有普通MMC卡的约一半大小。性能上和普通的MMC卡是相同的,且可以通过一个从后面安装的专用适配器,作为普通的MMC卡使用。RS-MMC主要应用于PDA、手机等移动存储设备上。由于后来推出SD卡标准中保留了设备对MMC卡的兼容,所以使用MMC卡的设备无法使用SD卡,但使用SD卡的设备却可以轻易使用MMC卡,如今已经推出最大容量为数GB的MMC卡。由于其对SD接口设备的高兼容性,所有应用于SD接口的设备都可以使用MMC卡。

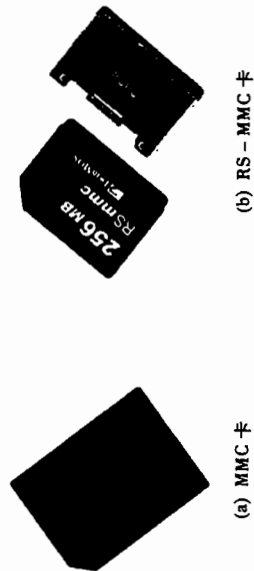


图 7-13 MMC卡和RS-MMC卡

多媒体卡协会(MultiMedia Card Association, www.mmca.org)于1998年创建,致力于促进MMC标准的开发和新产品的研制。

4. SD(Secure Digital Card)卡

SD卡由日本松下公司、东芝公司和美国SanDisk公司于1999年8月共同研制开发而成,具有大容量、高性能、高安全性等多种特点。SD卡是在MMC卡的基础上开发的,但是基本上两者还是不同的产品,只是在设计时考虑到兼容问题,所以在大多数情况下这两种产品能够互换使用。SD卡比MMC卡多了一个数字著作权保护功能,所采用的版权保护技术和DVD中使用的技术相同;多用于MP3、数码相机、电子图书、微型电脑和AV器材等。SD卡比MMC卡略微厚一些,尺寸为32 mm × 24 mm × 2.1 mm,如图7-14(a)所示。

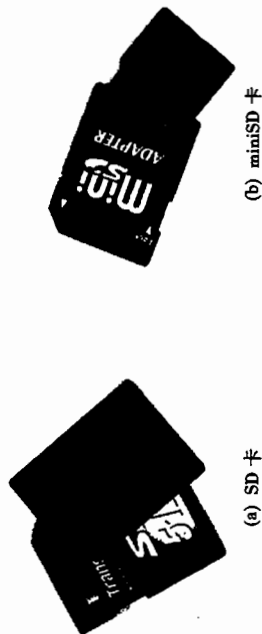


图 7-14 SD卡和miniSD卡

- SM (SmartMedia) 卡
- MMC (MultiMediaCards) 卡
- SD (Secure Digital Card) 卡
- 记忆棒 (Memory Stick)

1. CF (CompactFlash) 卡

CF 卡(如图 7-11 所示)是由美国 SanDisk 公司于 1994 年研制成功的,它是早些时候应用最为广泛的存储卡,可永久保存数据,并且无需电源,速度快,价格低于其他类型的存储卡。由于把存储模块与控制器集成在一起,这样 CF 卡的外部设备就可以做得比较简单而没有兼容性问题,特别是升级换代时也可以保证与旧设备的兼容性,保护了用户的投资,而且几乎所有的操作系统都支持这种卡。在笔记本计算机数码相机等数码产品中,CF 卡得到了非常广泛的应用。



图 7-11 CF 卡

CF 卡的大小为 43 mm × 36 mm × 3.3 mm。CF 卡同时支持 3.3 V 和 5 V 的电压,可以在这两种电压下工作。它有两种接口标准:CF Type II 和 CF Type I。从物理结构上来看,CF Type II 卡和 CF Type I 卡的每个插孔的间隔大小一样,但是 CF Type II 卡比 CF Type I 卡厚一些,为 5 mm。CF Type I 卡看起来只有 Type II 厚度的一半。另外,Type II 比 Type I 插槽要宽,所以 CF Type II 卡不能在 Type I 卡插槽上使用。目前已经推出容量高达数 GB 的 CF 卡。CF 卡主要在佳能、柯达、尼康等数码相机上使用。

CompactFlash 学会(<http://www.compactflash.org>)是一个发展该标准的国际组织。

2. SM (Smart Media Card) 卡

SM 卡(如图 7-12 所示)是日本东芝公司研制开发的小型存储卡,SM 卡的尺寸为 45 mm × 37 mm × 0.76 mm,且很轻、很薄,重只有约 1.8 g。与大部分存储卡不同的是,SM 卡由塑胶制成。SM 存储卡上只有闪烁存储器模块和接口电路,本身没有控制芯片,所以其兼容性相对较差,没有得到广泛推广,仅在富士、奥林巴斯和三星的老款数码相机中使用,新推出的数码相机已经基本不再采用 SM 卡的了。

固态硬盘专题组(<http://www.ssfdc.org.jp/english>)是一个发展 Smart Media Card 标准的国际组织。

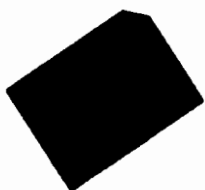
3. MMC (MultiMediaCards) 卡

MMC(多媒体卡)是由美国 SanDisk 公司和德国西门子公司于 1997 年共同开发的多功能存储卡,如图 7-13(a)所示,可用于手机、数码相机、数码摄像机、MP3 等多种数码产品。它具有体积小、重量轻、高兼容性的特点,外形尺寸只有 32 mm × 24 mm × 1.4 mm,重量在 2 g 以下,并且耐冲击,可反复读写 30 万次以上,驱动电压为 2.7 V ~ 3.6 V。为更好地捕捉市场变化,满足当今的移动数码时代体积更小、重量更轻、容量更大的需要,其还推出了更为小巧的迷你型 RS-MMC 卡,如图 7-13(b)所示,体积为 24 mm

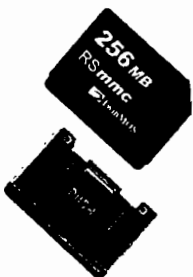


图 7-12 SM 卡

18 mm × 1.4 mm,只有普通 MMC 卡的约一半大小。性能上和普通的 MMC 卡是相同的,且可以通过一个从后面安装的专用适配器,作为普通的 MMC 卡使用。RS-MMC 主要应用于 PDA、手机等移动存储设备上。由于后来推出 SD 卡标准中保留了设备对 MMC 卡的兼容,所以使用 MMC 卡的设备无法使用 SD 卡,但使用 SD 卡的设备却可以轻易使用 MMC 卡,如今已推出的最大容量为数 GB 的 MMC 卡。由于其对 SD 接口设备的高兼容性,所有应用于 SD 接口的设备都可以使用 MMC 卡。



(a) MMC 卡



(b) RS-MMC 卡

图 7-13 MMC 卡和 RS-MMC 卡

多媒体卡协会 (MultiMedia Card Association, www.mmca.org) 于 1998 年创建,致力于促进 MMC 标准的开发和新产品的研制。

4. SD (Secure Digital Card) 卡

SD 卡由日本松下公司、东芝公司和美国 SanDisk 公司于 1999 年 8 月共同研制开发而成,具有大容量、高性能、高安全性等多种特点。SD 卡是在 MMC 卡的基础上开发的,但是基本上两者是不同的产品,只是在设计时考虑到兼容问题,所以在大多数情况下这两种产品能够互换使用。SD 卡比 MMC 卡多了一个数字著作权保护功能,所采用的版权保护技术和 DVD 中使用的技术相同,多用于 MP3、数码相机、电子图书、微型电脑和 AV 器材等。SD 卡比 MMC 卡略微厚一些,尺寸为 32 mm × 24 mm × 2.1 mm,如图 7-14(a)所示。



(a) SD 卡



(b) miniSD 卡

图 7-14 SD 卡和 miniSD 卡

214 第7章 外部存储器

SD 卡为 9 引脚,目的是通过把传输方式由串行变成并行,以提高传输速率,所以它的读写速率比 MMC 卡要快得多,最大的数据传输速率为 10 MBps,容量也达到了数 GB,这样在存储和传输一些较大的文件时具备一定的优势。从安全性的角度上讲,SD 卡可以通过特定的软件设置卡内程序或资料的使用权限,以避免其他人肆意复制,最大限度地保护了 SD 卡所有者的隐私。当然对于普通用户而言,这个功能没多大用处,而且有关的系统及加密资料会占用 SD 卡 900 KB 的容量。

从内部结构上来讲,SD 卡和 MMC 卡均采用了一体化固体介质,没有任何可移动的机械部分,不易损坏,所以不用担心一般程度上的物理撞击。同 RS-MMC 卡一样,SD 卡也推出更加小巧的 mini SD 卡,尺寸为 20 mm × 21.5 mm × 1.4 mm,重量仅为 1 g,通过插入形状大小和传统 SD 卡一样的专用适配器来实现与 SD 设备的兼容。

5. MS(Memory Stick)

MS 卡(图 7-15)由索尼公司于 1997 年研发推出,因其在形状上成长方形,所以我们又管它叫做“记忆棒”。目前只有索尼公司的产品支持记忆棒。与其他存储卡不同,记忆棒的规格是非公开的,目前只能在索尼公司的数码设备上使用。

记忆棒采用了单一平面的 10 引脚独立插槽设计,易于插拔而不易损坏。从尺寸规格上分,记忆棒有普通棒、高速棒(Memory Stick Pro)和短棒(Memory Stick Duo)三种。其中普通棒和高速棒的外形尺寸同为 50 mm × 21 mm × 2.8 mm,区别在于高速棒在存储卡的文件版权保护上比普通棒有所提高,加入了独立写保护开关,内含控制器,支持版权保护技术,拥有更高的读写速率,在存储容量上也有所提升。而短棒则像 mini SD 卡一样,把记忆棒的体积进一步缩减,长度约为普通棒的 1/2,通过一个适配器,可以像 RS-MMC 卡一样,把记忆棒的体积进一步缩减,长度约为普通棒的 1/2,通过一个适配器,可以像普通棒一样使用,但普通棒不能在短棒的机型上使用。

生产存储卡的公司在生产闪存卡时,也销售能将数据从存储卡读到计算机上的读卡器。典型的插入计算机 USB 端口的,可高速访问卡上数据的读卡器。

现在的读卡器大多数都是能读写多种卡的读卡器,简单的能够读写 MMC 和 SD 卡,包括这两种卡的小型版本卡,因为 MMC 卡和 SD 卡在尺寸上是相同的,而且接口兼容。全功能的读卡器几乎可以读写所有的常用存储卡。

存储卡的普及使用,使我们的移动存储方案多了一种选择。使用存储卡作为随身携带的移动存储设备,不但更加小巧,可以放在钱包里,而且可以在数码相机/摄像机或 MP3 播放机中使用。通过 USB 读卡器将存储卡连接到计算机上也很方便,操作和 U 盘一样。

习题

1/ 名词解释:

CRC、ECC、FAT、NTFS、CD-ROM、CD-R、CD-RW、CD-I、VCD、ISO 9660、UFD、DVD、DVD-ROM

DVD-Video、DVD+R、DVD-R、DVD+RW、DVD-RW、DVD-RAM、Combo、CAV、CLV、MMC、SD

2/ 硬盘内部由哪几部分组成,各部分有什么作用?

3/ 说明硬盘磁道、扇区、柱面之间的关系。

4/ 硬盘的“簇”是什么单位?它和硬盘的扇区是什么关系?

5/ 从硬盘的转速、介质传输速率和容量等几个方面比较标准和分区记录格式。

6/ 硬盘的缓存的作用是什么?

7/ 简述硬盘系统的 S.M.A.R.T. 技术。

8/ 分析比较 FAT 和 FAT32,它们各有什么优点?

9/ 简述 CD 光盘的扇区格式。

10/ 简述现代 CD-ROM 中使用的文件系统标准,Windows 2000/XP 支持哪些光盘文件?

11/ 从视频和音频质量、容量、记录时间等方面简述 VCD。

12/ 什么是 DVD5 和 DVD9?

13/ 比较 CD 和 DVD 的信息记录方面的技术指标。

14/ 分析 DVD 的信息记录数据,为什么双层 DVD 的容量不是单层 DVD 的两倍?

15/ 从容量、兼容性、成本等方面分析比较 DVD-RAM、DVD-R/RW、DVD+R/RW 三种可写 DVD 标准。

16/ 以 CD 或 DVD 为例,说明怎样计算光盘的准确物理容量。

17/ 计算 CD 和单面单层 DVD 的用户数据容量与物理容量的比值。

18/ 光盘驱动器的缓存容量对光盘的读写速率有什么影响?

19/ 分析比较光盘的 CAV 和 CLV 两种控制方式。

20/ 从性能和兼容性两个方面分析比较 MMC 卡和 SD 卡。

实验

1. 通过“优化大师”软件检测一台计算机的外存储器的配置情况:

(1) 硬盘型号、光盘驱动器型号;

(2) 硬盘的分区情况、接口类型、柱面数、磁头数、每磁道扇区数;

(3) 各个分区的总容量和剩余容量、文件系统、总簇数、簇容量、每个簇扇区数、扇区容量。

2. 使用软件 HARDINFO 测试硬盘、光盘和移动设备的参数:

(1) 文件系统、总容量和剩余容量;

(2) 光盘的传输速率;

(3) 硬盘和移动设备的寻道时间、缓存读写速率、介质读写速率。



第8章 常用外部设备

外部设备是计算机的重要组成部分,是人机交互的主要工具,方便、快捷、高效一直是人们研究和探索的主要课题。随着新技术的不断出现,新的外部设备也在不断地推出,例如,输入设备从键盘开始,发展到现在的鼠标、扫描仪、手写笔、光笔、条码阅读器、话筒、数码相机、摄像头等多种。

常用的外部设备主要包括显示系统、音频设备、打印输出设备、键盘与鼠标、扫描仪、触摸屏等,这些设备中多数是使用PC时必不可少的。本章将介绍这些常用外部设备的基本工作原理、主要技术指标和使用方法。

8.1 键盘与鼠标

键盘和鼠标是目前PC最广泛使用的输入设备,对于标准配置的PC而言是必不可少的。

8.1.1 键盘

键盘(Keyboard)是PC最常用也是最主要的输入设备。通过键盘,可以将英文字母、数字、标点符号等输入到计算机中,从而向计算机发出命令、输入数据等,使计算机完成不同的运算及控制任务。

计算机的键盘是从英文打字机键盘演变而来的,最早是以“电传打字机”的形式出现在计算机上的,是大型计算机和小型计算机时代最主要的人机交互式输入输出设备。20世纪70年代中期以后,随着显示器技术的成熟,电传打字机就逐渐退出了,而键盘则成为一种独立的设备。

键盘与主机的关系如图8-1所示,键盘一般通过设在主板上的专用键盘接口与主机相连,在主机的BIOS程序的控制下,将键盘输入的数据传送到主机,进而由CPU进行处理。

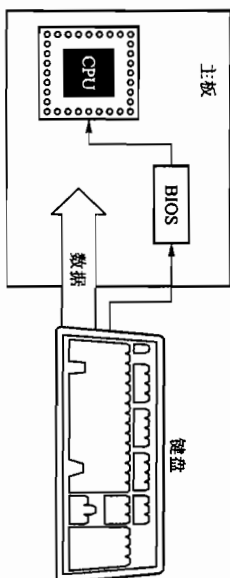


图 8-1 键盘与主机的关系

1. 键盘的工作原理

常见的键盘可以分为机械式和电容式两类。

(1) 机械式键盘的工作原理

图8-2所示的是一个最简单的键盘,其中每个键对应输入端口的一位。在没有键闭合时,各位均处于高电平,当某一键按下后,就会使对应的位接地而形成低电平。这样,CPU只要检测到某一位为“0”,便可识别出对应键已经按下。但这种结构的键盘有很大的缺点,就是当键盘上的键较多时,将会有许多引线,而且占用的输入端口较多。比如,一个有64个键的键盘,如果采用这种方法来设计,就需要64条连线和8个8位并行输入端口。所以,这种简单结构只用在仅有几个键的小键盘中。

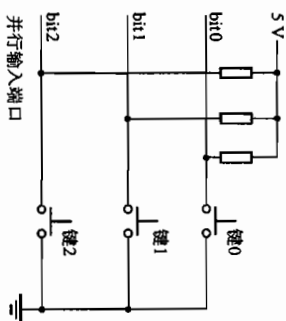


图 8-2 最简单的键盘结构

通常使用的键盘结构是矩阵式的,即由一组排列成矩阵形式的按键开关组成,如图8-3所示,这是一个有 $8 \times 8 = 64$ 个键的键盘,它需要16条线、1个8位并行输入端口和1个8位并行输出端口。为了识别键盘上键的闭合,常采用行扫描法,其原理如下:先使某行为低电平,其余行为高电平,通过检查列电平即可得知此行是否有键闭合,如果列为低电平则表示此行此列的键闭合;然后依次改变行,一直循环进行。例如,只有第1行、第2列上的键闭合了,当第1行为低电平时,第2列电平才会变成低电平,其他列仍为高电平,而其他时间所有列均为高电平。键盘中某一键闭合后,相应电路就给出的一组编码信息送到主机去进行识别及处理。

(2) 电容式键盘的工作原理

电容式键盘属于非接触式键盘,它们并不是依靠导电触点的机械连通来获得按键信号的,而是依靠按键本身的电参数变化来获得按键信号。图8-4所示的是一个电容式开关结构图,我们知道,电容的容量是由介质、两板的距离及两板的面积来决定的。所以当键按下时,两板的距离发生变化,这就引起电容量发生变化,当参数设计合适时,按键时就有振荡脉冲输出,这个输出再通过整形放大驱动编码器进行编码,编码后的按键信号通过键盘的通信部分传送给主机。

进入 20 世纪以后,机电打字机发明使得机械式打字机的铅字臂卡死不再成为一个严重的问题,众多的高速打字键盘也就应运而生,其中最著名的是德沃拉克(Dvorak)键盘。德沃拉克键盘是 August Dvorak 教授在 1930 年设计的键位方案。由于不再考虑按键的机械结构问题,所以按键排布完全按照理想化的击键率分布设计,手指运动的行程比柯蒂斯键盘要小得多,平均打字速度几乎提高了一倍。不过正如很多事情一样,习惯的力量是难以抵挡的,德沃拉克键盘至今只是在极少数专业场合使用。不过对于想试试的人来说,可以尝试一下 Windows 里自带的德沃拉克键盘方案。

键位设计的另一个概念就是附属键位的设计,从最早的 IBM PC 83 键盘到现在主流的 108 键,已经更新了几代,但总体上并没有根本性的变化。虽然其中有一些诸如紧凑型的设计,但从市场反应来看是不成功的。由此可见,目前的键盘键位设计经过了多年的实践检验,已经是非常成熟的理想设计。

2. PC 的键盘

PC XT/AT 时代的键盘主要以 83 键为主,随后又推出了 84 键的设计标准。虽然两种规格的键盘现在已经不多见了,但是目前键盘主要区域的划分仍然沿用当时的标准,至今没有什么变化。随之是 101 键和 104 键键盘,紧接着 104 键键盘出现的是多媒体键盘,它在传统的键盘基础上又增加了不少常用快捷键、音量调节装置等等,使 PC 操作进一步简化。对于收发电子邮件、打开浏览器软件、启动多媒体播放器等只需要按一个特殊按键即可,同时在外形上也做了重大改进,着重体现了键盘的人性化。起初这类键盘多用于品牌机,随着时间的推移,市场上也渐渐出现了独立的具有各种快捷功能的产品。

(1) PC 键盘的特点

PC 的键盘是一个智能终端,其内部有一个单片微处理器(单片微型计算机),负责检测键矩阵、消除按键信号的抖动、将按键信号转变成相应的扫描码和将扫描码传送给主机。早期的键盘几乎都使用机械金属弹簧作为弹力机构。这种键盘的手感硬、按键行程长、按键阻力变化明显,手感很接近打字机键盘。机械弹簧式金属键盘的机械弹簧很容易损坏,导致按键失灵。随着键盘技术的发展,一种更为简单可靠的薄膜开关成为当今键盘产品的主要类型。

图 8-5 所示的是目前流行的薄膜接触式键盘的内部结构。这种键盘在每个按键的位置上

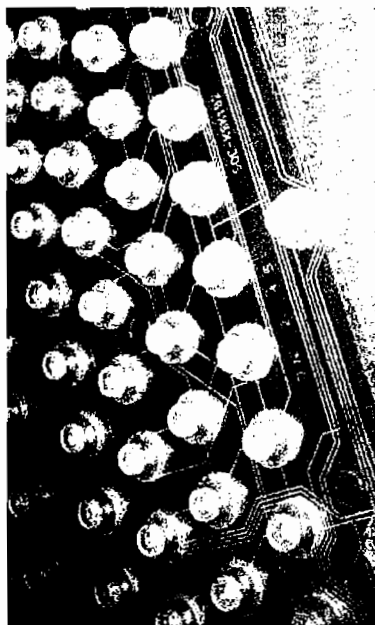


图 8-5 薄膜接触式键盘的内部结构

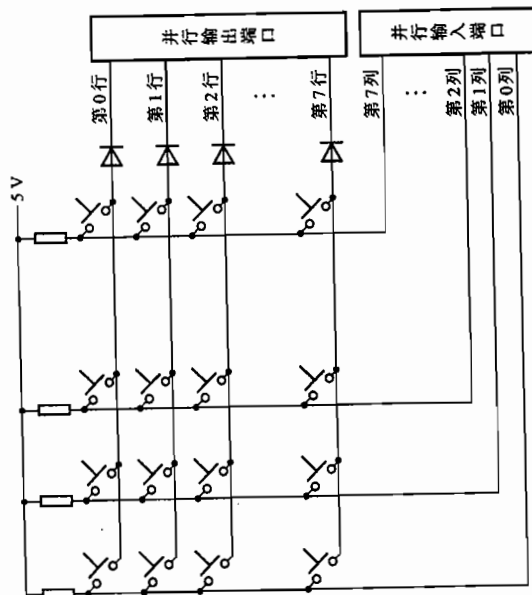


图 8-3 矩阵式键盘结构

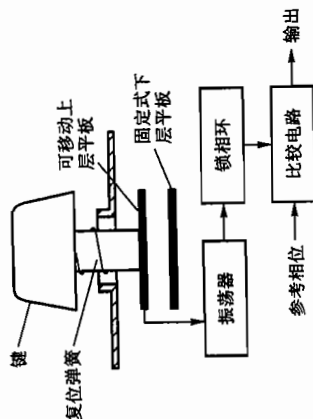


图 8-4 电容式键盘开关

(3) 键盘的键位设计

最常见的英文和数字键位设计就是俗称的“柯蒂斯”(OWERTY)键盘,是由 Christopher Latham Sholes 于 1868 年发明的键位方案。这样设计的根本原因在于机械式打字机的结构,其铅字杠杆的结构决定了当两个位置接近的铅字同时按下时就会卡死,但相对的两个相距较远的铅字就不会发生同样的问题。设计柯蒂斯键盘的最终目的是“在不会卡死的情况下尽量提高打字速度”。

有一个弹性键帽,是键盘的主要弹性元件。一款键盘的手感主要就是由这个部件的性状的材质决定的,因此其形状设计和橡胶成分都是各大键盘厂商的核心机密。其按键开关是3层重叠在一起的塑料薄膜,上下两层塑料薄膜上涂有起导线作用的线条,并在每个按键的位置上涂有触点,而中间一层塑料薄膜则不含任何导线,但在按键触点的位置上开有圆孔。在正常情况下,上下两层塑料薄膜上的触点被中间的塑料薄膜分隔开来,不会导通。但在按键时,上层塑料薄膜会向下弯曲,涂有触点的部位就会通过中间塑料薄膜的开口与下层塑料薄膜上的触点连接,从而产生一个按键信号。实际上它的基本原理和机械触点式键盘是一样的,依靠机械性的导电触点连过来产生按键信号。

目前的主流键盘除薄膜接触式键盘外,还有另外一种导电橡胶接触式键盘,其工作原理如图8-6所示。它只有一层导电薄膜,在每个按键位置上有不连通的两个触点,而橡胶弹簧的下部则使用导电橡胶来制作,当按下时就会将两个触点连通,产生按键信号。

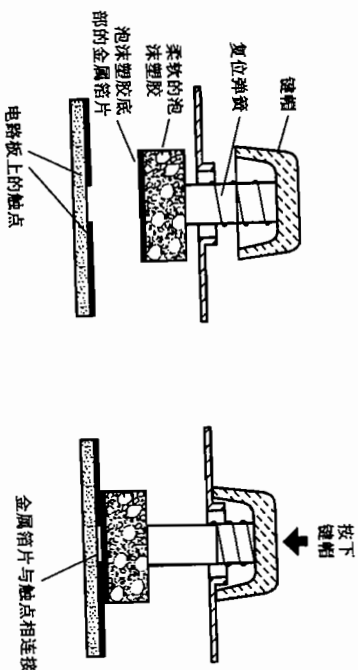


图 8-6 导电橡胶接触式键盘的原理

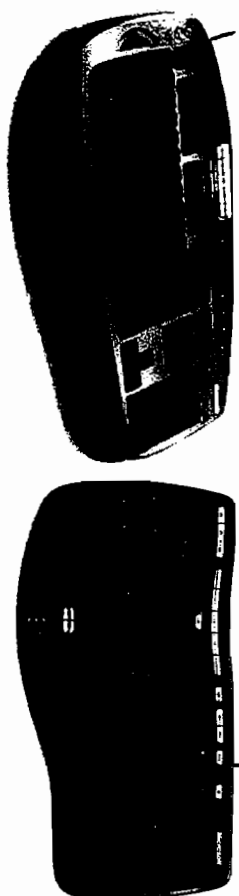
(2) PC 键盘的种类

在这个日新月异的世界,键盘的发展不仅仅从性能上加以改进,也越来越重视外形及多媒体运用,一些具有特殊功能的键盘应运而生。从按键的数量上看,可分为83键、84键、101键、102键、103键、104键、105键、108键、109键等。较早的键盘主要是83键,但随着这几年Windows系统的广泛应用,已经被淘汰。目前市场占有率主流地位的是104键和108键的键盘。108键在传统104键的基础上增加了4个Windows功能键:Power键、Sleep键、Wake Up键、Fn组合键,使操作计算机更加容易,得心应手。对于以前需要打开好几个窗口才能完成的操作,通过设定,只需一个按键指令即可轻松完成。

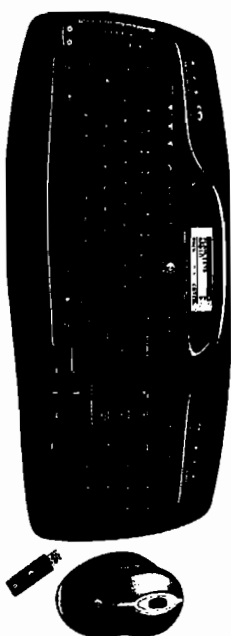
从外形和功能上看,键盘可分为人机工程学键盘、集成鼠标的键盘、无线键盘、防水键盘、橡胶软性键盘、集成USB键盘、手写板键盘、多媒体键盘等。

1) 多媒体键盘

多媒体键盘的特征是通过驱动程序设定,可以使用键盘所提供的特殊的快捷操作键实现CD播放、音量调整、键盘开关主机、休眠、唤醒等功能,如图8-7(a)所示。由于这些附加功能目



(a) 多媒体键盘



(b) 人机工程学键盘

(c) 无线键盘与鼠标

图 8-7 多功能键盘

前尚没有统一的标准,所以具体键盘提供的快捷键的数量和功能均不相同。具体的快捷键也有两种形式,一种是利用键盘上的原有但不常用的按键,如F功能键和数字键提供扩展定义;另一种比较彻底,采用独立的按键实现一键专用,更有甚者在键盘增添一个步进电位器实现音量的控制,比起采用增音键和减音键来调整更为直观。

2) 人机工程学键盘

一般键盘平直的键位排列使得手腕必须正对键盘,从而导致双臂不能自然放松;其次,由于所有的键都在一个平面上,这样双手的手掌就不能保持一个自然相对的姿势,而必须转动手腕使双手手掌都保持正面向下;再次,由于按键高于手腕,所以手指不能保持自然弯曲下垂的姿势,而必须努力上举。传统键盘的三大缺陷使得很多打字人员患上了肩周炎等职业病。

20世纪90年代中期,微软推出了其第一代人机工程学键盘,从而奠定了现代人机工程学键盘的设计标准。微软人机工程学键盘的主键盘部分设计成分开的两部分,如图8-7(b)所示,左右两部分呈弧形向两个方向分开,这样的排布使得手臂不必像传统键盘一样靠在一起,而可以呈自然的“八”字放置,解决了肩部的受力问题;微软人机工程学键盘的中部比左右两边高,这样手掌放在上面的时候,手掌心可以呈自然相对的姿态,这样就解决了手腕的扭曲问题;微软人机工程学键盘增加了腕托,而且各键中位置最高的是空格键,位置最低的则是Esc键和Backspace键,这样就符合手掌自然放置时手腕和拇指最高而其他四指自然下垂的姿态。

3) 集成鼠标的键盘

在键盘上集成的鼠标多采用轨迹球或压力感应板的形式,理论上可以节省桌面空间,对于那些未提供 PS/2 鼠标接口且串口资源又比较紧张的计算机,可以节省一个端口。这种键盘的操作与笔记本计算机键盘相似,比较适用于习惯使用笔记本计算机的操作者。

4) 无线键盘

顾名思义,无线键盘的键盘体与主机之间没有直接的物理连线,而是通过红外线或无线电波将输入信息传送给特制的接收器,如图 8-7(c)所示。接收器的连接与普通键盘基本相同,也只需简单地连接到 PS/2、COM 或 USB 接口上即可。无线键盘需要使用干电池供电。对于红外线型的无线键盘具有较严格的方向性,尤其是水平位置的关系更为敏感,而采用无线电的键盘要灵活得多。考虑到无线电是辐射状传播的,为了避免在近距离内有同类型(同频率)的键盘工作导致互相干扰,一般都有 4 个以上的频道,如遇干扰可以手动转换。无线键盘为了配合移动的需要,一般体积较为小巧并集成有鼠标的功能,键位与笔记本计算机相仿,省略了数字小键盘,将常用的功能键融合在键盘的边缘位置。

另外,还出现了带特殊功能的键盘,如带扫描仪的键盘,它之所以产生是因为在商务办公中有很多地方需要用到扫描,而平板扫描仪虽然扫描质量好但比较占用桌面空间,所以有了带扫描仪的键盘,它前部集成有带状式的扫描仪,特点是可扫描 A4 幅面的文件、可发真迹传真、精度较低。带有身份识别功能的键盘,它是在普通键盘上集成了条形读卡器,这种键盘在一些机要部门(如银行)很有用处。

(3) PC 键盘的接口

目前,键盘与 PC 主机的接口(见表 8-1)可分为 PC XT/AT 接口(俗称“大口”)、PS/2 接口(俗称“小口”)和 USB 接口。PC XT/AT 接口首先应用在 PC XT/AT 计算机上,一直延续至早期的 Pentium 计算机。实际上 PC XT/AT 接口是 AT 结构的主板的标准键盘接口,而 PS/2 接口键盘如今被广泛应用于 ATX 结构的主板上,PC XT/AT 和 PS/2 接口采用的是德国标准化协会(Deutsches Institut für Normung, DIN)制定的标准。

表 8-1 键盘的接口

接口类型	引脚号	信号定义	接口外形
PC/AT 接口	1	键盘时钟	
	2	键盘数据	
	3	空	
	4	接地	
	5	+5 V	

续表

接口类型	引脚号	信号定义	接口外形
PS/2 接口	1	键盘数据	
	2	空	
	3	接地	
	4	V _{cc}	
	5	键盘时钟	
	6	空	
USB 接口	1	V _{cc}	
	2	- Data0	
	3	接地	
	4	+ Data0	

3. 笔记本计算机的键盘和鼠标

(1) 笔记本计算机的键盘

由于体积的限制,笔记本计算机键盘在与台式机键盘具有同样功能的前提下,还要有更小的面积和厚度,同时又不能有太长的按键行程。在这种情况下,要做到同样舒适的手感,其设计就比台式机键盘更加复杂。

笔记本计算机的键盘一般为 83 键或 86 键,比台式机键盘少了小键盘及一些重复的控制键。虽然总键数有所减少,但不少功能键身兼数职,使得笔记本计算机键盘的功能并不比台式机键盘少,有些型号的笔记本计算机甚至把控制显示方式、音量调节等功能也安排在功能键上,省掉了专门的控制按钮。

尽管外表差别很大,但现在的笔记本计算机键盘也大都都是薄膜接触式键盘,其基本结构与标准的台式机键盘很类似。但是由于其按键行程短,底板受力大,同时也是为了散热的需要,所以笔记本计算机键盘的底板全是金属的。同时,由于笔记本计算机键盘的按键行程很短,只能使用橡胶弹簧。由于压缩幅度很小并不能提供足够的弹力回馈,因此在笔记本计算机上都使用了支架设计。

所谓支架设计,就是指在笔记本计算机键盘键帽的下方,除了橡胶弹簧以外,还有两个硬质的塑料或金属支架,当按键按下的同时,两个支架会受压而向侧面伸出,从而压缩支架侧面的橡胶,这样就将向下的压力转为向侧面的压力,由于侧面有足够的压缩空间,就能够在按键垂直行程很短的情况下也能获得足够的抗力变化。绝大多数笔记本计算机键盘使用的都是“X”支架,即键下的支架是两个呈“X”状交叉的支架。这种支架的稳定性高、耐用性好,所以使用非常普遍。联想最新一代的笔记本计算机上已经取消了“X”型支架的设计,而改用“几”字形的支架。

一般来说,在 13 英寸和更大尺寸的笔记本计算机上,都能安装标准尺寸的笔记本计算机键盘,它们的按键尺寸和键间距都与台式机键盘接近。而在 12 英寸笔记本计算机上,一般安装的

是95%尺寸的笔记本计算机键盘,它们与标准尺寸的笔记本计算机键盘相比缩短了键间距和按键行程,但按键尺寸一般没有变化,整体上与标准尺寸的键盘差别不明显。10英寸的超薄笔记本计算机上则使用的是80%尺寸的键盘,与标准的笔记本计算机键盘相比,在各方面都能明显感觉到其尺寸的缩小,特别是按键行程的减小极为明显。但在此类笔记本计算机中的一个特例是IBM。IBM一向坚持键盘手感的要求,所以即便在10英寸笔记本计算机上,它也坚持使用最小95%的键盘。

与台式机相比,笔记本计算机键盘有很多的快捷键和特殊的Fn按键。通过该键和其他按键的组合,可以提供诸如硬件调节、数字小键盘等功能,而且可以提供很多自定义的功能。一般而言,笔记本计算机键盘的手感是永远比不上台式机键盘的,特别是它的按键行程太短导致停止的冲击力太大,容易疲劳。

(2) 笔记本计算机的指点设备

目前笔记本计算机内置的常见指点设备主要有轨迹球、触摸板和指点杆,其外观都与标准鼠标大相径庭,但功能是一致的。

1) 轨迹球

轨迹球的特点是体积较大,比较重,容易磨损和进灰尘,且定位精度的能力一般,现在轨迹球已经被淘汰了。

2) 触摸板

触摸板是目前使用最为广泛的笔记本鼠标,许多品牌笔记本计算机均配有触摸板。触摸板由一块能够感应手指运行轨迹的压感板和两个按钮组成,两个按钮相当于标准鼠标的左右键。触摸板设有机械磨损,控制精度也不错,最重要的是,它操作起来很方便,初学者很容易上手,一些笔记本计算机甚至把触摸板的功能扩展为手写板,可用于手写字的输入。

3) 指点杆(Track Point)

指点杆是由IBM发明的,目前常见于IBM和东芝的笔记本计算机中。它有一个小按钮,位于键盘的G、B、H三键之间,在空格键下方还有两个大按钮,其中小按钮能够感应手指推力的方向和大小,并由此来控制鼠标指针的移动轨迹,而大按钮相当于标准鼠标的左右键。指点杆的特点是移动速度快,定位精确,但控制起来却有点困难,初学者不容易上手,但不少用户在掌握了指点杆的使用诀窍后,往往对它爱不释手。

触摸板和指点杆各有所长,它们都拥有不少喜爱者。于是惠普公司在其笔记本计算机中同时内置了这两种鼠标,以满足不同用户的需要。不过,无论是触摸板还是指点杆,它们使用起来都不如标准鼠标方便,不适合于一些操作复杂和精确要求高的场合,因此,必要时可以另接一个USB接口的标准鼠标。

8.1.2 鼠标

鼠标器简称为鼠标(Mouse),鼠标的点击可代替键盘那繁琐的指令,使计算机的操作更加简便。特别是具有图形用户界面(Graphical User Interface, GUI)的操作系统为人机交互提供了直观便捷的手段,通过诸如鼠标等指示设备可方便地对计算机进行操作。通过鼠标能方便地将光标准确定位在指定的屏幕位置,操作按键和滚轮完成各种功能。因此可以说,在图形用户界面操作系统中鼠标是必不可少的输入设备。

世界上第一个鼠标是由美国人 Douglas C. Engelbart 博士于1968年12月9日发明的。随着Windows操作系统的不断普及和升级,鼠标作为计算机的一个最不起眼的输入设备身价陡升,在某些场合它的重要程度甚至超过了键盘。经过几十年的发展,鼠标技术有了长足的进步,由最初的机械式鼠标,发展到光电式鼠标、激光式鼠标,其他如轨迹球、笔记本计算机上的指点杆和手指感应式鼠标也在不断发展。鼠标向着多功能、多媒体、符合人机工程学的方向继续发展。

1. 鼠标的工作原理

常见的鼠标可以分为机械式、光电式和激光式三大类。

(1) 机械式鼠标的工作原理

机械式鼠标的结构最为简单,由鼠标底部的胶质滚球通过摩擦作用带动X方向滚轴和Y方向滚轴,在滚轴的末端有译码轮,译码轮附有金属导电片与电刷直接接触,接触译码轮的电刷随即产生与二维空间位移相关的脉冲信号。

由于电刷直接接触译码轮和鼠标小球与桌面直接摩擦,所以精度有限,电刷和译码轮的磨损也较为厉害,直接影响机械鼠标的寿命。因此,机械式鼠标已被淘汰。

(2) 光电式鼠标的工作原理

光电式鼠标是一种光电和机械相结合的鼠标,是在机械式鼠标的基础上发展起来的,它将磨损最厉害的电刷和译码轮改进成为非接触式的光电机构。光电式鼠标的外形与机械鼠标没有区别,习惯上人们仍将光电式鼠标称为机械式鼠标。

光电式鼠标的工作原理非常简单,其结构如图8-8所示,主要由发光二极管(LED)、光栅轮和光接收器件组成。由鼠标底部的胶质滚球通过摩擦作用带动X方向滚轴和Y方向滚轴。滚轴连接圆盘状的光栅轮,其上带有类似百叶窗的扇形缝隙,当滚轴转动时,缝隙产生与二维空间位移相关的脉冲信号。由于采用的是非接触部件使磨损率下降,从而大大地提高了鼠标的寿命,也能在一定范围内提高鼠标的精度。

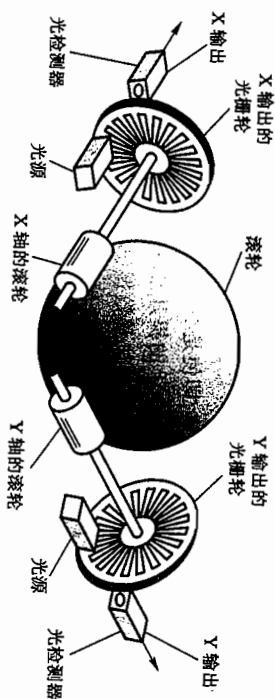


图 8-8 典型的光电式鼠标器结构

(3) 光电式鼠标的工作原理

光电式鼠标是目前市场上最常见的一种鼠标。

1) 第一代光电式鼠标

在光机式鼠标出现的同时,还出现了一种完全没有机械结构的全数字化的第一代光电式鼠标。这种鼠标的主要部件为发光二极管、感光芯片、控制芯片和一个带有网格的专用反射鼠标垫。

其原理是通过发光二极管、反射鼠标垫和感光芯片的协作来测量鼠标的位移。工作时,光电鼠标必须在反射鼠标垫上移动,X发光二极管和Y发光二极管会分别发射出光线照射在反射鼠标垫上,部分光束将经过镜头组件传递后照射在感光芯片上,感光芯片将光信号转换为对应的数字信号后送到定位芯片中专门处理,进而产生X-Y坐标偏移数据。

光电式鼠标由于接触部件较少,鼠标的可靠性和精度都得到较大提高。但它在后来的应用中暴露出大量的缺陷。首先,光电式鼠标必须依赖反射鼠标垫,它的位置数据完全依据反射鼠标垫中的网格信息来生成,若反射鼠标垫被弄脏或者磨损后,光电鼠标便无法判断光标的位置所在;若反射鼠标垫不慎被严重损坏或遗失,那么整个鼠标便就此报废。其次,光电式鼠标的使用方法非常不人性化,它的移动方向必须与反射鼠标垫上的网格纹理相垂直,用户不可能快速地将光标直接从屏幕的左上角移动到右下角。由于存在大量的缺陷,这种光电式鼠标并未得到流行,很快就被市场所淘汰。

2) 第二代光电式鼠标

与第一代光电式鼠标的工作原理完全不同,其核心部件是发光二极管、微型摄像头、光学引擎和控制芯片。工作时发光二极管发射光线照亮鼠标底部的表面,同时微型摄像头以一定的时间间隔不断进行图像拍摄。鼠标在移动过程中产生的不同图像传送给光学引擎进行数字化处理,最后再由光学引擎中的定位数字信号处理器(Digital Signal Processor, DSP)芯片对所产生的图像数字矩阵进行分析。由于相邻的两幅图像总会存在相同的特征,通过对比这些特征点的位置变化信息,便可以判断出鼠标的移动方向与距离,这个分析结果最终被转换为坐标偏移量实现光标的定位。鼠标内部的光二极管主要用于照明,以便微型摄像头拍摄到鼠标底部画面,进而传送到光学处理芯片。

3) 激光式鼠标

激光式鼠标原理与第二代光电式鼠标差不多,只是把发光二极管换成了激光二极管来照射鼠标所移动的表面。激光光线具有一致的特性,当光线从表面反射时可产生高反差图形,出现在传感器上的图形会显示物体表面上的细节,即使是在光滑的表面也可使用。

激光式鼠标的优势主要是表面分析能力上的提升,借助激光引擎的高解析能力,能够非常有效地避免传感器接受到错误或者是模糊不清的位移数据,更为精确的表面数据反馈将会非常有利于鼠标的定位,这样就可以在很多光电式鼠标无法使用的表面进行操作。精确度要比光电式鼠标提高约20倍。

2. 鼠标的种类

(1) 按键数分类

鼠标可以分为双键、三键和多键鼠标。根据微软最早的定义,鼠标只需要左右两个键。双键鼠标结构简单,应用广泛,一般无需另外安装驱动程序就可以在Windows 95/98操作系统下正常工作。三键鼠标是IBM在双键鼠标的基础上进一步定义而成的,又被称为PC鼠标。与双键鼠标

相比,三键鼠标上多了个中键,使用中键在某些特殊程序中往往能起到事半功倍的作用。例如,在AutoCAD软件中可以利用中键快速启动常用命令,使工作效率成倍提高。早期的三键鼠标一般都有一个微型拨动开关,用于切换双键与三键两种不同工作方式。为了与现有的操作系统兼容并发挥中键的作用,很多产品都配备自己的第三方驱动程序,在Windows系统中将中键设置成某一常用功能的快捷键。多键鼠标是在微软发布智能鼠标(IntelliMouse)之后,鼠标向多功能应用领域发展而产生的新一代多功能鼠标。微软智能鼠标带有滚轮,使得上下翻页变得极其方便,在Office软件中可实现多种特殊功能。随着应用的增加,之后其他厂商生产的新型鼠标除了有滚轮,还增加了拇指键等快捷按键,进一步简化了操作程序。多键多功能鼠标将是鼠标未来发展的目标与方向。

(2) 按接口分类

将鼠标连接到计算机所用的插头取决于接口的类型。最常用的鼠标接口有3种:串行接口、专用主板鼠标接口(PS/2鼠标接口)和USB接口。

1) 串行接口

早期的鼠标主要是通过串行接口(COM端口)与主机相连接的。同其他串行设备一样,它占了一个串行通信端口。由于大多数PC机带有两个串行接口,鼠标可以任意插在COM1或COM2上。设备驱动程序在系统启动过程中会搜索端口来判断鼠标现在所插的串行接口。目前串行接口的鼠标已基本淘汰。

2) 专用主板鼠标接口(PS/2鼠标接口)

目前生产的台式计算机的主板上都有一个内置的专用鼠标口,这种接口最先由IBM于1987年在PS/2系统中实现,所以也称为PS/2鼠标接口。此称呼并不表示这种接口只能在PS/2系统中,相反,只要主板上有这样的专用接口的任何系统都可以用这种鼠标。PS/2鼠标通过一个6引脚微型DIN接口与计算机相连,它与键盘的接口非常相似,使用时注意区分,PS/2鼠标接口一般为绿色。PS/2鼠标是目前市场上的主流产品,目前市场份额已经相当的大,而且种类繁多,造型多样。

3) USB接口

随着USB接口的普及,现在采用USB接口的鼠标越来越多。与前两种接口相比,USB接口的鼠标具有连接方便、热插拔、即插即用等一系列优点。

(3) 无线鼠标和3D振动鼠标

无线鼠标是为了适应大屏幕显示器而生产的,其工作原理与无线键盘类似。3D振动鼠标是一种新型的鼠标,此类鼠标的主要特点是具有全方位立体控制能力。3D振动鼠标具有前、后、左、右、上、下六个移动方向,而且可以组合出前右、左右、下等移动方向;具有振动功能,即触觉反馈功能。玩某些游戏时,当你被敌人击中时,你会感觉到你的鼠标也振动了;是真正的三键式鼠标,无论是在DOS或Windows环境下,鼠标的中间键和右键都大派用场。

3. 鼠标的性能指标

一个鼠标的性能是否优异,主要有几种参数:移动精度、刷新率、光学引擎、像素处理能力、微动等。

(1) 移动精度

移动精度也称为分辨率(单位: dpi, dots per inch), 表明鼠标的精确度。移动精度越高, 鼠标精确度就越高。机械式鼠标一般达 100 dpi ~ 300 dpi, 光电式鼠标一般达到 400 dpi ~ 900 dpi。

一般用户选用精度为 300 dpi ~ 500 dpi 的鼠标就行了。对于图形处理或某些特殊用途, 则需要更高的精度来进行精细的处理。

(2) 刷新率

刷新率(单位: 次/秒, 有些资料用 Hz 表示)也叫做内部采样率、扫描频率、帧速率等, 它是描述鼠标光学系统采样能力的参数: 发光二极管发出光线照射工作表面, 微型摄像头以一定的频率捕捉工作表面反射的快照, 交由 DSP 分析和比较这些快照的差异, 从而判断鼠标移动的方向和距离。刷新率目前为数百次每秒至近万次每秒。

很显然, 刷新率的高低决定了图像的连贯性好坏以及对微小移动响应的灵敏度。刷新率越高, 在相同的时间内获得的信息就越充分, 图像就越连贯, 帧之间的对比也就更有效、更准确, 表现在实际使用效果上则是鼠标的反应将更加敏捷、准确和平稳, 而且对任何细微的移动都能做出响应。

(3) 光学引擎

光学引擎的成像系统由光源、具有一定放大倍率的镜头和互补金属氧化物半导体(Complementary Metal-Oxide Semiconductor, CMOS)晶阵构成, 其性能是决定移动精度的主要因素。

1) 光学放大倍率

光学放大倍率决定了入射的可侦测面积与细节。光学传感器首先将工作表面的反射图像进行光学放大, 然后投射到 CMOS 晶阵上形成帧。在 CMOS 的面积一定的情况下, 若光学放大倍率较大则 CMOS 获得的图像细节较多, 提高了图像的分辨率; 但与此同时, 实际的捕捉面积则会缩小, 若移动速度过大则会造成丢帧, 这种弊端可以通过提高刷新率得以补偿。若光学放大倍率较小, 则有利于提高捕捉能力, 而不利分辨率的提高, 但可以通过提高 CMOS 的分辨率(即像素数)进行补偿。

2) CMOS 晶阵像素数

CMOS 晶阵像素数决定了经过放大后光学引擎实际成像的面积与细节。像素数 = CMOS 的面积 \times 密度。因此, 在光学放大倍率一定的情况下, 提高 CMOS 的像素数有两种方式: 增大 CMOS 面积以获得更大面积的图像, 或者提高 CMOS 晶阵密度以获得更丰富的细节, 或者两种方式同时采用。另一方面, 若通过提高密度来提高像素数则可提高移动精度, 若密度过低则会损失图像的细节; 若因 CMOS 面积过小导致的性能缺陷则需提高刷新率进行补偿, 但通常 CMOS 的面积会受到一定限制, 因此主要的改进方向在于提高密度。

3) 最大速度和最大加速度

像素处理能力虽然十分科学, 但是毕竟不够直观, 所以将其与移动精度参数相结合, 可以派生出最大速度和最大加速度两个参数。最大速度及最大加速度通常会被作为一项指标用于衡量鼠标的性能。

鼠标的最大加速度是通过 DSP 运算使鼠标在保证精确性的前提下能够达到的最大加速度。当使用者用鼠标单击屏幕上的不同位置时, 移动过程通常都带有突然的加速和减速。在玩游戏时, 鼠标经常需要快速移动, 这种加速度的变化就更加的常见。

根据实验, 入手在使用鼠标的时候, 最高的移动速度约为 30 英寸/秒, 但是早期的光电式鼠

标能够承受的最大位移只有 15 英寸/秒, 这也是造成丢帧的根本原因。罗技的 MX 光学感应器可以在 10 g 的加速度或 40 英寸/秒的速度下正常工作, 这已经远远超过了常人所能达到的极限。

4) 像素处理能力

罗技在发布 MX 引擎时引入了“像素处理能力”这一指标, 这一指标能够更加直观地说明光电式鼠标的性能, 其单位为像素/秒, 计算公式为: 像素处理能力 = 每帧像素数 \times 刷新率。

在提高处理能力的途径上, 罗技与微软走的是两条道路: 微软的方法是单纯提高帧速率(高达 6 kHz 以上), 而罗技则在提高帧速率(5.2 kHz)的同时提高了像素数。罗技 MX 引擎的像素处理能力为 4.7×10^6 像素/秒, 而通过换算, 微软 IntelliEye 光学引擎像素处理能力为 2.9×10^6 像素/秒, MX 引擎领先 62% 之多。

2 扫描仪

扫描仪是除键盘和鼠标之外被广泛应用于计算机的输入设备。扫描仪是一种光机电一体化的高科技产品。它是将各种形式的图像信息输入计算机的重要工具, 是继键盘和鼠标之后的第三代计算机输入设备。从最直接的图片、照片、胶片到各类图纸、图形以及各类文稿资料, 都可以用扫描仪输入到计算机中, 进而实现对这些图像形式信息的处理、管理、使用、存储、输出等。配合光学字符识别(OCR)软件, 还可以将扫描的文稿转换成可编辑的文本形式。目前扫描仪已广泛应用于各类图形图像处理、出版、印刷、广告制作、办公自动化、多媒体、图文数据库、图文通信、工程图纸输入等诸多领域, 极大地促进了这些领域的技术进步, 甚至使一些领域的工作方式发生了革命性的变化。

1. 扫描仪的分类

扫描仪有很多种, 按不同的标准可分成不同的类型。按扫描原理可将扫描仪分为以电荷耦合器件(Charge Couple Device, CCD)为核心的平板式扫描仪、手持式扫描仪和以光电倍增管为核心的滚筒式扫描仪; 按扫描图像幅面的大小可分为小幅面的手持式扫描仪、中等幅面的台式扫描仪和大幅面的工程图扫描仪; 按扫描图像的介质可分为反射式(纸材料)扫描仪和透射式(胶片)扫描仪以及既可扫描反射稿又可扫描透射稿的多用途扫描仪; 按用途可将扫描仪分为可用于各种图像输入的通用型扫描仪和专门用于特殊图像输入的专用型扫描仪(如条形码扫描器、卡片阅读机等)。

1) 手持式扫描仪

手持式扫描仪体积小、重量较轻、携带比较方便, 但扫描精度较低, 扫描质量和扫描幅面与平板式扫描仪相比都有较大的差距, 曾一度拥有价格优势, 也随着平板式扫描仪价格的大幅下降而不复存在, 已处在被淘汰的边缘。

2) 滚筒式扫描仪

大幅面扫描仪一般指扫描幅面为 A0(扫描宽度为 900 mm, 长度不限)或以上的扫描仪。因为扫描幅面大, 为减小机器体积多半采用滚筒式走纸机构。在扫描时扫描头是固定的, 图纸在

走纸机构控制下移动,将图纸全部扫描一遍,一幅完整的图像就输入到计算机中了。滚筒式扫描仪主要用于扫描彩色海报、建筑草图、工程图、地图、地形图、美术图等,这些应用领域都要求更为精确和高质量的扫描效果。大幅面工程图纸的输入,为CAD、工程图管理等方面应用提供了输入手段,另外在测绘、勘探、地理信息系统等方面也有许多应用。滚筒式扫描仪近年来发展很快,产品种类和用户都在迅速增加。目前国内的CAD应用正在飞速发展,生产、设计、研究等部门都有大量的图纸要输入计算机中进行处理,加上近年来滚筒式扫描仪的性能有了很大的进步,与工程图输入配套的矢量量化软件的功能也有了很大的改进,再加上新出现了一些在光栅方式下对工程图进行编辑处理的软件,都将促使对滚筒式扫描仪的需求进一步扩大,滚筒式扫描仪市场将稳步发展。

3) 平板式扫描仪

平板式扫描仪主要应用在A4和A3幅面,其中又以A4幅面的扫描仪用途最广、功能最强、种类最多、销量最大,是扫描仪家族的代表性产品。经过多年来的发展,目前平板式扫描仪的性能已经达到了很高的水平。分辨率通常为600 dpi~1200 dpi左右,高的可达2400 dpi。色彩数一般为30位,高的可达36位。扫描时将图稿放在扫描台上由软件控制自动完成扫描过程,速度快、精度高。有些平板扫描仪还可以加上透明胶片适配器,使其既可以扫描反射稿又可以扫描透射稿,实现一机两用。平板扫描仪已广泛应用于各类图形图像处理、电子出版、印前处理、广告制作、办公自动化等诸多方面,其性能几乎可以满足所有应用领域的要求。由于目前无论是在商用或是家庭方面,平板式扫描仪最为常见,因此以下主要介绍平板式扫描仪。

2. 扫描仪的组成部分及工作原理

扫描仪主要由光学成像部分、机械传动部分和转换电路部分组成,如图8-9所示。这几部分相互配合将反映图像特征的光信号转换为计算机可识别的电信号。扫描仪的核心是完成光电转换的光电转换部件。目前大多数扫描仪采用的光电转换部件是CCD,它可以将照射在其上的光信号转换为对应的电信号。其他主要部件有:光学成像部分的光源、光路和镜头;转换电路部分的A/D转换器;控制机械部分运动的控制电路和机械传动机构的步进电机、扫描头和导轨等。扫描仪工作时首先由光源将光线照在欲扫描的图稿上,产生表示图像特征的反射光(反射稿)或透射光(透射稿),光学系统采集这些光线并将其聚焦在CCD上,由CCD将光信号转换为电信号,

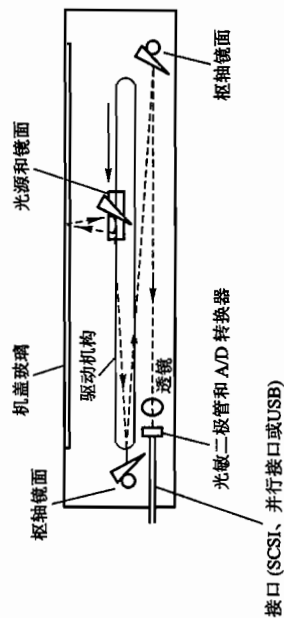


图8-9 扫描仪的结构

然后由电路部分对这些信号进行A/D转换及处理,产生对应的数字信号输送给计算机。当机械传动机构在控制电路的控制下带动装有光学系统和CCD的扫描头与图稿进行相对运动,将图稿全部扫描一遍时,一幅完整的图像就输入到计算机中了。

3. 扫描仪的主要性能指标

(1) 光学分辨率

光学分辨率是指扫描仪的光学系统可采集的实际信息量,也就是扫描仪的感光元件的分辨率。光学分辨率是扫描仪最重要的性能指标之一,它直接决定了扫描仪扫描图像的清晰程度。扫描仪的分辨率通常用每英寸长度上的点数(dpi)来表示。低档扫描仪的光学分辨率通常为300 dpi~600 dpi,中高档扫描仪的光学分辨率通常为600 dpi~1200 dpi。另外,除了光学分辨率之外,扫描仪的包装箱上通常还会标注一个最大分辨率,光学分辨率为300 dpi~600 dpi的扫描仪的最大分辨率一般为4800 dpi,而600 dpi~1200 dpi扫描仪的最大分辨率则高达9600 dpi。最大分辨率实际上是通过软件在真实的像素点之间插入经过计算得出的额外像素,从而获得的插值分辨率。插值分辨率对于图像精度的提高并无好处,事实上只要软件支持而你的机器又可提供支持的话,这种分辨率完全可以做到无限大。对于个人用户,300 dpi~600 dpi的扫描仪通常就能满足需要。

(2) 色彩位深与灰度值

在进行彩色图像扫描时,位深是一个重要的指标。较高的色彩位深可以保证扫描仪反映的图像色彩与实物的真实色彩尽可能一致,而且图像色彩会更加丰富。扫描仪的色彩位深一般有24位、30位、32位和36位几种。光学分辨率为300 dpi~600 dpi的扫描仪其色彩位深一般为24位或30位,而600 dpi~1200 dpi扫描仪的位深一般为36位,最高的可达48位。灰度值是指进行灰度扫描时对图像由纯黑到纯白整个色彩区域进行划分的级数。编辑图像时一般使用8位,即256级,而主流扫描仪通常为10位,最高可达12位。

(3) 感光器件

感光器件是扫描图像的拾取设备,相当于人的眼球,其重要性不言而喻,也是我们要进行重点介绍的部分。目前扫描仪所使用的感光器件有三种:光电倍增管、电荷耦合器件(CCD)和接触式图像传感器(Contact Image Sensor, CIS)。

1) 光电倍增管

光电倍增管实际上是一种电子管,感光材料主要是金属铯的氧化物及其他一些活性金属(主要是碱系金属)氧化物的参杂物,用这种材料制成的光电阴极,在光线的照射下能够发射电子,经栅极加速放大后冲击阳极,形成电流。在各种感光器件中,光电倍增管是性能最好的一种,无论是灵敏度、噪声系数还是动态范围都遥遥领先于其他感光器件,更难能可贵的是它的输出信号在相当大范围内保持着高度的线性输出,使输出信号几乎不用做任何修正就可以获得准确的色彩还原。同时,光电倍增管的温度系数极低,可以忽略不计,因此它几乎不受周围环境温度的影响。不过光电倍增管在各种感光器件中是生产成本最高的,而且由于一次只能扫描一个像素,因此扫描速度很慢,扫描一张图需要几十分钟,所以现在它一般只用在昂贵的专业滚筒式扫描仪上。

2) 电荷耦合器件(CCD)

CCD 与我们日常使用的半导体集成电路相似,在一片硅单晶上集成了几千到几万个光电三极管,这些光电三极管分为三列,分别用红绿蓝色的滤色镜罩住,从而实现彩色扫描。光电三极管在受到光线照射时可以产生电流,经放大后输出。采用 CCD 的扫描仪技术经过多年的发展已经比较成熟,是市场上主流扫描仪采用的感光元件。CCD 的优势主要在于:最近几年其成像质量和性能大幅提高,其高端产品的性能已经接近低档的光电倍增管产品;在物体表面进行成像,具有一定的景深,能够扫描凹凸不平的物体;温度系数比较低,对于一般的工作,周围环境温度的变化可以忽略不计。CCD 的缺陷主要有:要求有一套精密的光学系统配合,这使得扫描仪结构复杂、成本昂贵;由于数千个光电三极管的距离很近(微米级),在各光电三极管之间存在着明显的漏光现象;各感光单元的信号产生干扰,降低了扫描仪的实际清晰度;由于采用了反射镜、透镜,会产生图像色彩偏差和像差,需要通过软件进行校正,抗震能力较差;扫描仪体积不可能做得很小。

3) 接触式图像传感器(CIS)

现在几乎每家扫描仪生产厂家都推出了几种使用 CIS 作为感光器件的扫描仪。其实,这种技术与 CCD 技术几乎是同时出现的,它使用的感光材料一般是硫化镉,但由于尺寸太大,无法使用镜头成像,只能依靠贴近目标来识别目标,因此光学分辨率最高只能达到 200 dpi,曾广泛用在低档手持式黑白扫描仪上,但随着扫描仪彩色化、高精度化,CIS 迅速从扫描仪市场上销声匿迹了。1998 年后,CIS 技术有了重大突破,极限分辨率被提高到 600 dpi,再加上其生产成本只有 CCD 的 1/3,所以得到广泛应用。

CIS 采用了一种接触式图像感光器件——光敏传感器来进行感光,在扫描平台下一至两毫米处,一排由 300 或 600 个紧密排列的红、蓝、绿三色发光二极管(LED)所发的光混合在一起产生白色光源,取代了 CCD 扫描仪中的 CCD 阵列、透镜、荧光管或冷阴极射线管等复杂的结构,变 CCD 扫描仪光、机、电一体为 CIS 扫描仪的机、电一体,使扫描仪可以做得像笔记本计算机一样薄,甚至还要薄一些。使用 CIS 技术还有一个优点,其所用的光敏传感器耗电量远远低于 CCD 扫描仪中使用的灯泡,这使得 CIS 扫描仪可以依靠电池供电或使用 USB 接口供电,能提供更好的便携性及易用性;利用 USB 接口传输比 EPP 接口更快,能提高扫描速度。

但 CIS 技术也有一些不足之处。由于 LED 阵列是由数百个发光二极管组成,一旦有一个损坏就意味着整个阵列的报废,因此这种产品的寿命比较短;由于 CIS 固有的感光特性,决定了其依然需要一次扫描、三次曝光,所以扫描速度较慢;由于 CIS 不能使用镜头,没有景深特性,原稿必须与感光元件靠得很近,无法进行实物扫描;由于目前 CIS 感光元件的性能决定了 CIS 扫描仪的分辨率只能达到 300 dpi 或 600 dpi,加之 CIS 光源的均匀性不够好,使得 CIS 扫描仪的扫描图像质量和色彩逼真度不是太好,甚至比不上一些低档位的 CCD 扫描仪。

二极管间接曝光(LED Indirect Exposure, LIDE)技术是佳能公司独创的技术,为 CIS 技术的一种。

(4) 扫描速度和扫描幅面

扫描速度有多种表示方法,通常用在指定的分辨率和图像尺寸下的扫描时间表示。扫描幅面表示可扫描图纸的最大尺寸,常见的有 A4、A3、A0 幅面等。

(5) 扫描仪的接口

扫描仪的接口是指与计算机主机的连接方式,通常分为 SCSI、并行接口(即 PC 的打印机接口,一般工作在 EPP 模式)、USB 和 IEEE-1394 等 4 种。

早期的扫描仪都使用 SCSI 卡作为接口,SCSI 接口速度快、连接设备多而且系统资源占用率低。但是许多扫描仪厂商为了降低成本,都自己开发精简过的扫描仪专用 SCSI 卡,这样的 SCSI 接口与 EPP、USB 相比在传输速率上几乎没有优势,而且因为要拆开机箱进行安装,也显得比较麻烦。当然也有不少厂商使用标准的 SCSI 卡连接扫描仪,在扫描速度上会快很多,感觉非常明显。

EPP 并行端口扫描仪使用普通并行电缆即可与计算机相连接。一般这样的扫描仪上还会有一个适配器用于连接打印机,但同时只能有一个设备占用并行端口。如果同时进行打印和扫描,速度会慢得不堪忍受。EPP 并行端口扫描仪的优势在于安装简便、价格相对低廉,而且不需要设置中断、地址等,不会与其他硬件发生冲突,弱点是其传输速率比 SCSI 接口稍慢,但对于个人用户来说够用了。

目前,各种外设的接口都趋于使用 USB 接口,它的优点几乎和 EPP 并行端口一样,只是速度更快、使用更方便,对于一般个人用户,推荐使用 USB 接口的扫描仪。

IEEE-1394 接口由于可以支持比 USB 更高的传输速率,而且它不需要控制器,可以实现对等传输,同样支持即插即用。现在的大幅面彩色扫描仪一般都采用这种接口。

8.3 触摸屏

触摸屏很多人都用过,在一些公共场合的计算机查询系统上,只需用手指轻按屏幕上的相关条目,即可查到所需的内容。其实,你动手按的已不是普通的电脑屏幕,而是装在屏幕前面的触摸屏。它的功能类似于鼠标,主要起光标定位作用,通过点按触摸屏,不懂计算机的人也能很好地操作计算机了。触摸屏是最适合信息查询的输入设备。

触摸屏是一套透明的绝对定位系统。首先它必须保证是透明的,因此它必须通过材料科技来解决透明问题;其次,它采用绝对坐标进行定位,不像鼠标,后者采用相对坐标定位。

触摸检测装置一般安装在显示器的前端,主要作用是检测用户的触摸位置,并传送给触摸屏控制卡。触摸屏的基本工作原理是通过手指或其他物体触摸安装在显示器前端的触摸屏时,所触摸的位置(即绝对坐标)经触摸屏控制器检测后通过接口送到 CPU,从而确定输入的信息。触摸屏系统一般包括触摸屏控制器和触摸检测装置两个部分。触摸屏控制器的主要作用是从触摸点检测装置上接收触摸信息,并将它转换成触点坐标,再送给 CPU,它同时能接收 CPU 发来的命令并加以执行。

由于触摸屏本身的特点,对触摸屏的要求除了要求非常透明、精确定位之外,还要求它长时间保持准确、工作稳定可靠、不影响美观和不容易损坏。根据触摸屏所用的介质以及工作原理,可分为表面声波式、电阻式、电容式、红外线式等,其特性如表 8-2 所示。

2. 电阻触摸屏

电阻触摸屏的品牌很多,但它们的工作原理都一样,所不同的是制作工艺及基材不同。电阻触摸屏的主要部分是一块与显示器表面相匹配的电阻薄膜,这是一种多层的复合薄膜,由一层玻璃或有机玻璃作为基层,表面涂有一层透明的氧化锡(Indium Tin Oxide,简称ITO,是电阻触摸屏及电容触摸屏都用到的主要材料)导电层,上面再覆盖一层外表面经过硬化处理、光滑且防刮擦的塑料层,其内表面也涂有一层导电层,在两层导电层之间有许多细小(小于千分之一英寸)的透明隔离点把它们绝缘。

电阻触摸屏的两层导电层工作面必须是完整的。在每个工作面的两条边线上各涂一条银胶,一端加5V电压,一端接地;一个工作面上的电压加在垂直方向,另一个工作面上的电压加在水平方向。当手指触摸屏幕时,两层导电层在触摸点位置就有了一个接触,经模拟-数字转换器(ADC)测量出接触点的模拟量电压值,根据它和5V的比例公式可计算出触摸点的X、Y坐标值。

(1) 第一代四线电阻触摸屏

四线电阻触摸屏与控制器的连接总共需要四根导线。四线电阻触摸屏的基层大多数是有机玻璃。由于ITO是无机物,有机玻璃是有机物,有机物和无机物是不能良好结合的,时间一长就容易脱落,而且存在透光率低、风化、老化的问题。另外,还存在安装风险,这是因为有机玻璃的刚度差,安装时不能用手捏边上的银胶,以免薄弱的ITO和相对厚实的银胶开裂。不能用力压或拉扯膜屏,以免压断或拉断ITO层。有些四线电阻触摸屏安装后显得不太平整就是因为这个原因。

由于四线电阻触摸屏靠外的那层塑胶及ITO涂层被经常触动,时间一长就容易脱落,用了一段时间后外层薄薄的ITO涂层就会有细小的裂纹。显然,导电工作面一旦有了裂纹,电流就会随之而过,工作面上的电压场分布也就不可能再均匀,这样,在裂纹附近触摸屏漂移严重,裂纹增多后,触摸屏有些区域可能就不再能触摸不到了。

(2) 第二代五线电阻触摸屏

五线电阻触摸屏的基层使用的是玻璃,内层ITO涂层直接与玻璃基层结合为一体,形成导电玻璃。导电玻璃自然没有有机玻璃作基层的种种弊端,同时在外层ITO涂层上有外层镀金导电层。镀金涂层延展性极好,可保护ITO涂层,只要它不断成两半,就仍能继续完成作为导体的使命。有触摸后靠检测内层ITO接触点电压的方法测得触摸点的位置。

五线电阻触摸屏内层ITO需4根引线,外层需要1根,因此共需要5根导线与控制器连接。五线电阻屏的使用寿命和透光率与四线电阻屏相比有了一个飞跃:五线电阻屏的触摸寿命是3500万次,四线电阻屏则只有不到100万次,且五线电阻触摸屏没有安装风险,同时五线电阻屏的ITO层能做得更薄,因此透光率和清晰度更高,几乎没有色彩失真。

3. 电容触摸屏

电容触摸屏是运用电场原理,利用人体的电流感应进行工作的。电容触摸屏由一块4层复合玻璃屏,玻璃屏的内表面和夹层各涂有一层ITO,最外层是一薄层砂土玻璃保护层,夹层ITO涂层作为工作面,4个角上引出4个电极,内层ITO为屏蔽层以保证良好的工作环境。当手指接触电容屏时,由于人体电场,手指和工作面形成一个耦合电容;对于高频电流来说,电容是直接导体,于是手指从接触点吸走一个很小的电流,这个电流从触摸屏的4个角上的电极中流出,并且流经这4个电极的电流与手指到4个角的距离成正比,控制器通过对这4个电流比例的精确

表 8-2 各种触摸屏的比较

类别	表面声波	电容	红外	五线电阻	四线电阻
特性					
清晰度	很好	字符图像模糊	由外壳材料确定	较好	字符图像模糊
反光性	很少	严重	由外壳材料确定	有	较少
透光率	92% (极限)	85%	由外壳材料确定	75%	55%
色彩失真	无	有	由外壳材料确定	无	有
分辨率(像素)	4 096 × 4 096	1 024 × 1 024	1 000 × 720	4 096 × 4 096	1 024 × 1 024
压力轴影响	有	无	无	无	无
漂移	无	漂移	无	无	无
防刮擦	非常好且不怕硬物	一般,怕硬物刮击	由外壳材料确定	一般,怕锐器	差
野室使用	不怕	一般	由外壳材料确定	好,但怕锐器	差
反应速度(ms)	10	15~24	50~300	10	10~20
材料	纯玻璃	四层复合膜	透光外壳	镀膜玻璃	镀膜有机玻璃
多点触摸	智能判断	中心点	左上角	中心点	中心点
光干扰	没有此问题	没有此问题	不能超范围	没有此问题	没有此问题
电磁干扰	没有此问题	有	没有此问题	没有此问题	没有此问题
防尘	不怕	不怕	不能挡住光路	不怕	不怕
寿命	远大于5 000万次,半永久性	2 000万次	太多传感器,易损坏	3 500万次	100万次
安装风险	不易碎	易碎	由外壳材料确定	不易碎	易损坏ITO涂层
市场返修率	2%	15%	20%	5%	8%
外观	不影响	不影响	影响外观	不影响	不平整
现场维护	不需要	需经常校准	清洁外壳	不需要	不需要

1. 表面声波触摸屏

表面声波是超声波的一种,能在介质(例如玻璃或金属等刚性材料)表面进行浅层传播。表面声波性能稳定、易于分析,并且在横波传递过程中具有非常尖锐的频率特性,近年来在无损探伤、造影和滤波器等领域发展非常成熟。

表面声波触摸屏的触摸部分可以是一块平面、球面或是柱面的玻璃板,表面没有任何贴膜和覆层,所以透光率高。将它安装在CRT、LED大屏、LCD或是等离子显示器屏幕的前面。在屏幕上X轴和Y轴方向的发射器及接收器,可实现电信号与声波之间的转换。当人体接触屏幕的时候,通过该点的声波信号被阻挡,其波形形成一个衰减,控制器通过分析接收到的衰减信号而确认被触摸的位置。

确计算,得出触摸点的位置。

4. 红外触摸屏

红外触摸屏是利用 X、Y 方向上密布的红外线矩阵来检测并定位用户的触摸位置。红外触摸屏显示器的前面安装一个电路板外框,电路板的四周布满了红外发射管和红外接收管,一一对应形成横竖交叉的红外线矩阵。用户在触摸屏幕时,手指就会挡住经过该位置的横竖两条红外线,因而可以判断出触摸点的位置。任何物体都可阻挡触点上的红外线而实现触摸屏操作。

红外触摸屏不受电流、电压和静电干扰,适宜恶劣的环境条件,红外线技术是触摸屏产品最终的发展趋势。采用声学和其他材料科学技术的触摸屏都有其难以逾越的屏障,如单一传感器的受损、老化,触摸屏面怕受污染、破坏性使用,维护繁杂等问题。红外线触摸屏只要真正实现了高性能和高分辨率,必将替代其他技术产品而成为触摸屏市场主流。

过去的红外触摸屏的分辨率由框架中的红外对管数目决定,因此分辨率较低,另外红外触摸屏对光照环境因素比较敏感,在光照变化较大时会误判甚至死机。而最新的技术第五代红外屏的分辨率取决于红外对管数目、扫描频率以及差值算法,分辨率已经达到了 $1\,000 \times 720$ 像素。至于说红外触摸屏在光照条件下不稳定,从第二代红外触摸屏开始,就已经较好的克服了抗光干扰这个弱点,但红外触摸屏毕竟是通过红外光线工作,只能承受有限的光干扰,因此在使用环境上有一定的限制。

第五代红外线触摸屏是全新一代的智能技术产品,它实现了 $1\,000 \times 720$ 像素高分辨率,多层次自调节和自恢复的硬件适应能力和高度智能化的识别,可长时间在各种恶劣环境下任意使用;并且可针对用户定制扩充功能,如网络控制、声感应、人体接近感应、用户软件加密保护、红外数据传输等。红外触摸屏可以选用任何客户认为满意的防爆玻璃而不会增加太多的成本和影响使用性能,这是其他的触摸屏所无法效仿的。

大多数红外触摸屏的控制器直接设计在框架中的电路板上,也有红外触摸屏把控制器设计在单独的小盒中。控制器通过键盘接口或者串行接口直接与主机通信,键盘接口的红外触摸屏用户甚至可以直接读取键盘口发来的触摸屏数据而无需任何驱动程序。

8.4 显示系统

在使用计算机时,人们通过显示器可以看到计算机的输出内容,显示器是计算机最重要的输出设备。计算机的显示系统实际上由显示适配器和显示器两部分组成,如图 8-10 所示。显示适配器通常以插卡的方式插在计算机的总线插槽上或显示专用接口上,也有的计算机主板上集成有显示适配器。显示适配器与显示器通过显示专用接口连接。本节着重以现在流行的显示系统标准介绍显示适配器、显示器和显示接口。

8.4.1 CRT 显示器

CRT(Cathode Ray Tube, 阴极射线管)显示器是计算机使用最广泛的显示器,随着计算机技术的发展和应用的拓展,CRT 显示器也在快速发展,其种类和档次也越来越多。

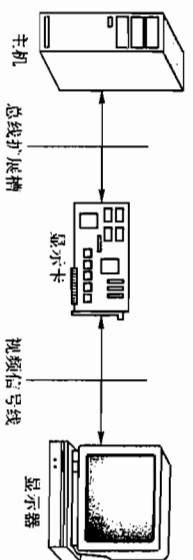


图 8-10 显示系统示意图

从 20 世纪 80 年代初到今天,显示器所用尺寸通常在 12 英寸至 21 英寸之间。其分辨率已从 320×200 像素发展到 $1\,024 \times 768$ 像素,有的达到 $1\,280 \times 1\,024$ 像素和 $1\,600 \times 1\,200$ 像素以上。颜色也由单色发展到彩色。显像管的荫罩点距从 0.6 mm 以上发展到 0.21 mm 以下;行扫描频率从 15.8 kHz 发展到 120 kHz 以上;输入信号从数字信号转换为模拟信号。显示屏尺寸从 12 英寸发展到 20 英寸以上,显示屏也越来越平面化。目前的显示器已朝着高分辨率、高亮度、平面化、大屏幕、低辐射和方便用户方向发展。在电路方面,目前国外的一些厂商已推出具有记忆和自动跟踪功能的智能化显示器。

可以依据显像管所能显示的颜色,将 CRT 显示器分为单色显示器和彩色显示器。彩色既可以显示彩色,也可以显示单色。但是,在实际使用中,人们习惯于以显示器连接主机所采用的显示适配器来划分,如 MDA 单色显示器、CGA 彩色显示器、EGA 彩色显示器、VGA 彩色显示器及 TVGA 彩色显示器等。

CRT 显像管是一种玻璃壳封装的真空电子器件,它的功能是把电信号转换成可见的光信号,而且在很大程度上决定着显示器的清晰度、对比度、色彩和亮度等,所以是显示器的重要组成部分。

1. 单色显像管

单色显像管只能显示一种颜色。它的基本结构是由电子枪、偏转系统和荧光屏三部分组成,如图 8-11 所示。

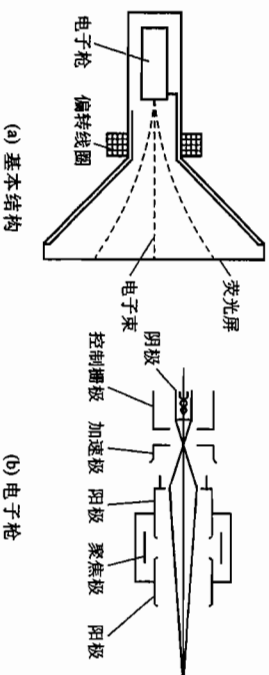


图 8-11 单色显像管原理

(1) 电子枪

电子枪发射出一束很细的强度可控的高能电子束。电子枪由灯丝、阴极、控制栅极、加速极、聚焦极和高压阳极组成。

灯丝装在阴极套管里面,用于加热阴极。阴极为一小圆筒,筒的顶端涂有发射材料,当阴极被灯丝加热后,即可发射电子。控制栅极也是一个圆筒,套在阴极外面,为便于电子流过,圆筒中间开一小孔,它的作用是控制阴极发射的电子束到达荧光屏的数量。由于电子带负电,所以控制栅极相对于阴极的电位越低,阻碍电子的能力越大,使得到达荧光屏上的电子数量就越少,荧光屏上的发光物质被激活的少,亮度就暗;反之,栅极相对于阴极的电位越高,则到达荧光屏的电子数量就越多,其对应的光点就越亮。由此可见,改变栅极相对于阴极的电位,就可以直接控制电子束的强弱,从而控制荧光屏上光点的亮度。加速极位于控制栅极的前面,为一圆盘,中间也有一小孔,其作用是控制阴极发射的电子束到达荧光屏的速度。通常在加速极上加有几百伏的正电压。如果不加正电压,阴极发射的电子束就不能通过控制栅极的小孔,荧光屏也就不能发光。聚焦极是一个直径较大的金属圆筒,一般加有几百伏的可调直流电压,用于调节阴极发射的电子束的聚焦程度。高压阳极也为金属制成的圆筒,通过弹性金属片与显示器的玻璃外壳的内壁石墨导电层相连,其上加上有 $9\text{ kV} \sim 16\text{ kV}$ 电压的强电场,以利于加速电子束的运动和提高荧光屏的亮度。由于阳极电压极高,因此高压不从显像管脚引出,而是在锥体玻璃外壳的侧面单独引出(又称高压嘴),与内壁石墨导电层和高压阳极连通,从而解决了高压的绝缘问题。而其他各极则由玻璃壳管颈的尾部管脚引出。

(2) 偏转系统

由电子枪产生并经聚焦的电子束只能在荧光屏上形成一个亮点。而使用偏转系统可以控制电子束在荧光屏上下左右移动,以构成不同的字符、图形或图像。

偏转系统有静电偏转和磁偏转两种。为了实现大角度的偏转,CRT显示器一般采用磁偏转系统。磁偏转系统通常在CRT显像管的管颈外面,安装相互垂直的水平(行)偏转线圈和垂直(场)偏转线圈。在偏转线圈中通以电流,形成磁场,电子束在磁场中运动时,由于受到磁场的作用,会产生偏转。假设人们面对荧光屏,水平偏转线圈产生的垂直磁场方向自上向下,若电子束的运动方向由屏里向外,即电流方向由屏外向里,根据左手定则,电子束从右向左偏转;如垂直偏转线圈产生的水平磁场方向自左向右,电子束运动方向不变,则电子束从上向下偏转。由此可见,只要改变磁场方向,就可以改变电子束的偏转方向,而改变偏转线圈中的电流方向就可使之产生的磁场方向改变。那么,偏转角度的大小与什么有关呢?很显然,偏转角度的大小与磁场产生的偏转力的大小有关。当偏转线圈的匝数一定时,流经偏转线圈的电流越大,磁场就越强,偏转力就越大,电子束偏转角度也就越大。因此,只要设计一种偏转电路,通过偏转电路改变偏转线圈中的电流方向与数值,就可以改变电子束的偏转方向和偏转角度,从而实现在屏幕上显示字符和图像。

在偏转系统的作用下,电子束在屏幕上具有规律的运动叫做扫描。计算机的CRT显示器一般采用光栅扫描,即把显示屏的左上角作为起点,通过偏转电路使电子束照射到左上角,然后让电子束沿水平方向,从左到右匀速地运动,电子束所经过的亮点连成一条直线,称为扫描一行,此为扫描一行。一行扫描完后,电子束光点再以极高的速度回到左上角的下一行的开始位置,回程时光点不亮,此为回扫过程。这一正扫和回扫过程构成一个水平扫描周期。之后,继续上述水平扫描周期,自上而下一行又一行,直到屏幕最末一行的右下角为止。一幅扫描完毕后,迅速返回到左

上角起点,同样回扫光点不亮,开始下一帧的扫描。这一帧的正扫与回扫构成了一个垂直扫描周期。在不断地重复水平和垂直扫描过程中,除各回扫期间,通过水平和垂直消隐信号抑制电子束发射不形成亮点外,在整个正扫期间,可使显示形成一条水平扫描线(又称光栅),这种有规律的扫描过程叫做光栅扫描,如图8-12所示。

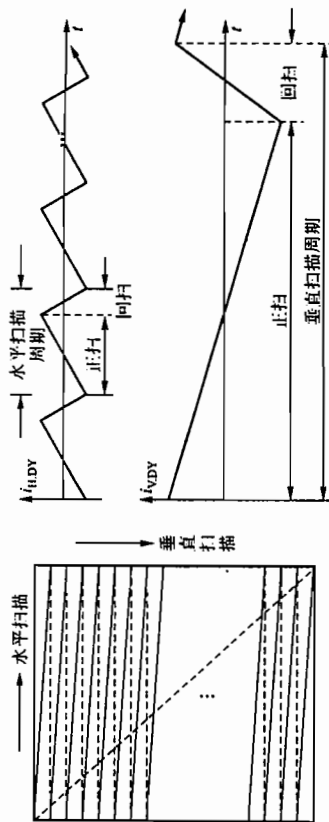


图8-12 光栅扫描与扫描电流

光栅扫描又分为逐行扫描和隔行扫描。逐行扫描是一行接一行地逐次扫描,一次扫描完成一幅画面。计算机显示器一般采用逐行扫描。隔行扫描是把一幅画面分为两次扫描完成,一次先扫描奇数行扫描线,形成奇数场;另一次再扫描偶数行扫描线,形成偶数场。一个奇数场和一个偶数场构成一幅,电视机一般都采用隔行扫描。有的计算机CRT显示器也采用隔行扫描。隔行扫描是为了增加帧频,以克服闪烁现象。如果采用逐行扫描,又不产生闪烁现象,则必须增加帧频,这样就会提高偏转系统的造价。

为了实现光栅扫描,必须使水平(行)偏转线圈和垂直(场)偏转线圈中通以线性变化的锯齿波电流,使所形成的磁场能分别控制电子束在水平和垂直方向的运动。

(3) 荧光屏

荧光屏是实现电信号转换为图像信号的器件,是用沉积法将荧光粉涂敷在玻璃屏幕上制成的。涂敷不同的荧光粉,具有不同的荧光特性。荧光粉的主要特性如下。

1) 发光颜色

不同的荧光粉在电子束的轰击下可以发出不同颜色的光,常用的有白、绿、黄、琥珀色等。由于人的肉眼对绿色的灵敏度较高,且不易疲劳,所以在单色显示器中多采用绿色荧光屏。

2) 亮度

荧光粉发光的总量与其接收到的电子束的能量成正比。因此,荧光屏某点发光的亮度与电子束运动速度、电流密度和电子束在该点停留的时间有关。电子束的运动速度越快,电流密度越大,电子束停留时间越长,则该点的发光亮度就越高。

3) 余辉时间

荧光屏在电子束停止轰击后,其发光有一个逐渐消失的过程,而不是立即消失。余辉时间就是指电子束停止轰击后,发光亮度下降到初始值的1%所经过的时间。余辉时间小于1 ms的称为短

余辉, $1\text{ ms} \sim 100\text{ ms}$ 的为中余辉, 大于 100 ms 的则为长余辉。CRT 一般采用中余辉的荧光粉。

4) 闪烁效应

CRT 显示器在显示字符或图像时, 电子束必须不断地在荧光屏上重复扫描。若重复频率过低, 尽管人的肉眼对字符或图像有视觉暂留现象, 还是会感觉显示的字符或图像有一亮一暗的现象, 这种情况叫闪烁效应。闪烁效应并不影响 CRT 显示的信息量, 但容易导致人眼的疲劳, 从而降低工作质量和效率。

字符或图像开始出现闪烁效应的重复频率, 称为临界闪烁频率。因此, 为了克服闪烁所带来的负面影响, 在 CRT 显示器中, 字符或图像的刷新频率必须大于临界闪烁频率。

2. 彩色显像管

彩色显像管可以显示彩色字符和图像。目前采用较多的彩色显像管都是基于三基色原理进行显示的。所谓三基色原理, 就是选择三种相互独立的单色光(多采用红(R)、绿(G)、蓝(B), 也有用青、紫、黄等), 把它们按不同的比例组合, 即可得到各种颜色。显然, 利用三基色的对称组合, 并通过不同亮度(I)的组合, 就可得到 16 种颜色。

彩色显像管有多种类型, 如荫罩式、栅网式、穿透式等。本书只介绍一种应用最广、发展最早、工艺最成熟的三枪三束荫罩式彩色显像管, 如图 8-13 所示。它有三个按正三角形排列的电子枪, 分别用来激发红、绿、蓝三种色的荧光粉发光。荧光屏上每一个像素也由按正三角形排列的红、绿、蓝三种小荧光粉组成, 并与电子枪相对应。为了保证三个电子枪发射出来的电子束准确地击中各自相对的荧光粉小点, 在距离荧光屏内侧约 1 毫米左右的地方设置了一块布满小孔的荫罩板(金属板), 其孔数与三色荧光粉点的组数相一致, 一般有几十万个。在设计安装时, 使红、绿、蓝三个电子枪发射的电子束穿过同一小孔正好打在各自对应的荧光粉上。

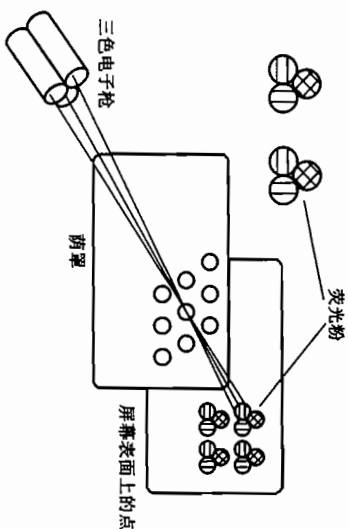


图 8-13 彩色显像管结构

通常, 把荧光屏上每一个荧光粉点组之间的距离定义为显像管屏幕的点距。显然, 这也是荫罩板小孔间的距离, 所以又叫孔距。显像管荧光屏的点距决定了屏幕的最高分辨率, 点距越小, 分辨率越高。目前, 在已有的 CRT 显示器中, 显像管荧光屏点距有 0.39 mm 、 0.33 mm 、 0.28 mm 、 0.24 mm 、 0.21 mm , 甚至更小。这些数值实际上反映了显示器的最高分辨率。当显示

器处于某种显示方式时, 一个像素点往往覆盖不止一个荧光粉点组, 即实际使用的分辨率不一定是显像管的最高分辨率, 它决定于显示方式、点时钟、视频带宽等因素的影响。例如, 14 英寸彩色显示器的水平有效显示尺寸为 250 mm 左右, 对于 0.31 mm 点距的显像管, 大约可显示 800 个点。在 320×200 像素显示模式下, 实际使用的分辨率要低于显示器能够达到的最高分辨率。另外, 有些 14 英寸显示器的说明中, 能够提供 1024×768 像素的显示模式, 只是说支持这种显示模式, 但并不等于在显示器的水平方向上真正能够显示 1024 个像素。若要显示器清晰地显示 1024×768 像素, 则必须使用 0.24 mm 荫罩点距的显像管。

彩色 CRT 显示器的基本组成如图 8-14 所示, 主要由视频放大驱动电路、行扫描电路、场扫描电路、高压电路、CRT 显像管和机内直流电源等六大部分组成。

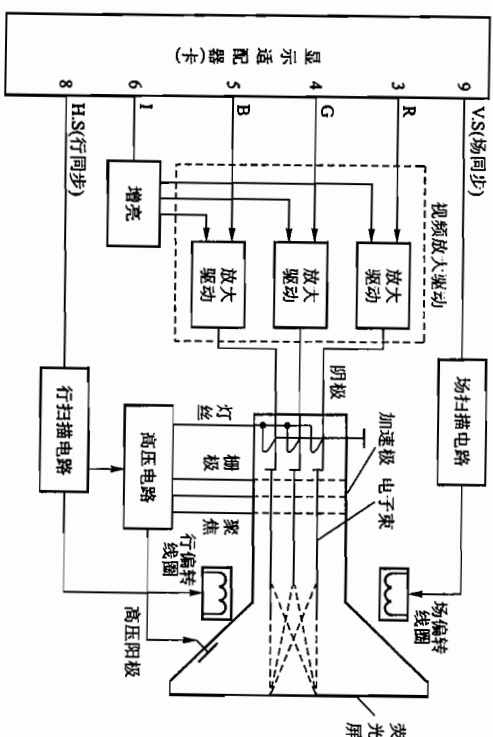


图 8-14 彩色 CRT 显示器的基本组成

视频放大驱动电路将主机经显示适配器发送来的视频信号放大驱动后, 送至显像管的阴极, 产生电子束轰击荧光屏而出现光点。由显示适配器发送来的水平同步信号和垂直同步信号分别送入扫描电路和场扫描电路, 为 CRT 的水平(行)偏转线圈和垂直(场)偏转线圈提供有一定幅度和线性良好的锯齿波电流, 产生垂直方向和水平方向的偏转磁场, 控制电子束作水平和垂直方向偏转, 以形成扫描光栅。

CRT 显示器的显像管是用来实现字符和图像显示的电-光转换器件, 要其正常工作, 必须按要求各极提供所需高、中、低电压。高压电路用来产生这些电压, 它利用行扫描电路产生行回程脉冲, 经过行回程变压器(又称高压包或行输出变压器)升压、整流、滤波后, 为显像管提供正常工作所需的各档电压, 并为视频等电路提供电源电压, 使显像管正常显示光栅或字符、图像。显示器内的直流电源一般均采用开关电源提供直流电压, 通过电路对交流电网电压进行变换, 输出几档直流电压, 为显示器内部各电路提供所需直流电压。一般来讲, TTL 数字集成电路需要 $+5\text{ V}$ 的直

流电压,视频信号处理及行、场振荡需要 +12 V 的直流电压,场输出需要 +25 V 左右的直流电压,视频放大驱动需要 +80 V ~ +150 V 不等的直流电压,行输出需要几十伏至一百多伏不等的直流电压。不同机型由于采用不同的电路,每组的电压值和电压的组数会有所差别。

3. CRT 显示器的主要技术指标

(1) 屏幕尺寸

屏幕尺寸是指屏幕对角线长度,如图 8-15 所示。一般有 14、15、17、19、20 和 21 英寸等。

(2) 点距

点距是指 CRT 上两个颜色相同的荧光点之间的距离,单位是毫米,如图 8-16 所示。



图 8-15 屏幕尺寸

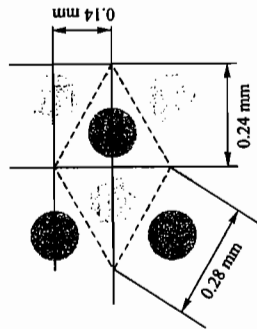


图 8-16 CRT 显像管的点距

(3) 像素和分辨率

像素是指组成图像的最小单位,也即上面提到的发光“点”。分辨率指屏幕上在水平方向和垂直方向的像素的数目。比如,640 × 480 的分辨率是说在水平方向上有 640 个像素,在垂直方向上有 480 个像素。为了控制像素的亮度和颜色深度,每个像素需要很多个二进制位来表示。例如,如果要显示 256 种颜色,则每个像素至少需要 8 位(一个字节)来表示。

(4) 带宽

带宽是指每秒钟电子枪扫描过的像素的个数,以 MHz 为单位,表明了显示器电路可以处理的频率范围。比如,在标准 VGA 方式下,如果刷新频率为 60 Hz,则需要的带宽为 $640 \times 480 \times 60 = 18.4$ MHz;在 1 024 × 768 像素的分辨率下,若刷新频率为 70 Hz,则需要的带宽为 55.1 MHz。以上的数据是理论值,实际所需的带宽要高一些。早期的显示器是固定频率的,现在的多频显示器采用自动跟踪技术,使显示器的扫描频率自动与显示适配器的输出同步,从而实现了较宽的适用范围。带宽的值越大,显示器性能越好。

(5) 安全标准

现在,人们越来越关心自己的健康和环保问题,从而开始把目光投向了显示器的辐射问题。目前,显示器的主要标准有:FCC 即美国联邦通讯委员会电磁兼容标准;DHHS 即美国联邦政府设立用来规范家用电器设备的电磁辐射标准;CSA 即加拿大标准协会的安全标准;ISO9247-3 即欧洲通用标准;MPR II 即瑞典国家测量局制定的标准,现已成为世界标准(主要针对有关静

电、电子及磁场制定的测试标准):TCO 是由瑞典专家联盟推出的对低辐射和环保最严格的标准,同期推出的还有 NUTEK 节电标准,这两项标准对显示器的要求非常高。低辐射显示器,其说明书上一般有“Low Radiation”字样,显示器的外壳上也有该字样。

(6) 节能标准

“能源之星”是符合美国环保局(Environmental Protection Agency, EPA)环保标准的认证标志,标有能源之星标志的电子设备符合 EPA 环保节能标准。符合“能源之星”标准的绿色环保显示器,这可以从显示器所附的说明书上看出,机壳上也大都标有“能源之星”的标志。

8.4.2 液晶显示器

液晶显示器(Liquid Crystal Display, LCD)是一种数字显示技术,通过液晶和彩色过滤器过滤光源,在平面板上产生图像。

在笔记本计算机市场占据多年的领先地位之后,基于液晶显示技术的显示屏正逐步地进入台式计算机市场。LCD 拥有许多传统的 CRT 显示技术所不具备的优势,能够提供更加清晰的文本显示,而且屏幕无闪烁,从而能够有效降低长时间注视屏幕所产生的视觉疲劳。尽管 LCD 显示器有其诱人的独到之处,但不可否认,与主要的竞争对手 CRT 显示器相比,LCD 显示器在高质量的色彩显示方面仍存在不足。

与传统的 CRT 显示器一样,应用于台式计算机的 LCD 显示器大多也被设计成接收模拟信号,而非直接接收由计算机产生的数字脉冲信号。这主要是因为,目前台式计算机中的绝大多数标准适配器仍然是在将视频信息由最初的数字信号转换为模拟信号之后,再传送给显示器进行显示。然而 LCD 显示器本身只能处理数字信息,因此当从显示适配器接收到模拟信号之后,LCD 显示器需要将模拟信号再还原为数字信号后进行处理。为了解决上述问题带来的显示上的不足,最新的台式计算机 LCD 显示器采用了一种特殊的带有数字连接器的图形卡直接向 LCD 显示器传送数字信号。

1. LCD 显示器的工作原理

早在 1888 年,人们就发现液晶这一呈液体状的化学物质,像磁场中的金属一样,当受到外界电场影响时,其分子会产生精确的有序排列。如果对分子的排列加以适当的控制,液晶分子将会允许光线穿越。

LCD 显示屏都是由不同部分组成的分层结构。位于最后面的一层是由荧光物质组成的可以发射光线的背光源(工作时约有 2~4 根灯管始终发光)。背光源发出的光线在穿过第一层偏振过滤层之后进入包含成千上万水晶液滴的液晶层。液晶层包含了两片相当精密的无钠玻璃基板,称为 Substrates,中间夹着一层液晶。当光束通过这层液晶时,液晶本身会有序排列或扭转呈不规则状,因而阻隔或使光束顺利通过。液晶层中的水晶液滴都被包含在细小的单元格结构中,一个或多个单元格构成屏幕上的一个像素。当 LCD 中的电极产生电场时,液晶分子就会产生扭曲,从而将穿越其中的光线进行有规则的折射,然后经过第二层过滤层的过滤在屏幕上显示出图像。LCD 显示器是通过直接控制每个液晶的偏转角度来形成图像,而不是像 CRT 那样靠电压引起电子流的偏移来形成图像,故完全没有闪烁和完全没有电磁辐射。

对于单色 LCD 显示器,上述结构已经足够了。但是对于彩色 LCD 显示器来说,还需要有专

门处理彩色显示的彩色过滤层。通常,在彩色 LCD 面板中,每一个像素都是由三个液晶单元格构成,其中每一个单元格前面都分别有红色、绿色或蓝色的过滤器。

2. LCD 显示器的分类

LCD 的驱动方式可分为:静态(Static)驱动、单纯矩阵(Simple Matrix)驱动和主动矩阵(Active Matrix)驱动 3 种。其中,单纯矩阵又可称为被动矩阵,主要有扭转向列型(Twisted Nematic, TN)和超扭转向列型(Super Twisted Nematic, STN),而主动矩阵型大致可分为薄膜晶体管型(Thin Film Transistor, TFT)及二端子二极管型(Metal Insulator Metal, MIM)两种方式。

从目前液晶显示技术所应用的范围和层次来看,主动式矩阵驱动技术以 TFT 为主,多应用于笔记本电脑及动画、影像处理产品;而单纯矩阵驱动技术则以 TN 和 STN 为主,目前的应用多以文字处理器和消费性产品为主。其中,生产 TFT LCD 显示器所需的资金投入和技术条件较高,而 TN LCD 显示器和 STN LCD 显示器相对较低。

TN、STN 及 TFT LCD 显示器因其利用液晶分子扭转的原理不同,在视角、色彩、对比及动态显示品质上有高低层次的差别。大多数通用无源阵列显示器采用 STN 设计,因此称为 STN LCD 显示器。有源阵列面板采用 TFT 设计,因此称为 TFT LCD 显示器。

1) TN LCD 显示器

TN LCD 显示器因技术层次较低,价格低廉,多应用于 3 英寸以下的小尺寸产品,而且仅能呈现出黑白单色及做一些简单文字、数字的显示,主要应用于电子表、计算器、简单的掌上游戏机等消费性电子产品。

2) STN LCD 显示器

STN LCD 显示器较 TFT LCD 显示器工艺简单,成品率较高,价格相对便宜,适于对比强烈与画面转换反应时间较快的产品,因此多应用于信息处理设备。如果在液晶面板前加一片彩色滤光片,则可显示多种色彩,甚至可达全彩化程度。此种产品多用于文字、数字及绘图功能的显示,例如低档的笔记本电脑、掌上计算机、股票机和个人数字助理(PDA)等便携式产品。

3) TFT LCD 显示器

TFT LCD 显示器因为响应速度更快,适用于动画及图像显示,故广泛应用于数码相机、液晶投影仪、笔记本电脑、台式计算机显示器。由于其在色彩品质及响应速度方面优于 STN 型产品,因此成为目前市场上的主流产品。

TFT 液晶显示器主要由荧光管、导光板、偏光板、滤光板、玻璃基板、配向膜、液晶材料、薄膜晶体管等构成。首先,液晶显示器必须利用背光源,也就是荧光灯管投射出的光线,这些光线会经过一个偏光板然后再经过液晶,这时液晶分子的排列方式改变了穿透液晶的光线的角度。这些光线接下来还必须经过前方的彩色滤光膜与另一块偏光板。因此,只要改变刺激液晶的电压值就可以控制最后出现的光线强度与色彩,并进而能在液晶面板上变化出不同深浅的颜色组合了。

3. LCD 的主要性能指标

(1) 屏幕尺寸

LCD 显示器与 CRT 显示器的尺寸的标示方法不一样。CRT 显示器的标示尺寸指的是 CRT 显像管外壳的对角线长度;而 LCD 显示器则以可视范围的对角线长度作为标示尺寸。例如,

CRT 显示器在规格中标示为 17 英寸,但实际上可视尺寸却绝对达不到 17 英寸,大约只有 15 英寸多些;而对于 LCD 显示器,若标示为 15.1 英寸,那么可视尺寸就是 15.1 英寸。

(2) 分辨率

分辨率是显示器的主要衡量指标,因为显示器必须支持软硬件所需要的分辨率。传统 CRT 显示器支持的分辨率较有弹性,可以自己选择高或低的分辨率。这是因为 CRT 显示器的影像主要是由像素所组成的点和线而产生的,因此像素的多寡是影响分辨率的重要因素。

LCD 显示器的分辨率与 CRT 显示器不同。由于受 LCD 液晶层中实际单元格数量的影响,一般不能任意调整,它是制造商所设置和规定的。LCD 显示器一般只能提供固定的显示分辨率,也就是所谓的“真实分辨率”。LCD 显示器只有在真实分辨率下才能表现最佳影像效果。如果用户需要将 800 × 600 像素的分辨率提升到 1024 × 768 像素的话,只能借助于特定软件的帮助实现模拟分辨率。当 LCD 显示器工作在分辨率较低的显示模式时,有两种方式显现。第一种为居中显示,例如,若想在 XGA 1024 × 768 像素的屏幕显示 SVGA 800 × 600 像素的分辨率,只有水平方向居中的 800 个像素,垂直方向居中的 600 条网线可以被呈现出来,整个画面看起来好像是影像被缩小,外围还有阴影环绕。另一种为扩展显示,此种显示方法的好处是,不论用户使用分辨率是多少,所显示的影像一定会运用到屏幕上的每一个像素,而不至于产生阴影边缘环绕。然而,由于影像是被扩展至屏幕上的每一个像素,因此影像难免会受扭曲,清晰度也会受到影响。

所以,选购 LCD 显示器时一定要确保它能支持用户系统所使用的软硬件支持的原始分辨率,千万不要盲目相信分辨率要越高越好。现在,主流产品的分辨率主要有:VGA(640 × 480 像素)、SVGA(800 × 600 像素)、XGA(1024 × 768 像素)、QVGA(320 × 240 像素)、SXGA(1280 × 1024 像素)、SXGA+ (1400 × 1050 像素)和 UXGA(1600 × 1200 像素)等。现在 15.1 英寸 TFT LCD 显示器基本都能够支持到 1024 × 768 像素;17 英寸以上的 TFT LCD 显示器可以达到 1280 × 1024 像素。

(3) 显示色彩

早期的彩色 LCD 显示器在颜色表现方面,最多只能显示高彩(16 位)。因此许多厂商使用所谓的帧速率控制(Frame Rate Control, FRC)技术,以仿真的方式来表现出全彩的画面。目前, LCD 显示器最起码能够显示出高彩,支持全彩(32 位)也是轻而易举的事。

(4) 响应时间

响应时间反映了液晶显示器各像素点对输入信号的反应速度,即像素由暗转亮或由亮转暗的速度。响应时间越短,在观看运动画面时就越不容易出现拖影的现象。一般会称响应时间分为两个部分:上升时间(Rising)和下降时间(Falling),而表示时以两者之和为准。

结构不同的 LCD 显示器,其响应输入信号的速率差异很大。对于速率的要求取决于用途。用于文本显示,响应时间 250 ms ~ 500 ms 就可以满足要求;而如果用鼠标器驱动光标显示,就需要 175 ms 以内的响应时间才能防止拖影;用于运动图像,要求低于 125 ms。实时视频显示,则必须低于 50 ms。只有有源点阵 LCD 具有视频响应能力,它的响应时间为 30 ms ~ 50 ms。目前,无源点阵 LCD 远达不到此要求。

(5) 视角

视角是指人们观察显示器的范围。它用垂直于显示器平面的法向平面角度来度量。这与

LCD 是 STN 还是 TFT 有很大关系。因为前者是靠屏幕两边的晶体管扫描屏幕发光,后者是靠自身每个像素后面的晶体管发光,其对比度和亮度的差别,决定了它们的视角有较大区别。STN LCD 一般只有 60° ,而 TFT LCD 则有 80° 。

单就当前市面上出售的 LCD 显示器来说,可视角度都是左右对称的(即由左边或右边可以看见屏幕上图像的角度是一样的)。通常上下可视角度都小于左右可视角度,而且并不一定是对称的。

(6) 亮度和对比度

LCD 的亮度取决于其结构和背景照明的类型。亮度的测量单位通常为英尺朗伯 (foot lambert, fl) 或坎德拉每平方米 (cd/m^2 或 nit, 即烛光每平方米)。其换算关系是:

$$1 \text{ fl} = 3.425 \text{ cd}/\text{m}^2$$

或

$$1 \text{ cd}/\text{m}^2 = 0.292 \text{ fl}$$

对大多数环境来说, $25 \text{ cd}/\text{m}^2$ 的亮度一般可以满足要求。目前大多数的 TFT LCD 显示器的亮度范围介于 $150 \text{ cd}/\text{m}^2 \sim 200 \text{ cd}/\text{m}^2$ 之间,而大部分的 CRT 显示器最高亮度只在 $150 \text{ cd}/\text{m}^2$ 左右,不过也有少数机种高达 $250 \text{ cd}/\text{m}^2$ 。

目前,提高亮度的方法有两种,一种是提高 LCD 面板的光通过率;另一种就是增加背景灯光的亮度,即增加灯管数量。需要注意的是,较亮的产品不见得就是较好的产品,显示器画面过亮常常会令人感觉不适;一方面容易引起视觉疲劳,同时也使纯黑与纯白的对比降低,影响色阶和灰阶的表现。因此,提高显示器亮度的同时,也要提高其对比度,否则就会出现整个显示屏发白的现象。

此外亮度的均匀性也非常重要,但在液晶显示器产品规格说明书里通常不做标注。亮度均匀与否,和背光源与反光镜的数量与配置方式息息相关。品质较佳的显示器,画面亮度均匀,柔和不刺目,无明显的暗区。

对比度通常是指开状态像素与关状态像素亮度的比率,对比度越高,色彩越鲜艳饱和,还会呈现出立体感。高分辨率显示器的对比度的比率范围一般为 $10:1 \sim 100:1$,有的可达到 $300:1$,甚至更高。当对比度达到 $120:1$ 时,就可以很容易地显示生动、丰富的色彩;而对比度高达 $300:1$ 时,则可支持各色阶的颜色。尽管人们的肉眼接受对比度的能力取决于多种因素,但在许多场合下, $7:1$ 或更高一点的对比度就能满足要求。对比度高于 $20:1$ 时,人的肉眼就觉察不到它们的差别所在了。

从目前来看,还没有一套有效且公正的标准来衡量对比度和亮度指标。所以最好的识别方法还是利用自己的双眼来判定,即将 LCD 显示器调到最亮和最暗,看看感觉如何,现在也只能利用这方法来找到比较合适的 LCD 显示器。

亮度和对比度对 LCD 显示器的影响要大于 CRT 显示器。高亮度的 LCD 显示器使用起来感觉比较好,但也要提供足够高的对比度,以确保色彩的真实度和色阶的准确度。

(7) 刷新频率

对于 CRT 显示器来说,刷新频率是相当重要的指标,因为刷新速度越快,画面越不容易闪烁。由于人的肉眼能够察觉 CRT 的扫描频率高低,因此扫描频率至少要在 75 Hz 以上,如果能调整到 85 Hz 以上,就基本上看不出有闪烁的现象了。而刷新频率对于 LCD 显示器的重要性远

低于 CRT 显示器。在 LCD 显示器中,每个像素都持续发光,直到不发光的信号被送到控制器中,像素的亮灭状态只有在画面内容改变时才会有所变化。因此刷新频率的高低并不会使画面闪烁,即使刷新频率很低,画面也不会因不断充电、放电而产生闪烁。

大多数的 LCD 显示器在 60 Hz 刷新频率下就能达到最佳画质。因为现今大部分显示器仍以 CRT 显示器为设计对象,更高的兼容性将有助于 LCD 显示器切入市场,并取代 CRT 显示器。

(8) 防眩光、防反射

防眩光、防反射功能主要是为了减轻用户眼睛的疲劳。由于 LCD 屏幕的物理结构特点,屏幕的前景反光、屏幕的背景光与漏光以及像素自身的对比度和亮度,都将对用户眼睛产生不同程度的反射和眩光,特别是视角改变时,表现更明显。

4. LCD 显示器与 CRT 显示器之比较

与 CRT 显示器相比, LCD 显示器的特点可以归纳为“五长两短”。

1) 超精致影像画质

LCD 显示器的液晶技术,可产生比一般 CRT 显示器更清晰、更精准的影像画质与更真实、更饱和的色彩呈现。

2) 真正的平面显示

LCD 显示技术实现了真正的平面显示。

3) 节省空间

一台普通 17 英寸 CRT 显示器的厚度大约为 43 cm ,而相同尺寸的 LCD 显示器加上支架为 20 cm 左右。有一项研究结果表明,如果将一家公司的 CRT 显示器全部换成 LCD 显示器的话,同一办公空间可以多容纳 24% 的人员。在办公室租金越来越昂贵的今天,这一项费用的节省也不可小视。另外,空间节省带来的另一项直接好处是可以购买更加精简的办公桌。

4) 节省能源

CRT 显示器需要加热电极元件使电子枪以极高的速度发射电子束,这也是 CRT 显示器耗能的主要原因。而 LCD 显示器功耗大约是 CRT 显示器的 $1/3$ 。

5) 有利于健康

LCD 显示器无辐射,无闪烁,因而会使使用者眼睛感觉非常舒适,这对因长期使用 CRT 显示器而使健康受到影响的人员来讲无疑是一大福音。

两大短处如下。

1) 图像质量仍不够完善

LCD 显示器的图像质量仍不够完善,这主要体现在色彩鲜艳和饱和度上。

2) 响应时间仍然过长

液晶显示器的响应时间仍然过长,静止画面时也许还可以,若用来玩游戏或播放更新速度快的动画或视频,液晶显示器的弱点就暴露出来了。

8.4.3 显示适配器

1. 显示标准

显示适配器提供了计算机与显示器之间的接口,并将 CPU 发送来的数字信号转换为模拟信号在显示器上呈现出来。综观计算机的发展历史,对于显示系统特性的一系列的标准,呈现出屏幕分辨率和彩色深度上的稳步提高。表 8-3 所列的标准可以作为计算机视频显示技术的简要历史。IBM 制定了其中的大部分标准,为其他兼容 PC 制造商所采纳。这些标准中的不少已经被废弃,唯一的例外是 VGA,它仍然是当今市场上各种显示适配器的基础。表 8-4 是目前使用的主要显示标准。

表 8-3 PC 的显示标准

标准	显示模式	最高分辨率	最大色彩数	主机机型
MDA	字符	720 × 350	2(单色)	8086 IBM - PC
HGA	字符/图形	720 × 350	2(单色)	8086 IBM - PC
CGA	字符/图形	640 × 200	4	8086 IBM - PC
EGA	字符/图形	640 × 350	16	286 IBM - PC
VGA	字符/图形	640 × 480	256	386 - 486PC, PS/2
SVGA	字符/图形	1600 × 1200	16M	486 - Pentium

注:

- MDA: Monochrome Display Adapter, 单色显示适配器
- HGA: Hercules Graphics Adapter, 大力神图形适配器
- CGA: Color Graphics Adapter, 彩色图形适配器
- EGA: Enhanced Graphics Adapter, 增强型图形适配器
- VGA: Video Graphics Array, 视频图形阵列
- SVGA: Super Video Graphics Array, 超级视频图形阵列

表 8-4 VGA 的扩展标准

格式	分辨率(像素)	格式	分辨率(像素)
VGA	640 × 480	WQXGA	2 560 × 1 600 (16:10)
SVGA	800 × 600	QXGA	2 560 × 2 048, 4 × (1 280 × 1 024)
UVGA, XGA, EVGA	1 024 × 768(标准的 15"LCD)	WQXGA	3 200 × 2 048
XGA +	1 152 × 864	QUXGA	3 200 × 2 400 (4:3)
WXGA	1 280 × 768	WQUXGA	3 840 × 2 400, 4 × (1 920 × 1 200) (HDTV)
WXGA	1 280 × 1 024(标准的 17 英寸 LCD)	HXGA	4 096 × 3 200, 4 096 × 3 072 (4:3)

格式	分辨率(像素)	格式	分辨率(像素)
WXGA +	1 400 × 1 050	WHXGA	5 120 × 3 200, 5 120 × 3 072 (16:10)
WSXGA	1 680 × 1 024	HXGA	5 120 × 4 096 (5:4)
WSXGA +	1 680 × 1 050 ^a	WHXGA	6 400 × 4 096 (15:6:10)
UXGA	1 600 × 1 200, 1 600 × 1 280, 1 920 × 1 440(4:3)	HUXGA	6 400 × 4 800 (4:3)
WXGA	1 920 × 1 200(16:10)	UHDTV	7 680 × 4 320 (UHDTV)
QXGA	2 048 × 1 536, 4 × (1 024 × 768)	WHUXGA	7 680 × 4 800 (16:10)

注:

- Q——Quad
- X——eXtended
- S——Super
- U——Ultra
- W——Wide
- H——Hex
- G——Graphics
- A——Array
- HDTV——高清晰度电视

(1) 显示模式

由于计算机显示技术的发展非常迅速,因此应用中不可避免地会遇到新显示系统与老的应用软件之间的兼容性问题。一般显示适配器采用多种显示模式来解决这一问题。显示模式从功能上分为两大类:字符模式和图形模式。

字符模式也称字母数字模式,即 A/N 模式(Alpha Number mode)。在这种模式下,显示缓冲区存放显示字符的代码(一般采用 ASCII 码)和属性。显示屏被划分为若干字符显示行和列,例如,80 列,25 行。

由于字符模式在 CRT 上不是点控制,而是一个由 8 位代码(ASCII 码)控制的一块(比如 8 × 8 像素或 8 × 14 像素)显示区域。因此显示缓冲区较小,显示更新的速度非常快,但缺点是无法显示图形。目前流行的所有显示适配器都支持字符模式。

图形模式也称 APA 模式(All Points Addressable mode)。由于显示适配器种类繁多,因而出现了各种显示模式,支持各种用途(字符、图形)。其中,有些显示模式是标准模式,在不同类型显示器及不同厂家的显示适配器之间是通用的。还有些显示模式是专用的,而这些专用的显示模式也有两种情况:一种是这一类卡专用,另一种是某一类卡中某个厂家的产品专用。

(2) 标准 VGA 显示模式

实际上,当今市场上没有哪一种显示适配器被限制成只支持标准 VGA 显示模式(16 色 640 × 480 像素的分辨率)。此标准 VGA 显示模式已经成为计算机图形显示配置器的基线,是所有运行 Windows 系统的显示适配器必须支持的最一般标准。所有版本的 Windows 系统安装程序都把 VGA 设置作为默认的显示设置。除了 VGA 以外,依靠硬件的性能,大部分显示适配器能支持更高的分辨率和颜色深度。如果由于启动时的问题,Windows 9x 系统只能工作在安全模式,

系统将自动采用 16 色的分辨率为 640×480 像素的默认 VGA 模式。

2. 计算机显示系统的发展

从 CGA、HGA、EGA、VGA 到现在的各种 VGA，显示系统的发展特点是分辨率越来越高，颜色越来越丰富，显示适配器的智能化程度越来越高。

随着计算机的广泛应用，特别是 CAD 系统、计算机桌面印刷系统等要求计算机需配备更高分辨率和更多颜色的专用显示系统。随着高档计算机的普及应用，这种专用系统，将会变成新的通用系统。

3. 显示适配器的工作原理

(1) VGA 显示适配器的组成

VGA 显示适配器由 VGA 控制电路、显示存储器和视频 BIOS ROM 三大部分组成，如图 8-17 所示。

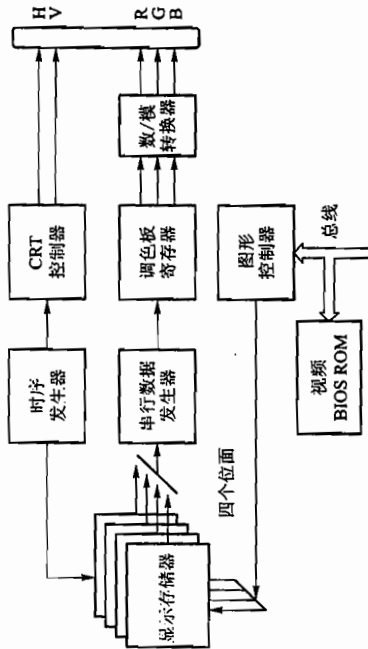


图 8-17 VGA 显示适配器的组成

VGA 控制电路包括由 CRT 控制器、时序发生器、图形控制器、串行数据发生器、调色板寄存器、数/模转换器等组成。CRT 控制器是一个可编程的大规模集成电路，负责显示器控制，确定扫描频率，确定 CPU、显示存储器和 CRT 三者之间的地址接口，输出 VGA 接口信号中的场同步 (V) 和行同步 (H) 信号。时序发生器用于产生 CRT 控制器和动态存储器所需的时序信号。图形控制器用于完成 CPU 对显示存储器的数据读写和运算操作。串行数据发生器用于在字符和图形的显示时，进行高速视频移位和属性处理。调色板寄存器用于将串行数据发生器送来的 4 位或 8 位数据转换成 18 位色彩数据 (R、G、B 各 6 位)。数/模转换器 (Digital to Analog Converter, DAC) 将 VGA 的数字颜色信号转换成模拟信号，驱动 CRT 显示 (因 VGA 使用的是模拟显示器)。

视频 BIOS 是固化在 ROM 芯片中的显示适配器的操作程序，VGA 显示适配器提供自己的专用操作程序库，一方面可以简化编程，同时使不同 VGA 显示适配器的操作完全一样。

VGA 的显示存储器用于存储显示数据，其容量取决于最高分辨率和色彩数。最初的 VGA 显示适配器装有 256 KB 动态存储器。标准 VGA 方式为 640×480 像素 16 色 (模式 12H)，每个

像素占用显示存储器半个字节，占用的存储器容量为 $640 \times 480/2 = 153\,600$ 字节，所以需要安装 256 KB 显示存储器。当使用 16 色时，采用如图 8-18 所示的位平面结构；若要使用 256 色 (模式 13H)，就要使用如图 8-19 所示的线性存储结构，每个像素用一个字节来表示。

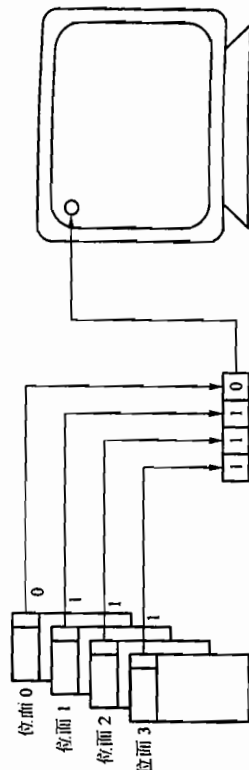


图 8-18 VGA 显示卡 16 色位平面结构

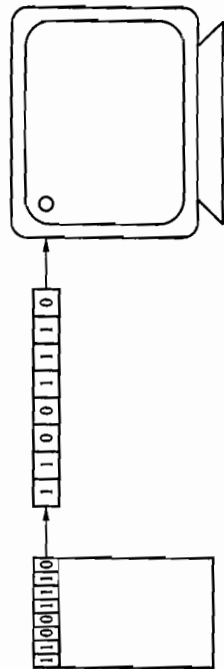


图 8-19 VGA 显示卡 256 色线性存储结构

(2) 显示适配器的总线

通常显示适配器是为特定系统总线而设计的，总线影响了系统处理视频信息的速度。为提高显示适配器与主机间的数据传送速率，Intel 设计了专用的视频总线——加速图形端口 (AGP)。AGP 总线本质上是对 PCI 总线的增强。AGP 总线提供对系统主存储器阵列的高速访问，使得显示适配器能直接从系统存储器中读取某些 3D 视频元素 (如纹理映射)，而不需要在处理前将数据复制到适配器存储器中。这样既节省了时间，也省去了为更好地支持 3D 功能而升级显示适配器存储器的需要。AGP 总线只能用于显示适配器。

随着 PCI_E 接口的普及，最新的显示适配器一般采用 PCI_E 接口。PCI_E 具有比 AGP 更高的数据传输速率。

(3) 显示适配器的接口

目前常见的显示适配器接口主要有 D-Sub 接口和数字可视化接口 (Digital Visual Interface, DVI)。除此之外，还有 13W3 接口 (SUN 公司的传统设计，为 SUN 图形工作站专用)，目前仍有采用该接口的显示适配器在继续开发、生产、使用，但已经同时做上了 D-Sub 或者 DVI 接口。

随着时代的发展，各式各样的接口会逐渐趋于统一。由于 DVI 同时具有数字和模拟型号的

特点,可以分别用于支持 CRT 显示器和越来越普及的 LCD 液晶平板以及等离子等数字化显示器,估计在不久的将来,会成为显示器和显示适配器领域的唯一事实接口标准。

1) D-Sub 接口

显示适配器所处理的信息最终都要输出到显示器上,显示适配器的输出接口就是计算机主机与显示器之间的桥梁,它负责向显示器输出相应的图像信号。

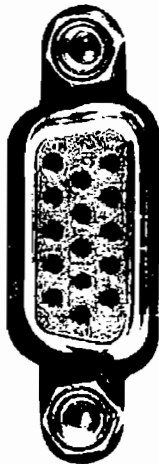


图 8-20 D-Sub 接口

CRT 显示器因为设计制造上的原因,只能接受模拟信号输入,最基本的信号包含 R、G、B、H 和 V(分别对应红、绿、蓝、行和场)5 个分量,这就要求不管以何种类型的接口接入,其信号中至少包含以上这 5 个分量。除了这 5 个必不可少的分量外,最重要的是在 1996 年以后的显示器中还增加 DDC 数据分量,用于读取显示器 EPROM 中记载的有关显示器品牌、型号、生产日期、序列号、指标参数等信息,以实现 Windows 所要求的“即插即用”功能,完整的 DDC 分量一般需要 3 根信号线传输。

目前常用的接口只是在原 VGA 的基础上进行了扩展,包括特色适配器。VGA 接口就是显示适配器上输出模拟信号的接口,采用 HD 型 15 引脚模拟接口,即 VGA 接口,也叫 D-Sub 接口,如图 8-20 所示。D-Sub 接口的引脚的功能如表 8-5 所示。

表 8-5 15 引脚 VGA 接口引脚的功能

引脚	功能	方向	引脚	功能	方向
1	红色视频	输出	9	键脚(插入孔)	
2	绿色视频	输出	10	同步地	
3	蓝色视频	输出	11	显示器标识 0	输入
4	显示器标识 2	输入	12	显示器标识 1	输入
5	TTL 地(显示器自测)		13	水平(行)同步	输出
6	红色模拟地		14	垂直(帧)同步	输出
7	绿色模拟地		15	显示器标识 3	输入
8	蓝色模拟地				

虽然 LCD 显示器可以直接接收数字信号,但很多低端产品为了与 VGA 接口显示适配器相匹配,因而采用 VGA 接口。VGA 接口是显示适配器上应用最为广泛的接口类型,绝大多数的显示适配器都带有此种接口。

目前大多数计算机与外部显示设备之间都是通过模拟 VGA 接口连接,计算机内部以数字方式生成的图像信息,由显示适配器中的数字/模拟转换器转变为 R、G、B 三原色信号和行、场同步信号,信号通过电缆传输到显示设备中。对于模拟显示设备,如模拟 CRT 显示器,信号被直接送到相应的处理电路,驱动控制显像管生成图像。而对于 LCD、DLP 等数字显示设备,显示设备中需配置相应的模/数转换器(Analog to Digital Converter, ADC),将模拟信号再转换为数字信号。

在经过 D/A 和 A/D 二次转换后,不可避免地造成了一些图像细节的损失。VGA 接口应用于 CRT 显示器无可厚非,但用于连接液晶之类的显示设备,则转换过程的图像损失会使显示效果略微下降。

2) DVI 接口

上述模拟信号的 VGA 接口对于常规的计算机应用(诸如网络浏览、文字处理、显示图片)来说已经足够了,但对于新兴的高清视频节目而言必须要通过 DVI 接口传输。

DVI 是 1999 年由 Silicon Image、Intel、Compaq、IBM、HP、NEC、Fujitsu 等公司共同组成数字显示工作组(Digital Display Working Group, DDWG)推出的接口标准。它是以 Silicon Image 公司的 ParallelLink 接口技术为基础,基于最小化传输差分信(Transition Minimized Differential Signaling, TMDS)电子协议的基本电气连接。TMDS 是一种微分信令机制,可以将像素数据编码,并通过串行连接传递。显示适配器产生的数字信号由发送器按照 TMDS 协议编码后通过 TMDS 通道发送给接收器,经过解码送给数字显示设备。一个 DVI 显示系统包括一个发送器和一个接收器。发送器是信号的来源,可以内置在显示适配器芯片中,也可以以附加芯片的形式镶嵌在显示适配器的电路板上;而接收器则是显示器上的一块电路,它可以接受数字信号,将其解码并传递到数字显示电路中,成为显示器上的图像。目前的 DVI 1.0 标准规定的最大单通道带宽为 165 MHz,也就是在大多数情况下可实现 1 600 × 1 200 像素的分辨率。

与 VGA 接口不同,DVI 存在很多标准,共分为 3 大类 5 种标准,如图 8-21 所示。

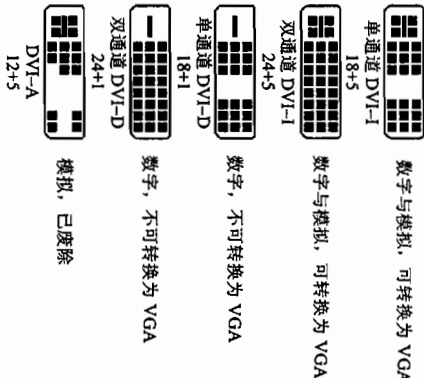
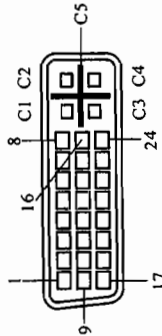


图 8-21 DVI 的标准

DVI 的引脚布局如表 8-6 所示。DVI 数字信号传输有单通道和双通道两种方式。对于单通道,仅用表 8-6 所示的 1/2、9/10、17/18 引脚传输,它的传输速率为 4.9 Gbps,而双通道可达 9.9 Gbps。

表 8-6 DVI 的引脚布局

DVI 连接器		引脚	功能	引脚	功能
	1	TMDS 数据 2 -	13	TMDS 数据 3 +	
	2	TMDS 数据 2 +	14	+5 V 直流电源	
	3	TMDS 数据 2/4 屏蔽	15	接地(+5 V 回路)	
	4	TMDS 数据	16	热插拔检测	
	5	TMDS 数据	17	TMDS 数据 0 -	
	6	DDC 时钟	18	TMDS 数据 0 +	
	7	DDC 数据	19	TMDS 数据 0/5 屏蔽	
	8	模拟垂直同步	20	TMDS 数据 5 -	
	9	TMDS 数据 1 -	21	TMDS 数据 5 +	
	10	TMDS 数据 1 +	22	TMDS 时钟屏蔽	
	11	TMDS 数据 1/3 屏蔽	23	TMDS 时钟 +	
	12	TMDS 数据 3 -	24	TMDS 时钟 -	
	C1	模拟垂直同步	C4	模拟水平同步	
	C2	模拟绿色	C5	模拟接地(RGB 回路)	
	C3	模拟蓝色			

在画面显示上,单通道的 DVI 支持的分辨率和双通道的完全一样,但刷新率却只有双通道的一半左右,会造成显示质量的下降。一般来讲,单通道 DVI 接口的最大刷新率只能支持到 1 920 × 1 080 像素/60 Hz 或 1 600 × 1 200 像素/60 Hz,即现有 23 英寸宽屏显示器和 20 英寸普通比例显示器的正常显示,屏幕尺寸再大的话就会造成显示效果的下降。

目前我们常用的液晶显示器还维持在 19 英寸普屏和 20 英寸宽屏左右,真正用上大屏幕显示器的人很少,因此 18 引脚的单通道 DVI 接口已经够用了。而使用大屏幕液晶显示器的话,24 引脚的双通道 DVI 是必须具备的条件。

考虑到兼容性问题,目前显示适配器一般会采用 DVD-I 接口,这样可以通过转换接头连接到普通的 VGA 接口。而带有 DVI 接口的显示器一般使用 DVI-D 接口,因为这样的显示器一般也带有 VGA 接口,因此不需要带有模拟信号的 DVI-I 接口。当然也有少数例外,有些显示器只有 DVI-I 接口而没有 VGA 接口。

在传统的 CRT 显示器中,使用的都是 VGA 接口,这是因为显像管的分辨率并没有严格的点阵划分,每一个点都是由红、黄、蓝三个小点呈三角形排列组成,并且颜色数也没有上限,无论使用哪种分辨率,显示出来的图像边缘过渡平滑自然。DVI 接口现在多用在 LCD 显示器上, LCD 显示器一般都有固定的最大分辨率,它的每一个点都是实际存在的,颜色数一般是 1670 万种。不过现在的 15 英寸液晶显示器一般只提供一个 VGA 接口,只有在大屏幕高档 LCD 显示器中才提供 DVI 和 VGA 双接口。现在的高档显示适配器上,一般也提供 DVI 接口,可以与 LCD 显示器相连。

目前常用的 DVI 接口分为两种,一个是 DVI-D 接口,只能接收数字信号,接口上只有 3 排 8 列共 24 个引脚,其中右上角的一个引脚为空,如图 8-22 所示。另外一种则是 DVI-I 接口,可同时兼容模拟和数字信号,如图 8-23 所示。

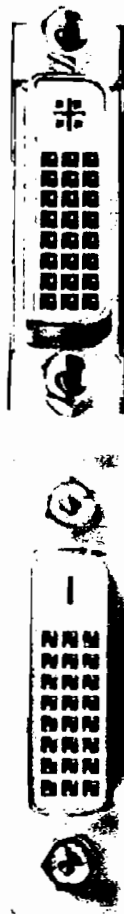


图 8-22 DVI-D 接口

DVI-I 接口可以通过转换头转换为 D-Sub,如图 8-24 所示,一般采用这种接口的显示适配器都会带有相关的转换头。

显示设备采用 DVI 接口具有以下优点。

1) 速度快

DVI 传输的是数字信号,数字图像信息不需经过任何转换,就会直接被传送到显示设备上,因此减少了繁琐的数字→模拟→数字转换过程,大大节省了时间,因此它的速度更快,有效消除了拖影现象,而且使用 DVI 进行数据传输,信号没有衰减,色彩更纯净,更逼真。

2) 画面清晰

计算机内部传输的是二进制的数字信号,使用 VGA 接口连接 LCD 显示器的话就需要先把信号通过显示适配器中的数字/模拟(D/A)转换成 R、G、B 三原色信号和行、场同步信号,这些信号通过模拟信号线传输到 LCD 显示器内部后,还需要进行相应的模拟/数字(A/D)转换,将模拟信号再一次转变成数字信号,才能在 LCD 显示器上显示出图像来。在上述的 D/A、A/D 转换和信号传输过程中不可避免会出现信号的损失和受到干扰,导致图像出现失真甚至显示错误,而 DVI 接口无需进行这些转换,避免了信号的损失,使图像的清晰度和细节表现力都得到了大大提高。

3) 抗干扰能力

现在各式各样的家用电器在不断增多,再加上广播、电视和手机等等,电磁干扰会越来越严重。如果你的计算机和电视、音响、微波炉、冰箱等电器放在同一间屋子里,当它们当中有一样同时工作的时候,一定会发现显示器或多或少地受到干扰。同时,计算机对其他设备也会产生干扰,这是不可避免的。在这种情况下,数字信号就可以发挥优势,它的抗干扰能力要比模拟信号强得多。举个简单的例子,我们在使用 GSM 移动电话时,经常会发现声音时大时小、断断续续,这就是信号受到了外界的干扰,而使用数字技术的 CDMA 电话就不会出现这种情况。所以,使用 DVI 数字接口要比使用 VGA 模拟接口好得多,而且支持的传输距离更长。

4) 易于升级

从模拟到数字的转变是发展的必然趋势,就像从模拟信号的磁带到数字信号 CD、MP3,

图 8-23 DVI-I 接口

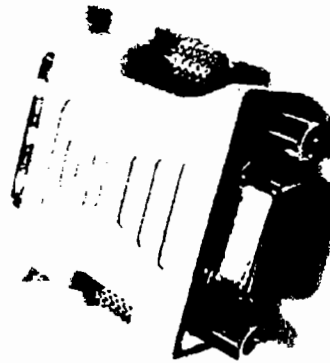


图 8-24 DVI-I 至 D-Sub 的转换器

从模拟的 VHS 录像机到数字化的 VCD、DVD 播放机转变一样,数字技术的优势是不言而喻的。如今信息技术的更新速度之快令人咋舌,今天刚刚流行的东西也许明天就处在淘汰的边缘,便于升级也就成了现在电子产品不可缺少的要素之一。由于 DVI 使用数字传输,所以仅仅需要修改显示适配器或显示器内部的程序就可以做到设备升级,而不用考虑中间的传输过程。

虽然如此,但 DVI 要在短期内淘汰 VGA 还具有一定的限制,因为匹配的 DVI 显示适配器和自身制造成本还相对较高,但 DVI 接口无疑是未来显示器信号接口的一个趋势和标准。所以,现在不少 LCD 显示器提供了 DVI 和 VGA 两个接口。

如今的高端微型计算机能利用专门的独立显示适配器(诸如 nVidia 等公司的产品)所具有的 DVI 来连接高清显示设备,而高清电视机也能利用 HDMI(High Definition Multimedia Interface)标准从机顶盒等接收设备(或通过自身内置相关模块)获得高清数字信号。DVI 和 HDMI 这两种标准能相互兼容,不过 HDMI 还多提供一些功能,诸如通用的遥控功能,并且支持仅通过一根线缆传输的多频道信号接收。

VESA 颁布了“即插即显”(plug and display, P&D)标准,使用户能够在系统中热安装显示器并且不必手动调整系统设置。这种标准还提供了数字化接口来取代陈旧的模拟 VGA 标准。由于 LCD 显示器本身就是数字化的,因此 P&D 省去了计算机和显示设备之间的数模转换和模数转换过程。

由 Intel、Apple、nVidia、SGI、Samsung 和 Foxconn 等公司发起成立的 UDI SIG(Unified Display Interface Special Interest Group),旨在为计算机和高清晰度显示器之间的连接制定并推广一种新的标准接口。该协会计划通过一种新的可以传输高清视频节目的连接技术来取代已经跟不上时代发展的传统 VGA 接口。UDI 标准承诺将对这些主流的普通计算机也推出 HDMI 所具备的全部功能和特性,并且将与 HDMI 和 DVI 一样共同采用 HDCP(High bandwidth Digital Content Protection)技术,达到完全兼容,使得采用 UDI 的计算机既可以与具备 DVI 接口的显示器连接,也可以向 HDMI 接口的高清电视机传输内容。

8.5 音频设备

微型计算机的音频设备主要是由声卡、音箱和话筒组成的音频信号处理设备,是多媒体计算机的重要组成部分。今天计算机的音频硬件通常是将声卡做成扩展卡的形式(也就是通常所说的声卡),插在计算机总线插槽内。像显示适配卡一样,现在有些系统将音频硬件直接集成在主板上。声卡上提供了一些插孔,可以连接喇叭、话筒以及其他设备,比如游戏杆和 MIDI 硬件。本节的重点是介绍目前计算机中的音频设备及其发展历史、工作原理、使用方法、技术指标和最新技术。

8.5.1 音频设备的发展历史及用途

早期的计算机,除了初级的蜂鸣声和音调外,不具备其他的音频功能。直到 ADLIB 声卡的诞生才使人们享受到了真正悦耳的音效。ADLIB 声卡是由英国的 ADLIB AUDIO 公司研发的,最

早的产品于 1984 年推出,它的诞生开了微型计算机音频技术的先河。在相当一段时间里,ADLIB 的声卡曾是多媒体领域的一个重要标准,直到 CREATIVE 崛起后,ADLIB 才逐渐退出历史舞台。

新加坡的 CREATIVE 公司在 20 世纪 80 年代后期推出的 Sound Blaster(声霸)系列的声卡开创了微型计算机语音处理的新时代,Sound Blaster 声卡是 CREATIVE 公司的第一代声卡产品,但是在功能上已经比早期的 ADLIB 声卡增强不少,其最明显的特点在于兼顾了音乐与音效的双重处理能力,虽然它仅为 8 位、单声道的声卡,在声音的回放效果上精度较低,但它却使人们第一次在计算机上得到了音乐与音效的双重听觉享受。此后 CREATIVE 又推出了后续产品——Sound Blaster PRO,它增加了立体声功能,进一步加强了计算机的音频处理能力。因此 Sound Blaster PRO 声卡在当时被编入了 MPCI 规格(第一代多媒体标准)。

在取得了音乐与音效的完美组合之后,CREATIVE 并没有满足现状,它们在技术上寻求新的突破。前面提到过,Sound Blaster 与 Sound Blaster PRO 都只有 8 位的信号采样精度,我们可以将其直接理解为音质的粗糙,虽然 SB PRO 拥有立体声处理能力,但依然不能弥补采样损失所带来的缺憾。Sound Blaster 16 的推出彻底改变了这一状况,它是第一款拥有 16 位采样精度的声卡,人们终于可以通过它实现 CD 音质的信号录制和回放,使声卡的音频品质达到了一个前所未有的高度。在此后相当长的时间内 Sound Blaster 16 成为多媒体音频部分的新一代标准。

从 Sound Blaster 到 Sound Blaster PRO,再到 Sound Blaster 16,CREATIVE 逐渐确立了自己声卡霸主的地位。期间技术的发展和成本的降低,也使得声卡得以从一个高不可攀的奢侈品高度(早期的声卡非常昂贵),渐渐成为普通多媒体计算机的标准配置。

从 Sound Blaster 一直到 Sound Blaster Ave 64 GOLD,声卡始终是采用 ISA 接口形式的。不过随着技术的进一步发展,ISA 接口过低的数据传输速率成为声卡发展的瓶颈。把接口形式从 ISA 转移到 PCI 成为声卡发展的大势所趋。PCI 声卡从理论上具有加大传输通道(ISA 为 8 MB/s,PCI 可达 133 MB/s),提升数据带宽的功能。从而在声卡上实现三维音效和杜比环绕立体声(DLS)技术,使得声卡的性能得到多方面提升,但总体成本却大幅度下降,可谓两全其美。

综上所述,我们可以将声卡的发展历史概括为——ADLIB 开创了声卡技术的先河;Sound Blaster 首次综合了音乐和音效;Sound Blaster PRO 和 Sound Blaster 16 则完善了这一系列的技术规格;而 Sound Blaster Ave 32 和 Sound Blaster Ave 64 开创了新的波表合成技术;PCI 声卡的出现标志着新技术和新挑战的不断涌现。

8.5.2 声卡的工作原理

近年来声卡发展快速,目前各类品牌的声卡非常多,但是就声卡的一些基本原理来看,其实都是一样的。所以我们将从声音的数字化开始介绍,进而说明声卡的基本原理及对声音的处理流程。

1. 声音的数字化

在自然界我们所听到的声音都是经由空气或一些介质所传播的,从物理学的观点来看,声音可以用波形图表示出音高(Pitch)、音量(Volume)和音色(Tone 或 Timbre)3 种组成要素。自然界的声音都是连续的模拟信号,然而计算机只能处理 0 或 1 的数字信号。由此可见声卡最基本的

功能,就是把所录到的模拟信号转换为数字信号(模/数转换),以及将处理完成的数字信号源信号转换成模拟信号输出(数/模转换),如图8-25所示。

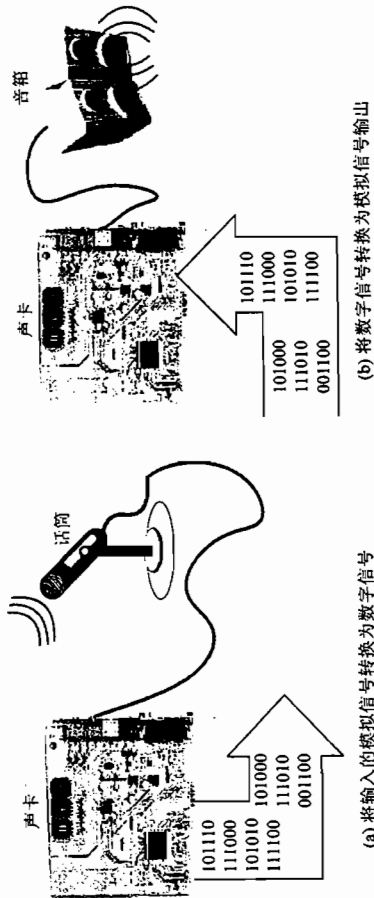


图 8-25 声卡的作用

2. 采样频率与采样精度

通常要把模拟信号转换成数字信号时,必须经过一个采样(sampling)过程,如图8-26所示。

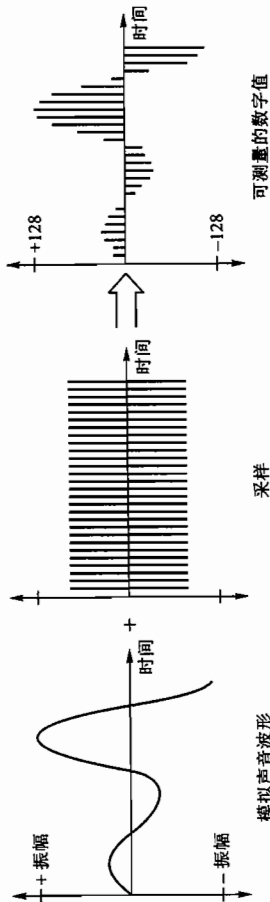


图 8-26 采样过程

在采样过程中,有几个参数会影响转换后声音的品质,其中之一就是采样频率(sampling frequency 或 sampling rate),其单位是赫兹(Hz),1 Hz就表示每秒采样一次。采样频率越高,数字信号就越接近原声。根据采样定理,采样频率只要达到信号最高频率的两倍,就能精确描述被采样的信号。一般来说,人耳的听力范围为20 Hz~20 kHz。因此,只要采样频率达到20 kHz \times 2=40 kHz时,就可以满足人们的要求。目前大多数声卡的采样频率都已达到44.1 kHz或48 kHz。另一个影响因素是采样精度(sampling resolution)。采样精度可以用位(bit)来表示。如果声卡是8位,就是每次采样的数据可解析成 $2^8=256$ 种表示等级,如果是16位的声卡就有 $2^{16}=65\,536$ 种。如果不考虑其他的因素,采样精度高的16位声卡的音质要好于8位声卡。CD

音质的采样精度至少要16位,采样频率要在44.1 kHz左右。

这里顺便提一下,常听一些人说:Sound Blaster Awe 32是32位声卡,而Sound Blaster Awe 64则是64位声卡,因此音质比16位声卡要好。其实这是认识上的一个误区,以上两款声卡的“32”、“64”指的是32种复音与64种复音。复音越多,音效越逼真,但这与采样位数无关。实际上,Sound Blaster Awe 32与Sound Blaster Awe 64均为16位声卡。

3. 模拟-数字转换和数字-模拟转换

CODEC是由COrder和DECoder组合而成的缩写字,即编码解码器。它的作用,也就是声卡最关键的两个作用——把来自计算机的数字信号转换为模拟音频信号;把声卡的模拟音频输入信号转换为数字信号,即模拟-数字和数字-模拟转换。这两种转换分别由ADC和DAC芯片完成。早期CODEC是内置在音效芯片之中,后来因AC'97(由Intel Architecture Labs于1997年开发的音频编码/解码标准,全称为Audio Codec'97)规范的品质要求,CODEC便从音效芯片中独立出来。CODEC决定了模拟输入输出的质量。

所谓AC'97的软声卡,是指在主板上集成了CODEC芯片,而其他的一切都交由CPU完成的,这节省了数字信号处理器(DSP)的费用,降低声卡的制造成本。

(1) 模拟-数字转换

输入的模拟声音信号经过ADC后会被转换成一系列的不连续信号,这也就是我们一开始所说的采样。通常信号摆动的范围必须在A/D转换器可以适用的范围内,而且采样精度不可太低,如此才能维持好的音质。

(2) 数字-模拟转换

数字-模拟转换是指将不连续的数字信号转换成连续的模拟声音信号。实际上,声音从原先的模拟信号转换成数字后,经过声卡的编辑处理,再经由数字-模拟转换,才可以从声卡输出。这一连串的处理过程,所输出的声音与原始的音源已经有所差别,即一般所说的失真(Distorsion)。

4. 数字信号处理器(DSP)

数字信号处理器(Digital Signal Processor, DSP)是一种专门进行数字信号处理的处理器,其功能通常包括采样频率的控制,对声音的录制与播放控制,处理MIDI指令等。有些声卡的DSP还有音频数据压缩的功能。另外,如果声卡有混音芯片(Mixer chip),就可以通过软件的操作来对声音做各种控制,如音量的高低控制、音场调整效果等。所以DSP是声卡中非常重要的芯片,所有数字音频信号的处理,都属于DSP的功能范围。至于声卡中,是将所有功能都制作在同一芯片里,或是各种功能独自做成单独的芯片,就完全取决于各声卡厂商的设计了。

并不是所有的声卡都有DSP芯片,很多声卡没有DSP,它们直接使用I/O控制芯片作为主芯片,而是只靠I/O芯片来协调音频的输入与输出。也有声卡带有DSP芯片,但并不让DSP去运行效果器插件,而是只负责声卡的工作。

DSP或I/O控制器决定了声卡内部数字信号的质量。

5. 音效合成

在声音信号未经过DAC之前都是数字式的,可以将几个数字信号音源组合出各种声音而产生音效,这就是一般所称的调频(Frequency Modulation, FM)。例如,Yamaha-3812采用两种音效合成的演算方法,一般也称作OPL2;而Yamaha-262和Yamaha-263采用两种音效合成的演

算方法,一般则称作 OPL3 及 OPL4。

6. MIDI 系统

乐器数字接口 (Musical Instrument Digital Interface, MIDI) 就是连接各种电子乐器及相关设备的通信标准。最早的 MIDI 统一规格来自日本,后来美国也成立了国际 MIDI 协会 (International MIDI Association, IMA), 协调 MIDI 标准的制定工作。

MIDI 是一种用于计算机与电子乐器之间进行数据交换的通信标准。MIDI 文件 (通常以 MID 为文件扩展名) 记录了用于合成 MIDI 音乐的各种控制指令,包括发声乐器、所用通道、音量大小等。由于 MIDI 文件本身不包含任何数字音频信号,因而所占的存储空间比 WAV 文件要小得多。MIDI 文件的回放需要通过声卡的 MIDI 合成器合成成为不同的声音,而合成的方式有调频与波表 (Wave table) 两种。早期的声卡及目前大多数廉价的声卡都采用调频合成方式。调频合成是通过振荡器产生正弦波,然后再叠加成各种乐器的波形。由于振荡器成本较高,即使是 OPL3 这类高档的调频合成器也只提供了 4 个振荡器,仅能产生 20 种复音。因此 MIDI 音乐听起来生硬呆板,带有明显的人工合成色彩,即所谓的电子声。与调频合成不同,波表合成是采用真实的语音样本进行回放。语音样本记录了各种真实乐器的波形采样,并保存在声卡上的 ROM 或 RAM 中。因此,要分辨一块声卡是否波表声卡,只需看卡上有没有 ROM 或 RAM 存储器即可。

7. 声音基本处理流程

声卡对声音的处理流程如图 8-27 所示。当一个音源输入后,会先经过滤波器做预先的采样,模拟-数字转换,再由 DSP 负责对此音源做各种处理,其中可能包括由调频芯片产生合成音效,或是从波表存储器取出音源;这些经过处理后的数字音源,再经过数字-模拟转换而输出。有些声卡含有混音处理芯片,可以处理多种音源的输入,并提供软件调整音量的功能。

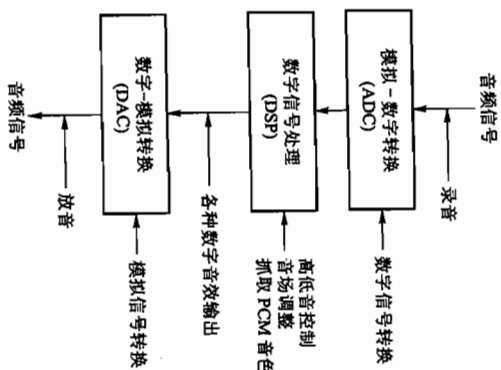


图 8-27 声卡处理流程

8.5.3 声卡的输入输出接口

大多数声卡输入输出接口都是一样的。通过小型插孔连接器将声音信号输出到喇叭、耳机或立体声音响系统,或从话筒、CD 唱机、磁带式录音机或立体声音响输入声音。图 8-28 显示了声卡上的 4 种类型连接器。

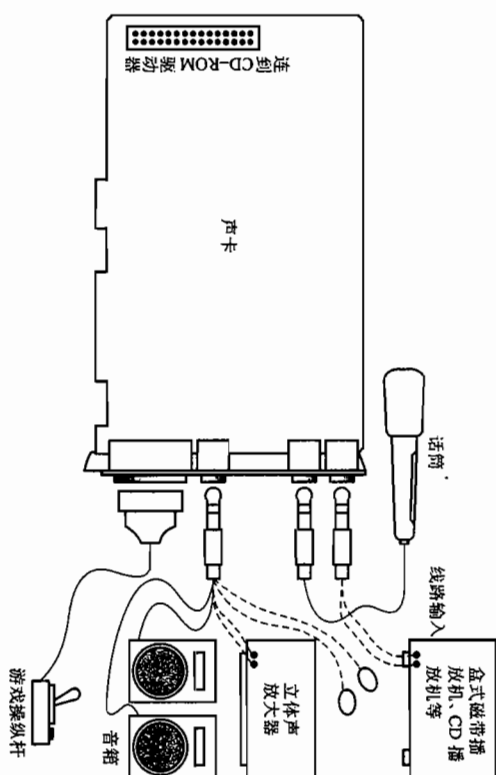


图 8-28 声卡输入输出连接器

1. 线路输出连接器

线路输出连接器用以将声音信号从声卡发送给计算机的外部设备。可以将线路输出接到立体声喇叭、耳机或者立体声音响系统。如果将计算机连接到立体声音响系统,就能听到经过放大的声音。有些声卡,例如 Microsoft 的 Windows 声音系统,提供两个插孔作为线路输出,一个输出立体声的左声道信号,另一个输出右声道信号。

2. 线路输入连接器

通过线路输入连接器,可以从外部音源 (例如立体声音响或录音机) 输入声音,进行录音或混音,并存储到计算机硬盘上。

3. 话筒输入连接器

话筒输入连接器用来连接话筒录制声音。很多声卡内部都有自动增益控制 (Automatic Gain Control, AGC) 功能,用来提高录音质量,能在录音过程中自动调整声音。动圈式或电容式话筒最适合这种插孔。

4. 游戏杆和 MIDI 连接器

游戏杆和 MIDI 连接器是一个 15 引脚的 D 形连接器,可以连接任何标准的游戏杆或游戏控制

8.6 打印机

打印机是计算机的传统输出设备。打印机的功能是将计算机中的数字信息以打印的方法转换为书面信息。按照打印机的工作原理,打印机分为击打式和非击打式两大类,其中击打式主要是针式打印机;非击打式种类较多,如激光打印机、喷墨打印机、热转印打印机、热敏式打印机等。其中常见的打印机主要有针式打印机、喷墨打印机和激光打印机。本节主要介绍目前使用最广泛的针式打印机、激光打印机和喷墨打印机的功能、工作原理、主要技术及打印机的使用方法等。

8.6.1 针式打印机

针式打印机在打印机历史上的很长一段时间内曾经占有着重要的地位。针式打印机以其便宜、耐用、可打印多种类型纸张等原因,普遍应用在多种领域,我们常用的 EPSON LQ-1600K、STAR CR-3240 等属于宽行针式打印机。EPSON LQ-100、NEC-P2000 则属于窄行针式打印机。

宽行针式打印机可以打印 A3 幅面的纸,窄行针式打印机一般只能打印 A4 幅面的纸张。针式打印机可以打印穿孔纸,它在银行、机关、企事业单位发挥着很大作用。另外,针式打印机有其它机型所不能代替的优点,就是它可以打印多层纸,这使之在报表处理中的应用非常普遍。但针式打印机的打印质量低,工作噪声大,打印速度慢,使它无法适应高质量、高速度的打印需要,所以在普通家庭及办公应用中有着逐渐被喷墨打印机和激光打印机所取代的趋势。目前,针式打印机主要用于银行、超市等部门打印单据。

针式打印机主要有打印头、字车机构、色带机构、输纸机构和控制电路组成。打印时打印机通过打印针头击打色带,使色带上的墨印在纸上形成文本或图形。针式打印机的关键部件是打印头。打印头由若干根钢针组成,常见的有 9 针、24 针等,现在的针式打印机通常都是 24 针打印机。这些打印钢针在垂直方向排成线,打印头从左向右移动,打印头里的钢针在驱动电路的控制下,击打色带和纸,从而形成一行字符。打印机每打印一行,打印走纸机构把纸推向前进行,而色带传动机构也将色带转动一定尺寸,使击打部位均匀地分布在整盘色带上。可以通过调整打印头与纸张的间距,从而适应打印纸的厚度,而且可以改变打印针的力度,以调节打印的清晰度,但注意色带用旧了要及时更换。

针式打印机通过 25 引脚电缆线连接到主机的并行端口(LPT),它使用独立的电源接线。在连接打印机到计算机上时,切记要在断电情况下操作,带电插拔打印电缆会烧坏打印机和计算机的连接端口。

一般可以用托纸架传送单页纸或用卡纸轮传送穿孔纸,还可以用手动旋钮微调打印纸的位置(注意,有些打印机在断电后,才可以方便地调节纸的位置)。

打印机的面板上都有控制按钮用于打印机操作,比如联机、进纸、退纸、微调等,通过打印机面板上的指示灯,可以及时了解打印机的状态,比如缺纸、联机、正在打印等。

制柄。如果购买了声卡选购件 Y 型适配器,有时游戏杆接口可以插入两个游戏杆。很多计算机在主板或扩展卡的多功能 I/O 电路中包含游戏杆接口,在连接游戏控制柄时要注意操作系统或者应用程序中用的是哪个接口。声卡一般将游戏杆接口也当做 MIDI 连接器,其中只有两个引脚用来与外部 MIDI 设备传输信号,例如电子键盘。在大多数情况下,必须从声卡厂商购买单独的 MIDI 插件插进游戏杆口,其上包括 MIDI 设备使用的两个圆形 5 引脚 DIN 连接器和一个游戏杆连接器。因为它们的信号使用不同的引脚,所以游戏杆和 MIDI 设备可以同时连接在上面。如果打算将计算机连接到外部 MIDI 设备,只要有该连接器就行了。还可以用声卡内部的合成器来播放网站上的 MIDI 文件。

5. 内部针形连接器

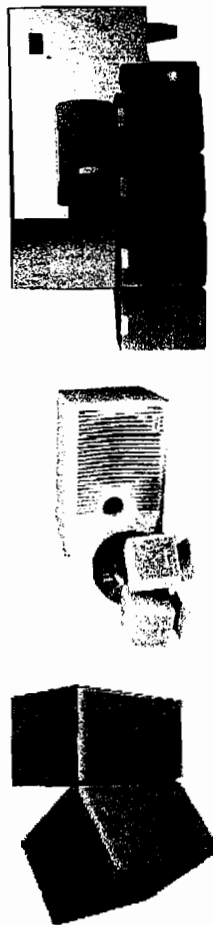
有些声卡带有内部针形连接器,可以通过带状电缆将内部 CD-ROM 驱动器直接连到声卡上。这种连接可以将音乐信号直接传送到 CD-ROM 驱动器,通过计算机的音箱进行播放。注意,这个连接器与有些声卡上的 CD-ROM 控制器连接器不是一回事,因为并不将 CD-ROM 的数据传送到系统总线,而只是通过声卡将 CD-ROM 驱动器的声音信号传送给音箱。如果声卡上没有这个连接器,仍然可以播放音乐 CD,只要用外部电缆将 CD-ROM 驱动器的耳机插孔与声卡上的线路输入插孔相连即可。

6. 话筒和音箱

在将声音录制成 WAV 文件时需要有一只话筒。选择话筒非常简单,要选择有 1/8 英寸小插头的,能插入声卡的话筒插孔或者线路输入插孔。多数话筒上都有开关。如果需要录制音乐,就要买一只高档话筒,但是相应的声卡必须能对话筒产生的信号进行适当的处理。一只高档话筒如果和一块廉价的 8 位声卡配合,不可能产生好的结果。

如果室内很嘈杂,就要买一只单方向的话筒,隔离外部的噪音。如果要给小组讨论录音,就要用全方向的话筒。

多媒体的视听时代发展到了今天,音箱作为一种必不可少的音频设备已经越来越被广大计算机用户所认识,它作为多媒体电脑的重要组成部分之一,在音频领域中有着不可替代的地位。当声卡完成了从 ISA 到 PCI 过渡的时候,随着声卡输出音乐质量的日益提高,特别是 4 声道声卡以及 DVD 杜比 AC-3 概念的出现,对音箱的要求也越来越高,计算机音箱已有了逐渐走向高档的趋势,4.1、5.1、6.1 和 7.1 声道的音箱已相继出现(图 8-29)。



(a) 普通音箱

(b) 带低音炮的音箱

(c) 5.1 声道的音箱

图 8-29 流行音箱的外形

8.6.2 喷墨打印机

喷墨打印机是在针式打印机之后发展起来的,采用非打击的工作方式。比较突出的优点是体积小、操作简单方便、打印噪音低、使用专用纸张时可以打出与照片相媲美的图片等。

1. 喷墨打印机的分类

目前,根据打印头的工作方式,喷墨打印机可以分为压电喷墨打印机和热喷墨打印机两大类。按照喷墨的材料性质又可以分为水质料、固态油墨和液态油墨等类型的打印机。

(1) 压电喷墨打印机

压电喷墨技术是将许多小的压电陶瓷放置到喷墨打印机的打印头喷嘴附近,利用它在电压作用下会发生形变的原理,适时地把电压加到它的上面,压电陶瓷随之产生伸缩使喷嘴中的墨汁喷出,在输出介质表面形成图案。用压电喷墨技术制作的喷墨打印头成本比较高,所以为了降低打印成本,一般都将打印头和墨水做成分离式结构,更换墨水时不必更换打印头。这种技术由Epson独创,因为打印头的结构比较合理,可通过控制电压来有效调节墨滴的大小和使用方式,从而获得较高的打印精度和打印效果。这种技术对墨滴的控制能力强,容易实现高精度的打印。现在1440 dpi的超高打印分辨率就是由Epson保持的。当然它也有缺点,假设使用过程中喷头堵塞了,无论是疏通还是更换喷头,费用都比较高而且不易操作,搞不好整台打印机可能就报废了。目前采用压电喷墨技术的产品主要是Epson公司的喷墨打印机。

(2) 热喷墨技术

热喷墨技术是让墨水通过细喷嘴,在强电场的作用下,将喷头管道中的一部分墨汁气化,形成一个气泡,并将喷嘴处的墨水顶出,喷出到输出介质表面,形成图案或字符。所以这种喷墨打印机有时又被称为气泡打印机。用这种技术制作的喷头工艺比较成熟成本也很低廉,但由于喷头中的电极始终受电解和腐蚀的影响,对使用寿命会有较大影响。所以采用这种技术的打印头通常都与墨水做在一起,更换墨盒的同时更换打印头。这样一来用户就不必再对喷头堵塞的问题太担心了。同时,为了降低使用成本,我们常常能看见给墨盒“打针”的情形(加注墨水)。在打印头刚刚打完墨水后,立即加注专用的墨水,只要方法得当,可以节约不少的耗材费用。

热喷墨技术的缺点是在使用过程中会加热墨水,而高温下墨水很容易发生化学变化,性质不稳定,所以打出的色彩真实性就会受到一定程度的影响;另一方面,由于墨水是通过气泡喷出的,墨水微粒的方向性与体积大小不易控制,打印线条边缘容易参差不齐,一定程度上影响了打印质量,所以多数产品的打印效果不如压电喷墨技术产品。采用热喷墨技术的产品比较多,主要为Canon和HP等公司所使用。

(3) 固态喷墨技术

固态喷墨打印机是TEKTRONIX公司的专利技术。它所使用的变相墨在室温下是固态的,工作时将蜡质的颜料块先加温溶化成液体,然后再按前面所述的喷墨方法工作。这类打印机的优点是颜料的防水性能比较好,并且不存在打印头因墨水干涸而造成的堵塞问题。但采用固态油墨的打印机目前因生产成本比较高,产品比较少。

2. 喷墨打印机的组成

喷墨打印机分为机械和电路两大部分,机械部分通常包括墨盒和喷头、清洗部分、字车机构、

输纸机构和传感器等几个部分。墨盒和喷头有两种类型,一种是二合一的一体化结构,另一类是分离式结构。两种类型各有所长。清洗系统是喷头的维护装置。字车机构用于实现打印位置定位。输纸机构提供纸张输送功能,运动时它必须和字车机构很好地配合才能完成全页的打印。而传感器是为检查打印机各部件工作状态而特设的。

3. 喷墨打印机的主要技术指标

(1) 分辨率

分辨率是衡量打印质量的一个重要指标。分辨率以每英寸长度上打印的点数表示,单位为dpi。单色打印时分辨率越高打印效果越好。而彩色打印时情况比较复杂,通常打印质量的好坏要受分辨率和色彩调和能力的双重影响。由于一般彩色喷墨打印机的黑白打印分辨率与彩色打印分辨率可能会有所不同,所以选购时一定要注意商家所说的分辨率是哪一种分辨率,是否为最高分辨率。一般至少应选择分辨率在360 dpi以上的喷墨打印机。

(2) 色彩调和能力

对于使用彩色喷墨打印机的用户而言,打印机的色彩调和能力是个非常重要的指标。传统的喷墨打印机,在打印彩色照片时,若遇到过渡色,就会在三种基本颜色的组合中选取一种接近的组合来打印,即使加上黑色,这种组合一般也不能超过16种,对色阶的表达能力是难以令人满意的。为了解决这个问题,早期的喷墨打印机又采用了调整喷点疏密程度的方法来表达色阶。但对于当时彩色分辨率只有300 dpi左右的产品,调整疏密程度的结果是过渡色效果很差,看上去会有很多斑点。现在的彩色喷墨打印机,一方面通过提高打印密度(分辨率)来使打印出来的点变细,从而使图片变得更为细腻;另一方面,在色彩调和方面进行了改进,常用的方法有:增加颜色数量、改变喷出墨滴的大小、降低墨盒的基本颜色浓度等。其中增加颜色数量的方法最为有效。目前通常采用五色的彩色墨盒,加上原来的黑色墨盒,形成所谓的六色打印。这样一来颜色的组合数提高了许多倍,效果改善自然非常明显。

改变喷出墨滴大小的原理是在打印中需要色彩浓度较高的地方用标准大小的墨滴喷出,而在需要颜色浓度较低的地方喷射小墨滴,同样实现了更多的色阶。而降低墨盒基本颜色浓度的方法其实是在高颜色浓度的地方反复喷墨来形成更多的色阶。

(3) 打印速度

喷墨打印机的打印速度一般以每分钟打印的页数(page per minute, ppm)来表示。但因为每页上打印内容的多少并不完全一样,所以这个数字一定不会准确,只是一个平均数字。对于家用打印机,由于打印数量一般不太大,选购时可以不特别在意打印速度。

(4) 打印幅面

一般喷墨打印机的打印幅面有A4和A3两种。一般家庭用户使用A4幅面的就可以了。

8.6.3 激光打印机

由于具有速度快、分辨率高、对纸张要求不高,以及防水、易维护等优点,再加上价格上的日渐低廉,激光打印机已成为现代办公室及商用等领域中最普遍的打印输出工具。在大批量黑白文件的打印领域,毫无疑问,激光打印机占有绝对优势。至于彩色激光打印机,虽然有着速度及品质的优势,但价格上过于昂贵,所以在彩色输出领域,其市场占有率远不及热门的彩色喷墨

打印机。

1. 激光打印机的构造及打印过程

激光打印机的主要部件包括激光扫描组件、感光器、墨粉盒、显影单元、转印电极、放电灯、定影装置和输纸机构。

激光打印机中的激光组件,有时也称光栅输出扫描器(Raster Output Scanner, ROS),它根据页面缓冲器中的图像在感光鼓上产生像素点的静电模型。激光组件包括激光器、旋转棱镜和透镜。激光器总是静止的,为了在感光鼓的水平宽度上产生像素点的模型,棱镜不断旋转,而透镜负责激光束的聚焦,防止感光鼓上的像素变模糊。垂直方向运动由鼓自身转动完成。

详细地说,当使用者在应用程序中下达打印的指令后,整个激光打印流程的序幕遂由“充电”动作展开,也就是先在感光鼓上充上负电荷或正电荷,然后再将打印机处理器处理好的影像资料,透过光束,照射“曝光”到感光鼓上,形成所谓的静电潜像(电位为零),接着让碳粉匣中的碳粉带电,此时快速转动的感光鼓上的静电潜像表面,经过碳粉匣时,便会吸附带电的碳粉,并“显像”出图文影像。然后再将打印机进纸匣引进来的纸张,透过“转像”的步骤,让纸面带相反的正电荷或负电荷,由于异性相吸的缘故,如此便能使感光鼓上的碳粉吸附到纸张上。为使碳粉更紧密地附在纸上,接下来则以高温高压的方式,将碳粉“定影”在纸上,这也是每张纸打印出来的纸张都热乎乎的原因。然后再以刮刀将感光鼓上残留的碳粉“清除”。最后的动作即为“除像”,也就是除去静电潜像,使感光鼓表面的电位恢复到初始状态,以便开始下一个循环动作。激光打印中整个动作可说是充电(Charging)、曝光(Exposure)、显像(Development)、转像(Transferring)、定影(Fusing)、清除(Cleaning)及除像(Erasing)七大步骤的循环。

2. 激光打印的基本元件

激光打印机的基本元件包括感光鼓、充电元件及显像单元等。感光鼓是镀上光电导体(Optical Photo Conductor)的一支铝管,光电导体具有光线照射时为导体,暗室时会成绝缘体的特性,利用这样的特性,便可完成曝光及显像的动作。

激光打印机中充电元件的作用是让感光鼓得以布上正电荷或负电荷,最常见的有电晕充电器和充电滚轮两种,但由于电晕充电器会产生大量的臭氧,所以渐为充电滚轮所取代。显像单元主要承担形成图文影像的任务,包括让碳粉带电、搬运碳粉、控制碳粉在不同区域的厚薄以及碳粉的定位等。

8.6.4 打印机的接口

打印机的接口是指它与计算机主机的连接方式。主流打印机的接口通常分为并行接口(即计算机的打印机接口)和USB接口两种。早期的打印机使用普通并行电缆即可与计算机相连接。并行接口的优势在于安装简便、价格相对低廉,而传输速度稍慢,但对于打印机来说已足够了。目前,打印机普遍使用USB接口,与并行接口相比,速度更快、使用更方便。

PictBridge打印机(支持PictBridge数码相机直连打印协议的照片打印机)给用户带来了极大的方便。PictBridge标准诞生于2002年12月,为了规范行业市场,整顿过去由于许多数码相机直连打印协议互不兼容所造成的市场混乱局面,Canon、Fujifilm、HP、Olympus、Epson及Sony六家公司联合推出了一个称为“DPS”的具有通用性的行业标准。相机与影像产品协会(CIPA)对

其进行审议并通过后,将其正式命名为“影像设备的数码照片方案”(Digital Photo Solutions for Imaging Devices),简称PictBridge,如图8-30所示。

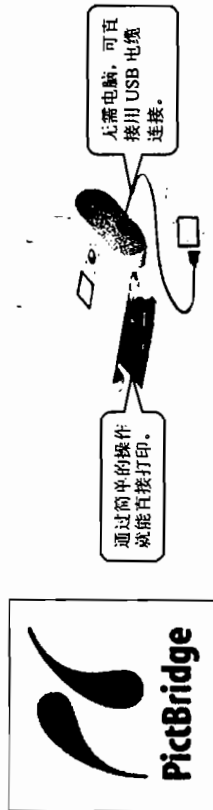


图8-30 PictBridge打印机的标志和工作原理

PictBridge标准的建立,使得各种品牌的数码相机与打印机可以相互兼容,不需要计算机和安装软件就可以打印成照片。今后只要用户的数码相机和打印机符合PictBridge标准,就可以免除插拔储存卡、计算机处理、控制输出等复杂的操作。2003年9月份以后推出的数码相机产品,基本上都会提供对此标准的支持,其中Canon和Sony作为此标准的发起者之一,具备更好的支持能力。

PictBridge通过USB接口进行物理连接,由打印机生成打印数据。数码相机向打印机发送图像数据和打印条件。打印机收到这些信息后,生成所要打印的图像,并进行打印。

蓝牙也正在成为高端便携式照片打印机相发展的接口。毫无疑问这一需求是为手机所定制的,随着摄像头手机的日益普及,厂商看到了甚至比相机市场更大的用户群体。



1. 列举几种常见的输入和输出设备。
2. 鼠标器的主要功能是什么?目前常见的是哪种类型的鼠标器?
3. 试述触摸屏的分类及其工作原理。
4. 简述扫描仪的工作原理。
5. 简述计算机显示系统的构成。
6. 简述CRT显示器的工作原理,彩色和黑白CRT显示器的主要区别是什么?
7. 简述LCD显示器的工作原理。
8. CRT显示器的主要性能指标有哪些?
9. LCD显示器的主要性能指标有哪些?
10. VGA适配器由哪几部分组成?VGA接口中有哪些信号?
11. 简述DVI接口的特点。
12. 显示适配器所支持的颜色与分辨率分别由哪几个因素决定?
13. 显示模式为“1024×768像素1670万色”时,按理论计算,显示适配器上至少需要多少显示内存?实际至少应配置多少内存?
14. 简述声卡的特性与功能。
15. 什么是采样频率和采样精度?如何决定数字声音的品质?

16. 声卡上的主要输入/输出接口有哪些?各自具备什么功能?
17. 声卡总线接口有哪一种?目前一般采用哪种接口?
18. 常见的声音格式有哪一种?
19. 通过分析针式打印机、喷墨打印机和激光打印机的工作原理,列出三种打印机的优缺点。



多媒体设备 第9章

9.1 多媒体的基本概念

媒体是信息的载体。多媒体——Multimedia 是由 multiple 和 media (medium 的复数形式)复合而成,核心词是媒体。媒体(media)在计算机领域有两种含义:一是指存储信息的实体,如磁盘、光盘、磁带、半导体存储器等,中文常译为介质;二是指传递信息的载体,如数字、文字、声音、图形和图像等,中文译作媒介。多媒体技术中的媒体是指后者。人类在信息交流中要使用各种信息载体,多媒体就是指多种信息载体的表现形式和传递方式。

9.1.1 多媒体技术的定义

多媒体技术的定义有多种,但基本概念是一致的。2001年国际电信联盟(ITU)的多媒体定义为:使用计算机交互技术、数字通信网络技术处理多种表示媒体——文本、图形、图像和声音,使多种信息建立逻辑连接,集成为一个交互式系统。国内有学者则定义为:以计算机为平台将数据、文字、图像、图形、视频和声音等处理技术结合在一起,构成生动而有效的信息系统,或将计算机系统中文本、图形、图像、声音、视频等多种信息媒体综合于一体进行编排处理的技术。

综合上述定义,并结合当今多媒体技术网络化、智能化以及与艺术紧密结合的发展趋势,可以将多媒体技术定义为:是以数字技术为基础,把通信技术、广播技术和计算机技术融于一体,对文字、图形、图像、声音、视频等多种媒体信息进行存储、传输、处理和控制,在不同媒体间建立逻辑连接,集成为一个具有交互性的系统,以提供丰富的生动的艺术表现来改善人们使用媒体体验的一种综合性的信息技术。简单地说,就是计算机综合处理声、文、图信息,并具有集成性和交互性。

多媒体技术将不同的媒体数据表示成统一的结构码流,然后对其进行变换、重组和分析处理,以进行进一步的存储、传送、输出和交互控制。所以,多媒体的传统关键技术

主要为以下四种:数据压缩技术、大规模集成电路制造技术、大容量光盘存储器、实时多任务操作系统。其中包括了多种技术:媒体处理与编码技术、多媒体系统技术、多媒体信息组织与管理技术、多媒体通信技术、多媒体人机接口与虚拟现实技术、多媒体应用技术、多媒体同步技术、多媒体操作系统技术、多媒体中间件技术、多媒体交换技术、多媒体数据库技术、超媒体技术、基于内容的检索技术、多媒体通信中的QoS管理技术、多媒体会议系统技术、多媒体视频点播与交互电视技术、虚拟现实空间技术等。

20世纪80年代中后期开始,多媒体技术成为人们关注的热点之一,是一种迅速发展的综合性电子信息技术,它给传统的计算机系统、音频和视频设备带来了方向性的变革,对大众传媒产生了深远的影响。多媒体计算机加速了计算机进入家庭和社会各个方面的进程,给人们的工作、生活和娱乐带来了一场深刻的革命。

9.1.2 多媒体技术的主要特征

多媒体技术具有以下的主要特征。

1. 集成性

多媒体技术不仅集成了多种媒体,而且集成了多种技术,包括计算机技术、通信技术、电视技术和其他音像处理技术。通过同一个文件中把来自多个通道的信息统一获取、组织、存储和合成,使得文字、图像、声音等各种媒体信息能在播放时同步地作用于我们的听觉、视觉等感官,从而取得最佳的效果。

2. 交互性

这是多媒体技术的关键特征之一。在多媒体系统中,用户可以借助交互活动控制信息的传播,甚至参与信息的组织过程,使之对感兴趣的画面或内容进行记录或者专门的研究。

3. 数字化

多媒体技术是建立在计算机基础上的,而计算机只能识别0、1组成的二进制数据。所以,在多媒体系统中,所有的多媒体信息都是用数字信号表示的。

4. 实时性

随着多媒体技术的进步,多媒体系统已经具备对多媒体信息进行实时处理的能力。具体应用如可视电话、电视会议、远程医疗等,使得千里之外的人物与场景犹如近在咫尺。

5. 多维性

多媒体技术具有信息处理范围的空间扩展和放大能力。它能将输入的信息变换、加工,增强输出信息的表现能力,丰富显示效果。这种信息空间的多维性使得信息的表达方式有声有色,生动逼真。

6. 非线性

线性为传统传递信息的模式,其读写方式大都采用章节页的框架,学习知识循序渐进。多媒体的非线性特点表现在多媒体技术借助超链接的方法,把内容以一种灵活的、多变的方式表达出来,使得用户可以按照自己的需要、兴趣、任务要求和认知特点来使用信息,选取图、文、声等信息表达形式,这改变了人们传统的读写模式。

9.1.3 多媒体系统

一般的多媒体系统由如下4个部分组成:多媒体硬件系统、多媒体操作系统、媒体处理系统工具和用户应用软件。

1. 多媒体硬件系统

多媒体硬件系统包括多媒体计算机,如多媒体个人计算机(MPC)、笔记本计算机和多媒体工作站;多种媒体输入/输出设备及信号转换装置、通信传输设备及接口装置等,如扫描仪、视频卡、音箱、光盘刻录播放器、触摸屏、数码相机、数码相机顶盒、无线远端控制产品等。其中,最重要的是根据多媒体技术标准研制的多媒体信息处理芯片和板卡、光盘驱动器。

2. 多媒体操作系统

多媒体操作系统也称为多媒体核心系统,具有实时任务调度、多媒体数据转换和同步控制、多媒体设备的驱动和控制以及图形用户界面管理等。

3. 媒体处理系统工具

媒体处理系统工具也称为多媒体系统开发工具软件,是多媒体系统的重要组成部分。

4. 用户应用软件

用户应用软件是指根据多媒体系统终端用户要求而定制的应用软件或面向某一领域的用户应用软件系统,是面向广大用户的系统产品。

9.1.4 多媒体个人计算机

所谓多媒体个人计算机(Multimedia Personal Computer, MPC)就是具有多媒体处理功能的个人计算机(PC),它的硬件结构与一般所用的计算机并无太大的差别,只不过是多了一些软硬件配置而已,如图9-1所示。一般来说,目前计算机的配置已基本符合多媒体计算机的硬件要求,还可根据需要进行增添一些其他设备,如摄像头、数码相机、扫描仪、MIDI设备、触摸屏、绘图板、光笔、手写输入板、投影机、打印机等人机交互设备;并增加一些专用扩展卡,包括音频声卡、多功能显示卡、视频卡(可细分为视频捕捉卡、视频处理卡、视频播放卡以及TV编码器等)。

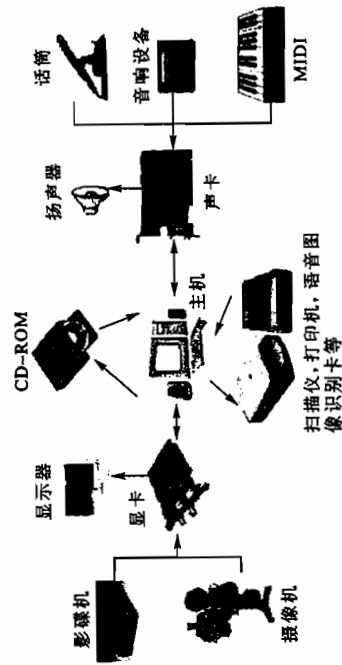


图9-1 多媒体计算机示意图

有些多媒体设备(如光盘驱动器、触摸屏、扫描仪、普通声卡、音箱、打印机等)已在前几章中做了介绍。下面简要介绍一些其他常用的多媒体设备,包括音频卡、摄像头、多功能显示卡、视频卡、屏幕采集卡、VGA 远程传输设备、视频分配器、视频切换器和投影机。

9.2 音频卡

音频卡,即专业声卡,与普通声卡的基本功能和工作原理是相似的。所谓的专业声卡是指为录音和音乐制作而设计的声卡,而普通声卡则是为娱乐、欣赏而设计的。普通声卡可以满足一般的声音处理需求,如需要更高质量的声音处理就要借助专业声卡了。专业声卡与普通声卡的主要区别在以下几个方面:采样频率与采样精度、输入输出接口、声卡芯片以及安装方式等。

9.2.1 采样频率与采样精度

数字-模拟和模拟-数字转换器是声卡的关键部分。普通声卡使用符合 AC'97 1. X 标准的 CODEC(数字信号编解码器, DAC 和 ADC 的结合体)。由于 AC'97 1. X 标准定义了输入输出的采样频率都是 48 kHz,所以如果 CODEC 接收到其他采样频率的音频流,便会经过采样频率转换器(Sample Rate Converter, SRC),将采样频率转换到统一的 48 kHz,在这个转换过程中,音频流中的数据便会由于转换算法而损失一部分细节,造成音质的损失,所以 AC'97 除了播放 48 kHz 的音频流音质还不错以外,播放其他采样频率的音频流时都不能得到很好的回放音质。

普通声卡的采样精度与采样频率一般最高达 16 位/48 kHz,而专业声卡目前可达 24 位/192 kHz。由于机箱内的电磁干扰较为严重,很多专业的音乐人,为了保证更好的音质,根本不使用声卡上的模拟-数字与数字-模拟转换芯片,而是购买专门的模拟-数字与数字-模拟转换器来使用;有一些高档的声卡将模拟-数字与数字-模拟转换器外置,如图 9-2 所示。声卡通过数字接口与外置模拟-数字与数字-模拟转换器连接。

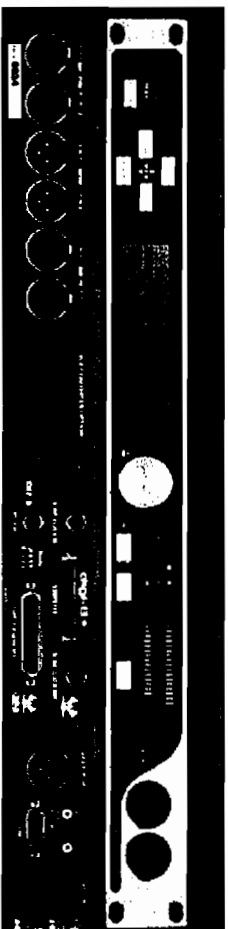


图 9-2 Apogee 的专业模拟-数字与数字-模拟转换器

9.2.2 输入输出接口

输入输出接口,俗称“几进几出”,如果说这个卡是两出,那就是左声道一个输出,右声道一个输出。一般来说,目前的所有声卡至少具备两进两出。声卡的输入输出接口当然是越多越好。多个输入端口的声卡意味着可以同时录制多轨,而多个输出端口的声卡可以同时提供多路输出

声源。

普通声卡一般都是双声道输入输出,使用 3.5 mm 的立体声接口就可以满足大多数情况的需要。自从 Direct Sound 3D 和 Dolby Digital/DTS 等多声道技术的出现以后,专业声卡后挡板的面积不足以容纳这么多接口了,所以,一些声卡厂商各自推出了非国际标准的多声道接口。一般使用转接线的方式提供输入输出端口,再多的,则使用外置接线盒,如图 9-3 所示。



图 9-3 多输入输出端口的连接方式

声卡的输出端口的驱动方式有两种,一种是平衡式,另一种是非平衡式。平衡式接口的抗干扰能力强,噪声低,明显好于非平衡式的接口。不过,在连接线长度不超过 10m 的情况下,两者并没有太明显的区别。

1. 模拟音频输入输出接口

非平衡接口一般使用 RCA 接头,也称为电话插头,也就是我们俗称的“莲花头”,而非平衡接口则使用 XLR 卡农(Cannon)接头,分别如图 9-4(a)和图 9-4(b)所示。

2. 数字音频输入输出接口

专业声卡基本都带有数字输入输出接口。数字接口的好处显而易见,它可以直接连接其他具有数字输入输出接口的设备,如数字调音台、带有数字输出的音源、合成器等。由于采用的是数字信号,音质几乎没有任何损失,而且数字信号的抗干扰能力远大于模拟信号。数字接口也有



(a) RCA 接头

(b) XLR 卡农接头

图 9-4 RCA 接头和 XLR 接头

平衡和非平衡之分。一般的专业声卡都有 SPDIF (Sony/Philips Digital Interface, 非平衡数字接口, 通常使用 RCA 或 BNC 通过同轴电缆连接, 也可使用光纤, 如 Toslink) 如图 9-5(a) 所示。而一些更专业的声卡, 则是使用 AES/EBU (Audio Engineering Society/European Broadcast Union) 接口, 属于平衡接口, 通常使用 XLR 接口。在某些情况下 AES/EBU 的专业设备和 SPDIF 的用户设备可以利用简单的平衡-非平衡转换器直接连接, 用一个小型脉冲变压器制成的专用转换器效果会更好些。但是一般不推荐这样的“跨协议”连接。很多专业声卡还具备一种专业的接口: ADAT 光纤接口, 如图 9-5(b) 所示。ADAT 是美国 Alesis 公司最早开发的 ADAT 数字多轨录音机接口规格, 它可以用一条光纤同时传送 8 路数字音频信号。专业声卡上的 ADAT 光纤接口可以用来连接数字录音机以及独立的模拟-数字转换和数字-模拟转换设备。



(a) SPDIF 接口

(b) ADAT 接口

图 9-5 SPDIF 接口和 ADAT 接口

3. SPDIF

SPDIF 是数字接口, 需要相应的数字化设备用于连接声卡。如果有数字解码器、数字功放或者带有数字音频解码的音箱, 就可以使用 SPDIF 接口进行数码音频输出。使用外置的 DAC 进行解码, 可以达到更好的音质。

SPDIF 接口一般有两种, 一种是 RCA 同轴接口, 另一种是 TOSLINK 光缆接口, 其中 RCA 接

口是非标准的。在国际标准中, SPDIF 需要采用 BNC 接口和 75 Ω 传输电缆, 然而很多厂商由于各种原因通常使用 RCA 接口, 甚至使用 3.5 mm 的小型立体声接口进行 SPDIF 传输, 久而久之 RCA 和 3.5 mm 接口就成为一个“民间标准”了。

多声道声卡连接多声道音箱的方法有很多种, 但主要分为两种: 一种是多声道声卡使用多声道模拟信号接口连接多声道扬声器, 这时候, 声卡传输到扬声器的信号完全是模拟类型的, 这种方法可以兼顾 3D 游戏和 DVD 电影。另一种是采用 SPDIF 数字信号传输经过编码的多声道信号 (Dolby Digital/DTS), 这需要多声道音箱具有 Dolby Digital/DTS 解码功能, 这时候, 声卡与扬声器只需要一条数字信号线连接即可。RCA 同轴电缆/3.5 mm 转接线/光缆都可以用于 SPDIF 信号的传输。除了经过多声道编码的音频信号能够被正确还原至多声道音箱以外, 你听到的其他所有声音都只有双声道, 因为这时声卡输出的只有一路 PCM SPDIF (未经压缩的 SPDIF, 仅仅支持双声道) 数字信号, 而非经过编码的多声道数字信号。

4. MIDI 接口

一般专业声卡都具备至少有 1 进 1 出的 MIDI 接口, 即 MIDI IN 和 MIDI OUT 接口。有些专业声卡还带有 MIDI THRU 接口, 此接口的功能是将 MIDI IN 输入的信号再原样地转发给另外一台设备。与普通声卡采用 D 型插座不同, 专业声卡的中 MIDI IN 和 MIDI OUT 接口一般分别采用一个与键盘/鼠标相同的 DIN 插座。图 9-6 所示是专业声卡的一些常用接口。

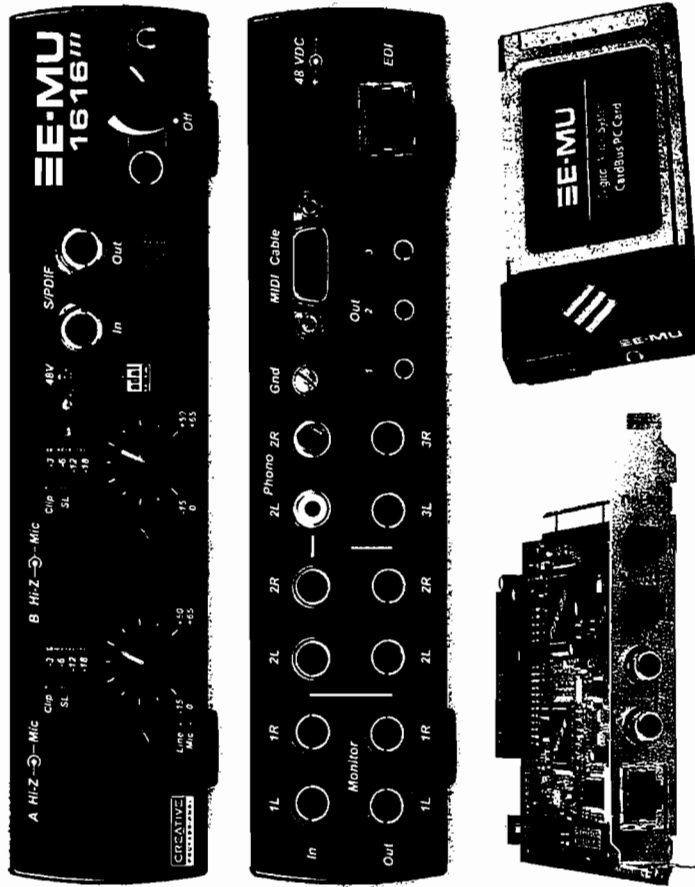


图 9-6 专业声卡的接口

9.2.3 声卡芯片

声卡芯片是声卡的核心部件,它决定了声卡的处理能力、音效、档次与价格。选购声卡芯片要注意,同一种芯片可能会有不同的版本,而价格与效果也会存在差异,像 CM18738 就有 6 声道和 4 声道等版本,选购时一定要了解清楚。

音质是判定一块声卡好坏的重要标准,其技术指标包括信噪比、采样位数、采样频率、总谐波失真等,这些参数都不同程度地影响了声卡的音质效果。我们知道,声卡处理音频信号时会出现背景静电、工作电流等噪声,有用信号功率与噪声信号功率的比值就是信噪比(Signal-to-Noise Ratio, SNR),它关系到播放的声音是否纯净。就目前的声卡而言,一般都能达到 96 dB 的信噪比,基本上不会出现明显的噪声。声卡的保真度也是非常重要的一项指标,其单位为分贝(dB),数值越低说明声卡的失真度越大,性能越差,因此,要尽量选择保真度较高的产品。采样频率是声卡对声音的采集能力的重要指标,指的是声卡每秒钟采集信号的次数,一般来说,也是频率越高越好,当它达到 48 kHz 时,音质就可以和 DVD 媲美了。不过有一点需要注意的是,不仅需要有一双专业的耳朵,还要选用质量档次较高的音箱,才能更好地测试出音质的水平。

好的回放音效,除了音质要好,芯片功能要强之外,最重要的是声卡还必须具备优秀的 3D 音效,它的性能会直接影响到声卡的音效表现力。常见的有 EAX、A3D、DirectSound 3D 等 3D 音频技术(API 规范),它们主要的特点如下。

(1) Creative 的环境音频扩展集(Environmental Audio Extensions, EAX)

Creative 公司的 EAX 最为出色,它是 Creative 公司推出的环境音频扩展开放性 API,着重于 3D 环境音效,特别是多声道效果十分突出,但对声音的定位能力不如 A3D。

(2) 帝盟的 A3D 和 A3D Surround

帝盟公司的 A3D 可通过双通道实现较好的 3D 效果,定位感不错,目前已发展到 A3D 3.0 版。A3D 是 Aureal 公司开发的一项 3D 音频技术,这一技术可以由用户决定音效的变化,从而达到用户想要的 3D 环绕效果。A3D Surround 是在 A3D 的基础之上加强了对声音定位精确性。它可以只通过两个音箱在三维空间中表现出 5 音箱的环绕效果,也可以叫做虚拟音箱。这一技术已经被杜比(Dolby)实验室授予 Virtual Dolby 认证。

(3) Microsoft 的 DirectSound 3D

DirectSound 3D 是 Microsoft 推出的音频 API 规范,借助 Windows 系统有了统一的接口和极好的兼容性,目前的 PCI 声卡几乎都支持这一技术。从它的名字可以看出,DirectSound 3D 是 Microsoft 所制定的 DirectX 音频 API 规范。但是 DirectSound 3D 仅仅是一个 API,并不能直接影响到声卡音效的好坏,而是要看声卡自身采用的 HRTF 算法能力来解决音效的高低。

3D 音频技术都是基于 HRTF(头部相关听觉传递函数)的原理,因此,只用两只音箱是无法达到三维的听觉效果的,这点在玩极品飞车游戏时玩家就有很大的感受了。近年兴起的多声道显然具有更好的环境定位能力,更能给人以身临其境的真实感受。

9.2.4 外置声卡

目前,除了装在机箱里的 PCI 声卡之外,专业声卡中还有一类外置声卡,也可称为音频接口。外置声卡目前分为 USB 2.0 接口和 IEEE-1394(火线)接口两种。火线接口的声卡比 USB 接口

的要更加稳定,传输速率也更高,很受音乐人的欢迎。其实外置声卡已经根本不能称之为“卡”了,如图 9-7 所示。

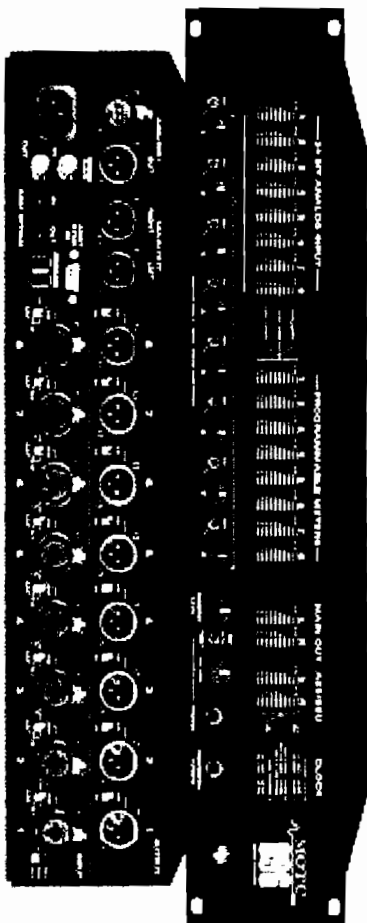


图 9-7 MOTU 896 外置火线声卡

外置声卡最大的优点就是可移动性好,接线方便,因此很受笔记本电脑用户的欢迎。移动录音作业,比如现场录音之类的工作,使用一款外置火线音频接口,是最好的选择。

在专业声卡中还有一类是专门为一些软件所设计的,而这些软件也必须要有这个声卡才能运行,比如 ProTools 系列(见图 9-8)就是如此,如果没有专门的卡,就无法使用这些软件。

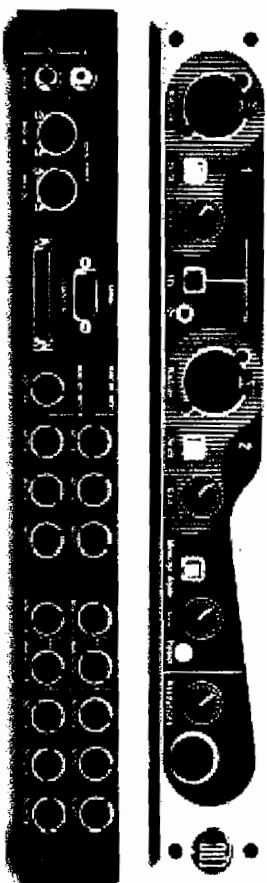


图 9-8 为 ProTools 设计的 Digidesign Dig1001 声卡

还有一类多功能的外置声卡,它们不仅仅包含了声卡的功能,同时还具备其他功能。比如 Yamaha 01X(图 9-9(a)),它既是一个火线接口声卡,本身又相当于一个数字调音台,还相当于软件控制台,还自带效果器。这类产品使用起来无疑更加方便,移动性也比较好;Novation X-Station 25(图 9-9(b))集成成器、声卡、MIDI 键盘、MIDI 控制器、MIDI 接口、效果器等于一身,一台笔记本电脑再加上它,就什么都有了,很适合外出携带。

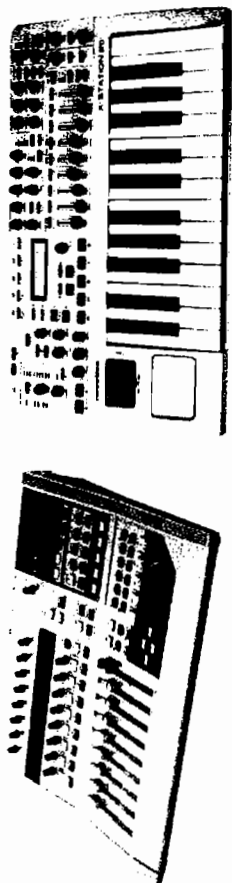


图 9-9 多功能外置声卡

(a) Yamaha 01X

(b) Novation X - Station25

9.2.5 与专业声卡相关的一些概念

1. AC'97 规范

1996年,以Intel为首的5个PC厂商: Intel、Creative Labs、NS、Analog Device与Yamaha,共同提出了Audio Codec'97(意为“音频多媒体数字信号编/解码器”,简称AC'97)的规范。其主要的创新“双芯片”结构,使ADC与DAC可以脱离系统,这样就可以避免数字/模拟信号转换时所产生的大部分杂波,从而得到较好的音质。

其实目前市面上所有的声卡都支持AC'97规范。到目前为止,AC'97软声卡已经发展到8声道,号称“HD-Audio”。其间AC'97总共历经4次大的修改:

- ① AC'97 1.x: 固定的48 kHz采样输出。
- ② AC'97 2.1: 扩展了部分音频特征,开始支持多种采样频率输出及多声道输出。
- ③ AC'97 2.2: 完善和扩展了部分音频特征,开始支持SPDIF输出。
- ④ HD-Audio: Intel最新ICH6支持的HD-Audio,用于取代AC'97标准,支持最高7.1声道音频输出,并拥有32位/192 kHz的高指标。

2. 声卡的输出声道数和内部通道数

按照输出声道数分类,一般可以分为双声道声卡、4声道声卡、5.1声道声卡、6.1声道声卡和7.1声道声卡。5.1以上的音箱要求与相应的声卡相配,其中的“.1”是实际存在的(如5.1声卡即为“6声道”声卡)。而2.1音箱和4.1音箱中的“.1”,在声卡中并不存在,而是音箱从各个声道中提取低频部分的信号产生的。2.1音箱使用的声卡是双声道声卡,4.1声道音箱使用的声卡是4声道声卡。

声卡的内部通道与声卡的输出通道不是一回事,它主要是构建灵活的内部音频流环境。比如,多个音频通道可以支持用户运行多个音频软件而互不干扰。专业声卡都具有多个内部通道。

3. 声卡的内录

专业声卡还有一个功能就是内录。有相当一部分专业声卡并不支持无损内录,也有一部分

声卡支持。其实,现在的音乐软件都是使用直接运算导出音频,包括GigaStudio使用Rewire模式也是直接导出音频的,已经根本不需要内录。内录方便、直接,也是有一定优点的。目前不少声卡都支持内部虚拟连线,完美地支持无损内录。

4. 声卡的驱动程序

驱动程序是声卡的灵魂之所在。现在的专业声卡的所谓“驱动程序”,往往并不只是一个驱动程序,还带有声卡本身的软件界面,提供强大的功能。专业声卡往往都支持专业标准驱动程序,比如,ASIO是所有专业声卡都支持的标准,另一个得到广泛支持的驱动程序是GSIF。所谓GSIF,是指声卡是否支持著名的采样器软件GigaStudio。但现在的GigaStudio已经不再需要GSIF,而是使用Rewire模式来运行了,因此也有一些专业声卡并不支持GSIF。

9.3 摄像头

摄像头也就是我们常说的数码摄像头、网络摄像头。如果按原理分,摄像头可分为模拟摄像头和数字摄像头两大类。

模拟摄像头是指那种可以直接通过视频接口(通常为S端子或AV端子)连接显示设备(一般指电视机或监视器)完成摄像功能的摄像头,也可经过特定的视频捕捉卡将模拟视频信号转换成数字模式,并加以压缩后才可以传输到计算机上运用。特点是模拟影像清晰而连贯,不受分辨率影响,模拟摄像头以中低价位黑白摄像头为主。

数字摄像头可以直接捕捉影像并转换为数字信号存储在计算机里。其信号传输接口发展由早期的串行接口、并行接口发展到如今的USB 2.0和IEEE-1394(火线)接口。除此之外,还有一种与视频采集卡配合使用的产品,但目前还不是主流。

模拟摄像头主要出现在家电市场上,而在计算机市场上销售的摄像头都是数字摄像头。摄像头主要由核心IC控制芯片(包括USB接口和图像处理功能)、图像传感器(CMOS或CCD感光器件)、镜头三大核心部件及其周围元器件组成。

摄像头的工作原理如图9-10所示。景物通过镜头生成的光学图像投射到图像传感器表面后转换为电信号,经过A/D(模拟/数字)转换后变为数字图像信号,送到数字信号处理芯片中加工处理,然后再通过USB接口传输到计算机中处理,最后通过显示器就可以看到图像了。

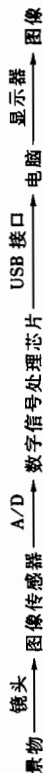


图 9-10 摄像头的工作原理

9.3.1 图像传感器

图像传感器是摄像头中最重要部件之一,主要有CCD和CMOS两种。CCD(Charge Coupled Device,电荷耦合器件)由一种高感光度的半导体材料制成,能把光线转变成电荷。CMOS(Complementary Metal-Oxide Semiconductor,互补金属氧化物半导体)是采用硅和锗这两种元素制成的

半导体,其中共存在着N(带负电)和P(带正电)极的半导体,这两个互补电极所产生的电流即可被处理芯片纪录和解读成影像。在相同像素情况下,图像传感器的面积越大越好,这种差异在弱光情况下表现得尤其明显。

一般来说,CMOS摄像头反应快、能耗低,对光源有一定依赖性,是主流大众级产品;CCD摄像头反应稍慢、耗电稍大、价格比较高,但在动态范围、分辨率、噪声控制等方面都优于CMOS传感器,CCD摄像头成像质量更好,图像明锐通透、细节丰富,色彩还原度好,曝光准确。

目前CCD产品因具备成像质量更好的优势,但限于成本、制造工艺更复杂(只有Sony、Fujifilm、Panasonic等少数几个厂商掌握这种技术)以及摄像头的功用,市售的CCD摄像头产品不仅在价格上非常昂贵,而且实际成像质量也很一般,明显比不上一些做工稍好、采用优质DSP芯片+CMOS的产品。所以,限于目前的技术水平,由于CCD摄像头产品性价比不高,已被完全淘汰出目前的摄像头领域,CMOS产品已成为绝对主流。

目前,市场销售的数码摄像头中,基本是以CMOS产品为主。通过一系列新技术以及DSP主控芯片技术的进步,采用CMOS为感光元件器件的摄像头,通过采用影像光源自动增益补偿技术、自动亮度与白平衡控制技术、色饱和度和对比度、边缘增强以及伽玛校正等先进的影像控制技术,已完全可以达到与CCD摄像头相媲美的效果。

9.3.2 数字信号处理器

数字信号处理器(DSP)的作用是将感光芯片获取的数据及时快速地传到计算机中并刷新感光芯片,因此控制芯片的好坏,直接决定画面品质(比如色彩饱和度和清晰度)与流畅度。是否有不断升级的驱动程序也比较重要,新驱动程序可以赋予其更强大的功能和使用效果。

DSP主要包括以下几个功能模块:图像信号处理器(Image Signal Processor,ISP)、JPEG图像解码器(JPEG encoder)和USB设备控制器(USB device controller)。ISP的性能是决定影像流畅的关键,JPEG图像解码器的性能也是关键指标之一,比如中星微(VI)301P是硬件JPEG压缩方式,而松翰602采用软件RGB压缩方式,因此301P图像处理速度要比SONIX 602A快一些。

9.3.3 镜头

镜头是仅次于CMOS芯片的、影响画质的第二要素,由几片透镜组成,一般可分为塑料透镜(plastic)和玻璃透镜(glass)。当然,所谓塑料透镜也并非纯粹塑料,而是树脂镜片,其透光率等光学指标不如上镀膜玻璃镜片。此外,我们注意到最外部的镜头曲率较大,类似于“鱼眼”,这种设计目的是提高镜头广角视野,不过随之而来的问题是边缘几何变形,也算摄像头类产品的通病了。

通常摄像头用的镜头构造有:1P、2P、1G1P、1G2P、2G2P、2G3P、4G、5G等(P—plastic,G—glass)。透镜越多,成本越高,相对成像效果会更出色;而玻璃透镜又比树脂贵。因此一个品质好的摄像头应该是采用多层玻璃镜头!现在市场上的多数摄像头产品为了降低成本,一般会采用廉价的塑料镜头或一玻一塑镜头(如1P、2P、1G1P、1G2P等)。好的全玻璃镜头都有较大的光圈,可达到F2.0甚至F1.8,直径为6mm,并有多层镀膜或加滤光片,以增加通光量、减少反光,使成像清晰,画质更鲜艳。

商家在销售摄像头时,都会说自己的产品是真正加滤光片5玻璃镜头。由于目前还没有一

种方法能通过肉眼准确判断镜头材质的方法,消费者只能根据实际效果来评价摄像头的品质。

9.3.4 摄像头的主要技术指标

1. 图像分辨率

图像分辨率(Resolution)与第8章所述的显示卡的分辨率相似,主要有以下几种:

- SXGA(1280×1024像素),又称130万像素。
- XGA(1024×768像素),又称80万像素。
- SVGA(800×600像素),又称50万像素。
- VGA(640×480像素),又称30万像素(35万像素是指648×488像素)。
- CIF(352×288像素),又称10万像素。
- SIF/QVGA(320×240像素)。
- QCIF(176×144像素)。
- QSIF/QVGA(160×120像素)。

有些产品上标注的分辨率是指这些产品通过软件插值计算后所能达到的分辨率,虽然插值计算也能适当提高所得图像的清晰度,但和硬件分辨率相比还是有着一一定的差距的。

2. 图像格式(Image Format/Color space)

RGB24和I420是目前最常用的两种图像格式:

- RGB24:表示R、G、B三种颜色各8位,可以再现256×256×256种颜色。
- I420:色差分量YUV格式之一。

其他格式有:RGB565、RGB444、色差分量YUV4:2:2等。

3. 自动白平衡调节(AWB)

自动白平衡调节(AWB)要求在不同色温环境下,照白色的物体,屏幕中的图像应也是白色的。色温表示光谱成分,即光的颜色,色温低表示长波光成分多。当色温改变时,光源中三基色(红、绿、蓝)的比例会发生变化,需要调节三基色的比例来达到色彩的平衡,这就是白平衡调节的实际作用。

4. 图像压缩方式

目前多数使用JPEG格式进行图像的压缩,这是一种有损的压缩格式。压缩比越大,图像质量也就越差。当对图像清晰度要求不高且存储空间有限时,可以选择这种格式。

5. 色彩深度

色彩深度(色彩位数)反映对色彩的识别能力和成像的色彩表现能力,主要取决于A/D转换器的量化精度,是指将信号分成多少个等级。常用色彩位(bit)数表示。色彩深度越高,获得的影像色彩就越艳丽动人。

6. 图像噪声

图像噪声指的是图像中的杂点干扰,表现为图像中有彩色杂点。

7. 视角

视角简单说就是成像范围。

8. 输出接口

摄像头的输出接口主要有:

- ① 串行接口(RS-232C/422):传输速率较低,为115 Kb/s。
- ② 并行接口(EPP):传输速率可以达到1 Mb/s。
- ③ 红外接口(IRD):传输速率也是115 Kb/s,笔记本计算机上一般配有这种接口。
- ④ 通用串行总线(USB)接口:USB的传输1.1速率可达12 Mb/s,USB 2.0可达480 Mb/s。
- ⑤ IEEE-1394:即火线接口,亦称iLink,其传输速率可达100 Mb/s~400 Mb/s。

9.4 多功能显示卡

随着技术的发展,消费者对显示卡(即显示适配器,也可简称为显卡)的功能亦愈加愈多,于是出现 VIVO (Video In & Video Out),即显示卡同时具备了显示、视频输入与视频捕捉等功能。

9.4.1 多功能显示卡的发展

IBM PC 上市之初,具有图形显示能力的是 CGA 显示卡,其所能支持的分辨率只有 $320 \times 200 \times 4$ 像素或 $640 \times 200 \times 2$ 像素。以后 Hercules(大力神)公司研发出了 MGA 显示卡,可提供细腻的单色文字与图形显示能力。由于 MGA 除了本身具有 720×348 像素的分辨率,也支持 CGA 的分辨率,所以许多早期为 CGA 设计的计算机游戏,绝大多数都可以在 MGA 下执行,使得后来的 286、386 甚至 486 的入门级计算机,大多还是配置 MGA 显示卡与单色显示器。在 1980 年代或者说 DOS 时代,MGA 显示卡最普及的显示卡。

后来有些厂商推出了把打印机端口整合进 MGA 卡的 MCP 显示卡,将当时把最常使用到的显示与打印功能整合在一张卡上,不仅节省另外购买 Multi-I/O 适配卡的预算,也节省一个 ISA 插槽。MCP 可以说是多功能显示卡的一个前身。

当 1984 年 IBM 推出 IBM PC/AT 时配置了 EGA 显示卡,可以提供 640×350 像素、16 色的分辨率,色彩呈现能力比 CGA 有所提升,图像也细腻多了。随之国内的一些公司也在此时推出含字形 ROM 的显示卡(汉卡),提供了彩色中文显示功能,这种显示卡算是有附加功能的显示卡。不过由于当时的 EGA 显示卡与显示器的价格过高,并不普及。后来,IBM 在 1987 年推出 PS/2 计算机,采用的是 VGA 的显示标准。VGA 除了颜色数提升到 256 色以外,分辨率也提高到 640×480 像素,到此真正提供彩色显示功能的显示卡出现了。

由于当时计算机的普及率不高,制定新的显示标准的速度并不高。一些显示卡厂商推出了采用 SVGA 标准的显示卡,以更高的分辨率(800×600 像素、1 024 \times 768 像素)、更多的色彩数(32K/64K 色甚至 1677 万色的全彩能力)来吸引用户。但由于众多显示卡厂商采用的标准不同,没有统一标准的绘图模式,造成了软件厂商在应用软件开发、支持的困难,所以视频电子标准协会(Video Electronic Standard Associates, VESA)制定了相关的显示标准,统一确定了显示卡的分辨率和色彩标准,于是各显示卡厂商有了共同遵循的标准。有了统一标准后,软件的开发脚步加快,许多为彩色

计算机环境所开发的绘图、应用软件如雨后天晴般出现,这也加速了显示卡的发展。

Windows 3.x 使计算机进入了图形界面的时代。为满足大量且频繁的图形界面操作的需要,出现了 Windows 加速卡(Windows Accelerator)。最早推出 Windows 加速卡的厂商是 S3 公司。从 1991 年之后所推出的显示卡,几乎都具备 Windows 加速卡的架构,并提供 Windows 3.x 驱动程序。而此时 MPEG 制定的动态图像压缩标准,带动了 VCD 的流行。不过当时 CPU 的速度较慢,无法以软件来实时播放 MPEG 影片,厂商随即推出了可解压 VCD 的硬件 MPEG 解压卡。搭配这种卡,可以通过覆盖(overlay)的显示方式,将 VCD 影像画面呈现出来。由于硬件 MPEG 解压卡价格高昂,于是有显示卡厂商推出内置 MPEG 硬件解压功能的显示卡,搭配专属的 VCD/MPEG 播放软件,就能直接播放 VCD,这是具有真正意义的多媒体显示卡。

进入 Pentium MMX 时代后,CPU 对多媒体的处理能力有了一定提高,可以通过软件对 MPEG 进行实时解压,机器只要配置普通的显示卡就可以直接播放 MPEG。这时内置 MPEG 芯片的多媒体显示卡已经不能吸引用户们的目光了,于是厂商开始为显示卡开发另外的附加功能。针对用户想将电视节目或 V8 影片录到计算机上的需求,厂商推出视频采集卡,通过硬件 MPEG 压缩芯片,提供实时录制影片的功能,这样就可以录制自己的 VCD。不过以上这些都是针对不同需求所开发出来的独立产品,提供视频输入、采集、电视播放等功能。针对此类用户,有的显示卡厂商将上面的一些或全部功能,整合或附加到自己的显示卡中,这样就诞生了多功能显示卡。

ATI 在 1997 年率先推出了 All-in-Wonder 的多媒体显示卡,首度将电视视频、影音输入输出和采集等功能,全都整合在一张卡上,实现了以往多媒体计算机平台上必须插上显示卡、MPEG 解压卡、视频采集卡、TV-Tuner 卡才能达到的效果。All-in-Wonder 整合所有视频需求于一身,等于是为显示卡开辟了一个新的市场,目前 ATI 是这方面的领头者。

nVIDIA 根据其产品的特点开发了 PersonalCinema (NPC)如图 9-11 所示。NPC 是 nVIDIA 在影音多媒体方面和 ATI All-in-Wonder 系列相抗衡的产品。



图 9-11 NPC 的标志

9.4.2 多功能显示卡的常见功能与应用

不同厂商生产的多功能显示卡功能不尽相同,下面是一些常见的功能与应用。

1. 双屏显示

现在不少多功能显示卡具备了两个输出接口(一个 D-Sub 和一个 DVI 或者双 DVI),用户可以根据显示器的接口类型来进行连接,图 9-12 就是连接双屏的例子。此多功能显示卡为一个 D-Sub 接口和一个 DVI 接口,首先将主显示器连接到 D-Sub 接口上,再将副显示器通过 DVI-I 转 D-Sub 转接头连接到显示卡的 DVI-I 接口上。

连接好显示器后开机,在开机自检过程中,两个显示器都会开启,并且显示同样的内容,进入 Windows XP 操作系统后,其中一个显示器会自动关闭,这是因为 Windows XP 中双屏显示功能默认是关闭的,其中保持开启的显示器是接在主(Primary)显示接口上的,关闭的显示器是接在从(Slave)显示接口上的。

此时需要安装显示卡最新驱动程序才能设置双显示器输出参数,如 nVIDIA 显示卡利用 ForceWare 驱动程序中的 nView 来实现双显示器功能,而 ATI 显示卡则通过 HydraVision 技术来实现。以 ATI 显示卡为例,进入系统后,在显示属性设置里可以看到两台显示器(如图 9-13 所示),如果是第一次使用,第 2 台显示器是灰色的,表示没有启用。双击第 2 台显示器,系统会弹出对话框,询问是否启用该显示器,按提示即可完成设置。



图 9-12 双屏连接图

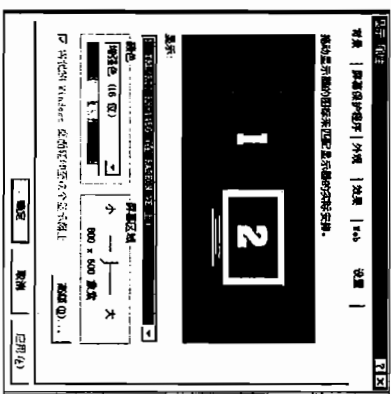


图 9-13 Windows XP 下的双屏设置

2. 接收电视

以前如果想在计算机上收看电视节目,一般的做法是安装电视卡。如果你有一块具有电视接收功能的多功能显示卡就不必多此一举了,只需将电视的信号接到多功能显示卡的射频接口上,就可利用采集程序来接收电视信号了。一般多功能显示卡的软件都提供自动搜台和录像(实时采集、压缩和存储)的功能,如图 9-14 所示。

3. TV-Out

TV-Out 功能可以直接将计算机的内容输出到电视机上,可以用来播放 VCD、DVD,也可以通



图 9-14 接收电视信号的高频头和自动搜台功能

过此功能将网上播放的电影转接到电视上与家人共赏。此外,如果使用的是高清数字电视,也可以通过 TV-Out 功能在电视上玩计算机游戏。很多显示卡采用 S-Video 视频输出接口(俗称 S 端子),此时电视机也必须具备 S 端子接口,如果没有这个接口,可以通过相关转接头进行连接。进入系统之后,可以对电视机的显示属性进行设置,如图 9-15 所示,来达到更好的视觉效果。

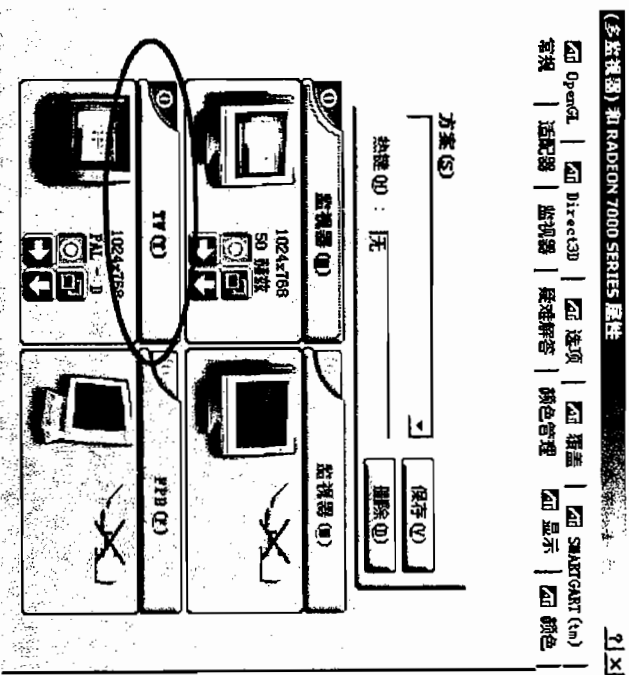


图 9-15 电视机显示属性的设置

4. VIVO-In/VIVO-Out

多数多功能显示卡具有 VIVO-Out 和 VIVO-In 接口。将 VIVO-Out 接口连接到录像机的 Video-In 接口,就可以在播放的同时用录像机进行录制;如将录像机的 Video-Out 接口连接到 VIVO-In 接口,就可将录像带上的影像采集下来,并通过视频捕捉编辑软件对影像文件进行编辑(例如剪辑、加入过场画面或加入字幕等),制作自己的影像文件,需要时还可以刻成 VCD。

多功能显示卡都会配送一些视频编辑软件,如 CyberLink PowerDirector、Video Studio 及 Win Producer 等。此外,WinFast 为了突出自身产品的易用性及特色,自行设计了自己的视频编辑程序——MyVIVO。最后根据需要编辑好视频文件,使用 Nero 等软件刻录成 VCD 光盘即可。

5. 采集和编辑

多功能显示卡在软件的支持下,都具有实时采集、动态影像捕捉、软件或硬件压缩/解压缩和存储等功能,可以直接对由电视、VCD 或摄像机等所传入的静/动态画面进行采集、压缩,并具备编辑功能。

9.5 视频卡

9.5.1 视频卡的基本概念

视频的定义很宽泛,包括摄像头、DV、电视卡、Web 摄像头等设备采集的画面、视频片断、网络视频等。同时视频卡具有一些基本的控制功能,比如亮度、对比度、色度、饱和度和分辨率等。

视频卡也叫视频采集卡,计算机通过视频卡可以对各种模拟视频信号进行采集并量化成数字信号,然后压缩编码成数字视频序列。大多数视频卡都具备硬件压缩的功能,在采集视频信号时首先在卡上对视频信号进行压缩,然后才通过 PCI 接口把压缩的视频数据传送到主机上。一般的视频采集卡采用帧内压缩的算法把数字化的视频存储成 AVI 文件,高档一些的视频采集卡还能直接把采集到的数字视频数据实时压缩成 MPEG 格式的文件或 ITU-T 视频编码 H.26X。

由于模拟视频输入端可以提供不间断的信息源,视频采集卡要采集模拟视频序列中的每帧图像,并在采集下一帧图像之前把这些数据传入计算机。因此,实现实时采集的关键是每一帧所需的处理时间。如果每帧视频图像的处理时间超过相邻两帧之间的间隔时间,则会出现数据的丢失,也即丢帧现象。采集卡都是把获取的视频序列先进行压缩处理,然后再存入硬盘,也就是说视频序列的获取和压缩是在一起完成的,免除了再次进行压缩处理的不便。不同档次的采集卡具有不同质量的采集压缩性能。

常见的压缩方式有硬件压缩和软件压缩,前者的压缩比一般不超过 1:6,后者的压缩比由软件而定。硬件压缩的优点就是不占用计算机资源,故较低配置的计算机也可以采集出好质量的视频文件,软件压缩则需要计算机有较高的配置。

德国 Miro 公司生产的 Studio Pctv 多功能视频卡是视频卡产品中的典型代表,它集电视卡、视频采集卡于一身,不但可以收看电视节目,而且可以将电视节目录制在计算机的硬盘中,有“硬盘录像机”的美称。此外,它还能够接收、录制在录像机、摄像机、影碟机中播放的内容。随

卡配有功能强大的 Studio 编辑软件,能实现 300 余种转场特技,再用上 TitleDeko 字幕软件,就可以编辑自己的图片及电影了。在影片编辑完成后,可以生成 MPEG 文件。

9.5.2 视频卡的基本类型

视频卡按其连接方式来划分,可以分为外置式和内置式两种。

内置式视频卡有以下优点:

- ① 价格便宜,一般 300 元左右,适合于一般用户的需要。
- ② 不会占用外部的桌面空间。现在在电脑桌上已经有许多外设(如音箱、打印机、调制解调器、扫描仪等),如果摆放得过多也是件比较烦人的事情。
- ③ 不需外接电源,不仅节省了一个电源插座的位置,也避免了电源连接过多而显得十分杂乱的情形,而且还可以避免因市电电压不稳而带来的视频质量下降的情况发生。
- ④ 大多数都可以在 Windows 系统下实现多任务,可以开一个小小的电视窗口,再同时打开 Word 继续工作,而大多数外置电视盒都只能在全屏状态下进行视频与 Windows 桌面的切换。

内置式视频卡有以下缺点:

- ① 易受到电脑内部元器件的电磁干扰,引起其播放质量下降。
- ② 安装时比较麻烦,必须拆开机箱才可以进行安装,且在安装软件时也容易和其他设备发生冲突。
- ③ 最致命的缺点是只能在打开计算机的前提下才可以收看电视节目,因为它需要计算机的电源来为其供电,不打开计算机就不能进行工作。

外置式视频卡的主要优点是显而易见。安装一般都比较简单,因受外界因素干扰的可能性较少,所以其性能比内置式视频卡稳定;通过其面板上的各种指示灯还可以对其运行状态了如指掌。此外,外置式视频卡大多可以在不打开计算机的情况下收看电视节目,只需要打开电视盒的开关和显示器即可;而且外置电视盒一般都配有类似普通电视所用的遥控器,这可以大大地提高操作的方便性。其主要缺点是价格要比同档的内置视频卡贵 200 元左右,且很多外置视频卡不支持窗口播放,因此不能在同一屏幕上进行多任务的操作;并且它还需要外接专门的电源和占用计算机上的一个接口。

图 9-16 所示为几种常见的视频卡。

9.5.3 视频卡的主要性能指标

1. 功能

现在视频卡功能的越来越多,也越来越完善。通常一个普通的视频卡只具有其中的几项功能,不必苛求高、新、全,只要实用、满足需求即可。

2. 视频信号源的制式

视频信号源的制式主要 PAL、NTSC 和 SECAM 三大类,一定要了解视频卡所支持的视频信号源制式。有些视频卡可以自动检测视频信号源的制式,使用比较方便。

3. 分辨率和压缩比

视频卡按其用途可以分为广播级、专业级和民用级,其区别主要是采集的图像指标不同。

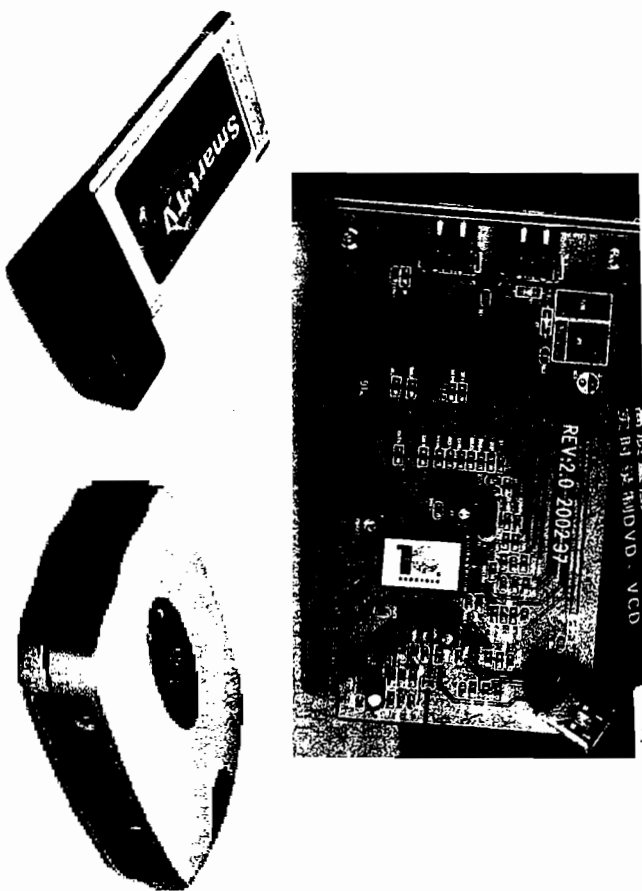


图 9-16 各类总线接口的视频卡

广播级视频卡的最高采集分辨率一般为 768×576 像素(均方根值)、PAL 制式、或 720×576 像素(CCR-601 值)、PAL 制式 25 帧每秒,或 640×480 像素/720 \times 480 像素、NTSC 制式 30 帧每秒,最小压缩比一般在 4:1 以内。这一类产品的特点是采集的图像分辨率高,视频信噪比高,缺点是视频文件庞大,每分钟数据量至少为 200 MB。广播级模拟信号采集卡都带分量输入输出接口,用来连接 BetaCam 摄录机,此类设备是视频采集卡中最高档的,用于电视台制作节目。

专业级视频卡比广播级视频卡的性能稍逊一些,分辨率相同,但压缩比稍高一些,其最小压缩比一般在 6:1 以内,输入输出接口为 AV 复合端子与 S 端子。此类产品适用于广告公司、多媒体公司制作节目及多媒体软件。

民用级视频卡的动态分辨率一般最大为 384×288 像素, PAL 制式 25 帧每秒。

另外有一类视频卡是比较特殊的,这就是 VCD 制作卡,从用途上来说它是应该算在专业级,而从图像指标上来说只能算是民用级产品。

4. 视频格式

如果视频卡主要是用来进行视频编辑处理的,那么就要注意采集和捕捉影像之后可以转存的视频格式。有的视频卡只能保存为 AVI 一种视频格式,并且没有影像压缩的功能。为了能适应多种格式的视频信息的编辑处理,最好选用可以保存为多种视频格式,且具有图像压缩功能的视频卡。

5. 视频接口

视频卡的视频接口目前主要有分量 YUV 接口、S 端子和 AV 复合端子。从传输的性能上来说,分量 YUV 接口最好, S 端子居中,而 AV 复合端子是最差的。具有电视接收功能的视频卡还应具有 RF(射频)接口。

9.6 VGA 设备

1. VGA 采集卡

屏幕采集,通常可采用如下所述的几种方法。

(1) VGA-AV 转换器与视频卡

先通过 VGA-AV 转换器将 VGA 信号转换为视频信号,再采用视频采集压缩卡进行处理。这种方法的特点是:VGA 信号(一般分辨率为 1024×768 像素或 800×600 像素)转换为视频信号(一般采用 CIF,即 352×288 像素)以后,即使不压缩,清晰度也大大降低,文字、网页等内容基本上已无法看清,如再经过压缩,信号质量将非常差。

(2) 软件拷贝

采用软件拷贝的方法存在两个致命的缺点。第一,将占用 CPU 时间,这势必严重干扰正常的软件运行和性能;第二,在播放视频文件时,无法采用软件拷贝的方式采集画面。这是由于显示卡在播放视频时一般采用 Overlay 方式,软件拷贝采集到视频部分的只是一个黑窗口。有些拷贝软件会采取禁用 DirectX 的方式来获取到所播放视频的内容,但这样会导致显示性能急剧降低,是实际使用所不允许的。

(3) VGA 采集卡

VGA 采集卡可以把输入的 VGA 模拟信号实时采集压缩,既能保证信号连续、实时,又可保证清晰不失真,从而完美地解决了 VGA 信号的实时采集压缩这一难题。

屏幕采集卡一般是一块具有 PCI 接口的内置板卡,外部有 VGA 输入接口,可以连接另一台计算机的显示卡的 VGA 输出,或者其他 VGA 输出设备,以便将 VGA 信号采集进计算机,成为数字视频信号,如图 9-17 所示。

2. VGA 远程传输设备

在显示广告工程、工业自动化控制系统、医疗显示系统、安防监控系统、多媒体网络教学系统、长距离切换器、远距离网络服务器管理等应用中,都需要将 VGA 信号进行远距离传输,因此必须有 VGA 远程传输设备,如图 9-18 所示。

VGA 远程传输设备一般由发送器和接收器组成,可以通过非屏蔽超 5 类双绞线传输 VGA 信号,无损耗传输距离可达 $200\text{ m} \sim 1000\text{ m}$,支持的 VGA 信号的分辨率可达 1600×1200 像素。

3. 视频分配器

视频分配器(图 9-19)的作用是将一路视频输入信号分成多路相同的视频输出。目前实际应用的视频分配器一般不止一路输入,而是多路输入和多路输出,输入和输出数用 $m \times n$ 表示。

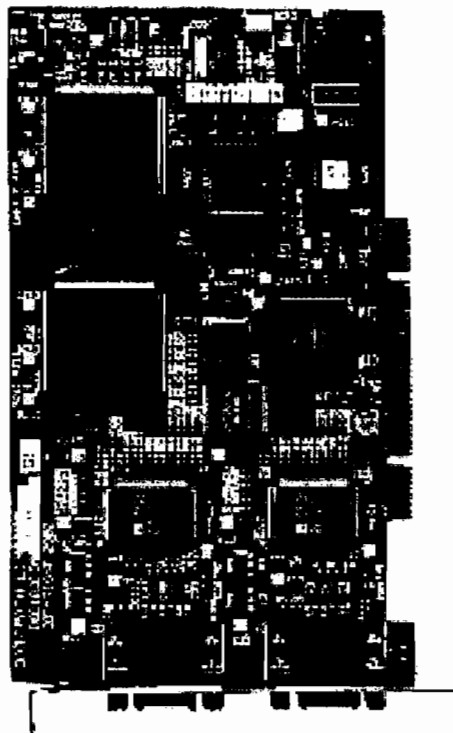


图 9-17 VGA 采集卡

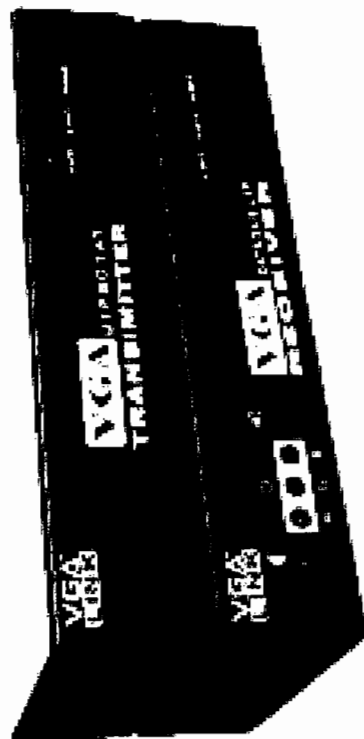


图 9-18 VGA 远程传输设备——发送器和接收器

视频分配器按功能可分为普通分配器和特殊分配器,如长距离分配器、多功能分配器(AV、VGA)及工作站专用分配器等。

4. 多计算机切换器

多计算机切换器,也称 KVM (Keyboard, Video and Mouse) 切换器,一般具有视频切换和多计算机控制的功能。视频切换是从多路输入信号中选择一路输出给显示设备。从多路输入信号源中选择两路或两路以上输出给不同的显示设备,这种视频切换叫做矩阵切换。

图 9-20 所示是通过一组键盘、鼠标和显示器实现操控 4 台主机的多计算机切换器。此设

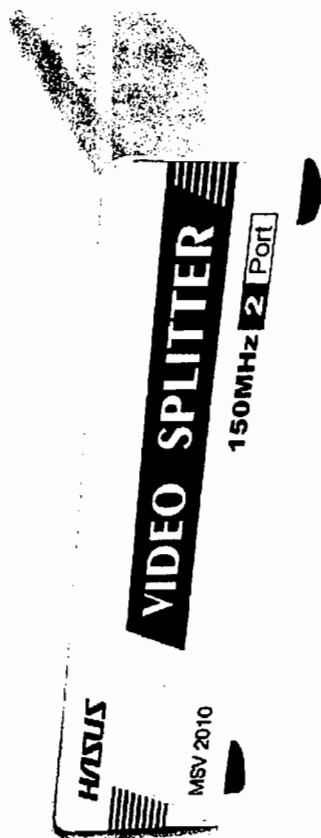


图 9-19 视频分配器 MSV 2010

备允许 4 台主机同时开机,具有热插拔功能。具有自动扫描每台主机画面的功能,主机切换可采用按钮或键盘热键,直接在键盘上切换某一台主机。切换时自动仿真并记忆每一台计算机的键盘及鼠标状态。

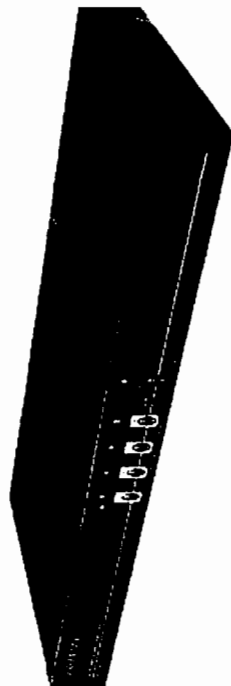


图 9-20 MPC4201 多计算机切换器

图 9-21 所示是通过一组键盘、鼠标和显示器实现操控 16 台主机的多计算机切换器。

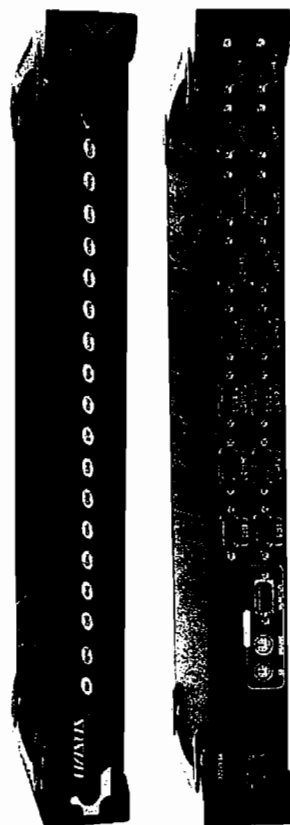


图 9-21 MPS1316 多计算机切换器

图9-22所示是采用双控制端设计,通过两组键盘、鼠标和显示器实现操控8台主机的多计算机切换器。此设备允许8台主机同时开机,具有热插拔功能。通过树状串接技术,可与相同系列机型串接,最多可控制64台计算机。具有多种切换方式,切换时自动仿真并记忆每一台计算机的键盘及鼠标状态。



图9-22 MP928多计算机切换器

9.7 投影机

随着多媒体技术和数字技术的迅猛发展,投影机也从早期的AV领域高端产品成为普通的计算机外部设备,特别是近年现代办公正朝着高效、便捷、移动的方向发展,投影机也得到了越来越多的应用。投影机可应用于多媒体教室、会议室、网络中心、指挥监控中心等一些固定场所,并且已扩展到家庭影院和移动办公领域,产品也越来越多元化。投影机可以与计算机、录像机、电视机、影碟机以及实物展台等相连接,可以说它是一种应用十分广泛的大屏幕影像设备。

9.7.1 投影机的工作原理

投影机主要有CRT、LCD、DLP(Digital Light Processor)、D-ILA(Direct-Drive Image Light Amplifier)等。CRT和LCD投影机技术成熟,应用时间较长,性能稳定。而DLP投影机应用时间较短,技术有待于进一步完善,但是该投影机采用微镜反射投影技术,亮度和对比度明显提高,体积和重量明显减小,具有较强的生命力和市场潜力。

1. CRT投影机

CRT是实现最早、应用最广泛的一种投影技术,其技术特点与大家所熟知的CRT显示器基本相同。CRT投影机把输入信号源分解到R(红)、G(绿)、B(蓝)三个CRT显像管的荧光屏上,在高压作用下将信号放大、会聚,并在大屏幕上显示出彩色图像。投影机通过光学系统和CRT管组成投影管,通常所说的三枪投影机就是指由三个投影管组成的投影机。CRT投影机由于使用内光源,也叫做主动式投影。

CRT投影机有两个重要的指标:会聚性能和CRT管的聚焦性能。对CRT投影机来说,会聚性能显得格外重要,这主要是因为CRT投影机拥有R、G、B三个CRT显像管,平行地安装在支架上,想要做到图像完全会聚则必须校正产生的各种失真现象。另外,投影机每一次的位置调整都会导致会聚性能发生变化,因此也需要重新调整会聚。会聚又有静态会聚和动态会聚之分,动态会聚主要包括倾斜、弓形、幅度、线性和梯形等功能,每一种功能均可在水平或垂直方向上进行调整。除此之外,还需进行非线性平衡、梯形平衡、枕形平衡的调整。CRT管的聚焦主要通过静电聚焦、磁聚焦和电磁复合聚焦三种。大家知道组成图像的最小单位是像素,像素越小,图

像的分辨率也就越高。在CRT投影机中,最小像素是由聚焦性能决定的。在这三种聚焦方式中以电磁复合聚焦最为先进,优点是聚焦性能好,尤其是在高亮度条件下聚焦。

CRT投影机的技术十分成熟,特别是CRT投影机在采用当前技术先进的CRT新型荫罩后,亮度也有了较大提高。但其重要技术指标图像分辨率与亮度相互制约,直接影响CRT投影机的亮度值。迄今为止,其亮度值始终徘徊在300lm以下。

CRT投影机的优点是显示的图像色彩丰富,还原效果好,具有丰富的几何失真调整能力。但CRT投影机操作复杂,特别是会聚调整繁琐,而且投影机体积较大,只适合安装在环境光线较弱、相对固定的场所。另外CRT投影机还必须由训练有素的专业人员进行安装和调节。而且CRT投影机价格昂贵,一般单位无法承受,因此在LCD、DLP等新型投影机的打压下,CRT投影机的市场份额逐年下滑。

2. LCD投影机

液晶是介于液体和固体之间的物质,本身不发光,工作性质受温度影响很大,其工作温度为 $-55 \sim +77^{\circ}\text{C}$ 。LCD投影机利用液晶的光电效应,即液晶分子的排列在电场作用下发生变化,影响其液晶单元的透光率或反射率,从而影响它的光学性质,产生具有不同灰度层次及颜色的图像。

LCD投影机是液晶显示技术和投影技术相结合的产物,它利用了液晶的电光效应,通过电路控制液晶单元的透射率及反射率,从而产生不同灰度层次及多达1670万种色彩的靓丽图像。LCD投影机的主要成像器件是液晶板。LCD投影机的体积取决于液晶板的大小,液晶板越小,投影机的体积也就越小。

LCD投影机的光源是专用大功率灯泡,发光能量远远高于利用荧光发光的CRT投影机,所以LCD投影机的亮度和色彩饱和度都高于CRT投影机。LCD投影机的像是液晶板上的液晶单元,液晶板一旦选定,分辨率就基本确定了,所以LCD投影机调节分辨率的功能要比CRT投影机差。

LCD投影机比用相同功率光源灯的DLP投影机有更高的ANSI流明光输出,在高亮度竞争中,LCD依然占着优势。LCD投影机明显缺点是黑色层次表现太差,对比度不是很高。

LCD投影机的另一个缺点是打出的画面看得见像素结构,观众好像是透过窗格子在看画面。SVGA(800×600像素)格式的LCD投影机,不管屏幕图像的尺寸大小如何,都能看得清楚像素格子,除非用分辨率更高的产品。

LCD投影机分为液晶光阀、液晶板和反射式液晶板三种。

(1) 液晶光阀投影机

液晶光阀投影机采用CRT管和液晶光阀作为成像器件,是CRT投影机与液晶光阀相结合的产物。为了解决图像分辨率与亮度间的矛盾,它采用外光源,也叫被动式投影。一般的光阀主要由三部分组成:光电转换器、镜子、光调制器,它是一种可控开关。通过CRT输出的光信号照射到光电转换器上,将光信号转换为持续变化的电信号,外光源产生一束强光,投射到光阀上,由内部的镜子反射,通过光调制器,改变其光学特性,紧随光阀的偏振滤光片,将滤去其他方向的光,而只允许与其光学缝隙方向一致的光通过,这个光与CRT信号相复合,投射到屏幕上。它是目前为止亮度、分辨率最高的投影机,亮度可达6000lm,分辨率为2500×2000像素,适用于环

境光较强,观众较多的场合,如超大规模的指挥中心、会议中心及大型娱乐场所,但其价格高,体积大,光阀不易维修。

(2) 液晶板投影机

液晶板投影机的成像器件是液晶板,也是一种被动式的投影。利用外光源金属卤素灯或UHP(冷光源),若是三片液晶板设计,则把强光通过分光镜形成R、G、B三束光,分别透射过R、G、B三色液晶板;信号源经过模拟/数字转换、调制加到液晶板上,控制液晶单元的开启、闭合,从而控制光路的通断,再经镜头放大,显示在大屏幕上。

而单片液晶板设计,光线不用分离,这种投影机体积小,重量轻,操作、携带极其方便,价格也比较低廉。但其光源寿命短,色彩不很均匀,分辨率较低,最高分辨率为 $1\,024 \times 768$ 像素,多用于临时演示或小型会议。这种投影机虽然也实现了数字化调制信号,但液晶本身的物理特性,决定了它的响应速度慢,随着时间的推移,性能有所下降。

液晶板的大小决定着LCD投影机的大小,液晶板规格越小,则LCD投影机的光学系统就能做得越小,从而使LCD投影机越小。但是在很小的液晶板上做到高分辨率,并且保持高亮度,其技术之难是可想而知的。目前 0.9 英寸和 0.7 英寸的面板产量最大,比例达到 70% 以上, 1.3 英寸产品比例在 15% 左右, 0.5 英寸、 0.79 英寸和 0.99 英寸和 1.0 英寸面板也开始用于投影机产品。在液晶板数量上,由于单片结构在性能和色彩方面的缺陷,目前已经基本被淘汰,主流为3片式LCD投影机。由于在性能和色彩方面表现出色,在很长一段时间内,都代表了投影机产品发展的最成熟水平。在同等亮度和分辨率的情况下,投影机体积越小价格相应越高。

(3) 反射式液晶板投影机

反射式液晶板(LCOS)采用的D-ILA技术,其核心部件是反射式有源矩阵硅-液晶板,所以业界将此技术称为反射式液晶技术。

在透射式LCD技术中的液晶板中,作为像素点开关控制的晶体管被集成在液晶板上相应位置上,在光源透射过程中,晶体管本身将阻挡部分光线,因此采用透射式液晶技术的投影机的光源的利用效率不高,很难实现高亮度。从原理上分析,在不增加液晶板尺寸的前提下,透射式LCD投影机的标称分辨率的发展空间已经不大。D-ILA在提供高分辨率和对比度方面显示出技术优势。

D-ILA技术中液晶板将晶体管作为像素点液晶的开关控制单元集成在一层硅基板上,硅基板(也称反射电极层)位于液晶层的下面,用于像素地址寻址的各种控制电极和电极间的绝缘层位于硅基板的下面,因此整个结构是一个立体排列方式。来自光源的光不能穿透反射电极层,而被反射电极层反射,避免了下面的各种结构层对光线的阻挡。因此采用D-ILA技术的液晶板的光圈比率可以达到 93% (DLP技术中DMD的光圈比率为 88% ,而透射式LCD的液晶板的光圈比率为 $40\% \sim 60\%$),因此采用D-ILA技术的投影机对光源的利用效率更高,可以实现更高的亮度输出。由于D-ILA技术的液晶板的液晶层采用电压控制可调制反射方式,在全开状态的光线全反射,几乎没有损失,而全状态式反射输出光线几乎为零,因此D-ILA可以实现非常高的对比度。

JVC推出的DLA-G150HT采用3块具有 $1\,365 \times 1\,024$ 像素的 0.9 英寸D-ILA显示板,可对VGA、SVGA、XVGA以及UXGA等画面,具有 $600:1$ 的高画面对比度,并且具有能够展现亮丽影像质感的 $1\,000$ lm的光输出强度。此系列产品正向更高的清晰度和亮度迈进,据悉下一代产

品的对比度将超过 $2000:1$ 。

JVC公司已经推出了 3.3 cm(1.3 英寸)、 3.04 cm(1.2 英寸)、 2.29 cm(0.9 英寸)和 1.78 cm(0.7 英寸)的D-ILA液晶板。

3. DLP投影机

DLP技术是美国德州仪器公司开发的一种专利技术,这一新的投影技术的诞生,使我们在拥有捕捉、接收、存储数字信息的能力后,终于实现了数字信息显示。DLP技术是显示领域划时代的革命,正如CD在音频领域产生的巨大影响一样,DLP将为视频投影显示翻开新的一页。DLP主要是以数字微镜设备(Digital Micromirror Device, DMD)作为光阀成像的器件,单片DMD是由很多微镜组成,每块微镜都对应着一个成像点,因此DLP投影机的物理分辨率取决于微镜的数目。DLP投影技术是一种全数字反射式投影技术,其主要的优点是数字优势。采用了数字技术之后,可以大大提高图像的灰度等级,使图像噪声消失,画面质量稳定,数字图像定位非常精确;另外它还具有反射优势,采用反射式DMD器件之后,成像器件的总光效率得到很大提高,对比度、亮度、均匀性都非常出色。

根据使用的DMD单元的数量,DLP投影机可分为单片式、双片式和3片式DLP投影机。DLP投影机清晰度高、画面均匀、色彩还原效果好,3片式DLP投影机的亮度可达 $1\,000$ lm以上。DLP投影机抛弃了传统意义上的会聚,可随意变焦,调整十分便利。由于DLP技术是种新技术,DMD芯片的优良率还不是很低,因此德州仪器在DMD芯片产量上还很难满足全球投影机生产的需要,所以DLP投影机在价格、维修难度和费用上没有任何优势。但DLP技术的最大优势是可以把投影机体积做得很小,超便携投影机便由此诞生。目前超便携DLP投影机的重量可以达到 1.5 kg以下。DLP投影机的光学、机械特性,也决定了它的移动、防震性能要比LCD投影机差一些。与LCD投影机一样,DLP的像元也是固化的,所以它的分辨率调整功能较差。

与LCD技术一样,DMD芯片尺寸是决定投影机体积和重量的重要因素。目前德州仪器推出了 0.55 英寸、 0.7 英寸、 0.9 英寸和 1.1 英寸多种尺寸的芯片。DLP投影机最常见的结构有单片式和3片式两种,其中3片式结构主要应用于影院系统和高性能产品中,市场上常见的普通应用的产品全部是单片式结构。人们普遍谈论的DLP技术和LCD技术的比较,也主要是对单片式DLP技术和3片式LCD技术之间的比较。单片式DLP投影机采用色轮来实现分色,三原色用同一个成像器件,与三原色各有一套成像系统的3片式LCD投影机相比,单片式DLP投影机在色彩饱和度和方面不如3片LCD投影机。第一代DLP投影机的色轮转速为 60 转/秒,第二代DLP产品的色轮转速提高了一倍,为 120 转/秒。新一代的DLP投影机的色轮转速仍为 120 转/秒,不过色轮采用了6分色(以前采用3分色),相当于把转速又提高了一倍,达到了 240 转/秒。因此目前的DLP投影机的色彩表现已经得到了很大提高,但是与LCD产品相比,大部分单片DLP投影机产品的色彩表现还有差距。

4. DLV投影机

数字光阀(Digital Light Valve, DLV)是一种将CRT透射式投影技术与DLP反射式投影技术结合在一起的新技术。该技术的核心是将小管径CRT作为投影机的成像面,并采用氙灯作为光源,将成像面上的图像射向投影面。通过CRT输出的光信号照射到光电转换器上,将光信号转换为持续变化的电信号;外光源产生一束强光,投射到光阀上,由内部的镜子反射,能通过光调制

器改变其光学特性,紧随光闸的偏振滤光片,将滤去其他方向的光,而只允许与其光学缝隙方向一致的光通过,这个光与CRT信号相复合,投射到屏幕上。因此,DLV投影机在充分利用CRT投影机的高分辨率和可调制性特点的同时,还利用氙灯光源高亮度和色彩还原好的特点。DLV投影机不仅是一款分辨率、对比度、色彩饱和度和很高的投影机,还是一款亮度很高的投影机。其分辨率普遍达到1250×1024像素,最高可达到2500×2000像素,对比度一般都在250:1以上,色彩数目普遍为24位,1670万种,投影亮度普遍在2000lm~12000lm,可以在大型场所中使用。

Panasonic 的 PT-UX88/PT-UX68 投影机的设计理念是高性能、易携带,因此,该投影机将所有的功能集于简洁、轻巧的机身中。PT-UX88/PT-UX68 通过 DLV 投影技术显著提高了画面质量。

在新型 DLV 投影功能下,投影机可根据房间光线的环境自动设计并调整投影图像的方式。这种技术大大提高了投影机在明亮房间里对色彩的感应度。它把因明亮环境而减弱的颜色分为6种基础色(红、绿、蓝、黄、青、紫红),分别加以精确的控制。DLV投影机将画面质量提高到了一个崭新的水平,使投影画面质量和计算机上显示的画面质量相差无几。PT-UX88/PT-UX68操作快捷简便,快速启动功能在用户按下启动键两秒内即可投影图像。当演示结束时,直接按下关机键即可切断电源离开,内置的电容存储了足够的能量来冷却灯泡。另外,一键自动设置功能和自动梯形校正功能在演示过程中大大节约了用户的时间,简化了操作。

5. 背投影机

其实背投是相对正投而言的,因此只要大家对正投的概念搞清楚了,背投的概念自然就清楚了。一般来说,投影机的正投就是观众和投影机位于投影屏幕的同一侧,那么从投影机发射出来的光照到投影屏幕上后,观众通过投影屏幕反射回来的光就能看到投影影像了。与此相对的是,背投是观众和投影机位于投影屏幕的两边,将投影机安装在机身内的底部,从投影机投射出来的光照射到半透明的背投屏幕时会有部分光透过屏幕,观众是通过透射出来的光来看到投影的影像的。由一台投影机、一块背投屏幕和反射系统以及结构构成的单机单画面投影显示系统,是最简单的背投影系统。将计算机中的信息或者其他输入端的信息输出到投影机,再通过背投影机投射到专用背投屏幕上,这种投影方案可以适用于商务演示、多媒体教学、会议室等许多场合。

与正投影机的分类一样,根据投影技术的不同,背投影机也可以分为CRT背投、LCD背投DLP背投以及LCOS背投等几种类型。不同类型的背投影机所实现的功能和应用场合可能也有所不同。

由于背投是将投影机做在箱体里的,投射到屏幕上的图像不会受到环境光的影响,因此在较亮的环境中也可以完好地显示图像。此外,倘若遇到在背投影机或者屏幕在安装中,发现投影空间场地有限的时候,大家还可使用背投屏幕专用反射镜系统,也就是说用反射镜把放映光束折返一次或两次,在放映投射距离不足的空间内达到规定的放映距离,以保证画面尺寸。

6. 几种投影技术的比较

几种常用的投影技术的特点如表9-1所示。从表中可看出,基于D-ILA技术的投影机在成本、分辨率方面有着独有的优势。尽管目前在量产方面仍有技术瓶颈,但随着众多厂商的量产努

力,这一工艺技术瓶颈迟早会被克服。相对于目前投影机厂商长期在DLP的产量不足、价格贵及专利授权的压力下以及液晶面板供货也不足的前提下,D-ILA技术的成熟和不断发展将是投影机厂商的另一解决方案。

表9-1 几种常用的投影技术的特点

	LCD 投影机	DLP 投影机	D-ILA 投影机
优势	制程技术较完整,具备量产技术,光学引擎结构较简单	易于轻型化,光利用率高,高分辨率	可利用半导体制程大量生产,从而大幅降低面板生产成本,开口率提高,高分辨率
劣势	光利用率低,开口率低,存在散热问题	制程复杂,优良率低,仅T1提供晶片组	相对于DLP技术不利于轻型化,较高的光学元件成本,相对于DLP技术影像对比度较差

9.7.2 投影机接口的分类及用途

投影机的接口大致可以分为两类:模拟信号接口和数字信号接口,有些机型还提供有音频输入输出端子。

1. 模拟信号接口

模拟信号接口包括复合视频信号输入接口(CVBS)、S-Video输入接口、欧洲的标准视频接口(SCART)、色差输入接口、RGBHV输入接口和VGA-PC的15引脚VGA输入接口等。CVBS主要用于输入电视信号,S-Video输入接口是随着录像机的出现而发展起来的视频接口,现在被广泛应用于录像机和影碟机中。SCART接口用于传输RGB信号。色差输入接口用于传输亮度信号和色差信号分开的影像,有逐行扫描和隔行扫描之分。RGBHV接口传输的信号和VGA相当,但其5RCA的插座通常用于专用显示领域。VGA-PC的15引脚VGA输入接口是最常见的接口之一,用于连接计算机的视频输出。

2. 数字信号接口

数字信号接口包括USB接口、IEEE-1394接口、DVI接口和HDMI接口等。USB接口通常只是用来传送图片文件,而不能传送视频文件。IEEE-1394接口则可用于传送MPEG2数码图像,但由于技术较复杂而且比较昂贵,所以配置IEEE-1394接口的投影机也比较少见。DVI接口可传输高达2048×1536像素、75Hz的视频图像,因此被广泛应用于投影机和等离子显示器上,市面上大多数中高端投影机均配备有DVI数字接口,但最近也有些新产品采用HDMI接口代替了DVI接口。HDMI接口是在DVI接口的基础上,增加了数字音频输入,从而成为专用的多媒体信息接口,而且支持1920×1080像素的高清晰数字信号,可望成为未来的视频设备的标准接口。

与投影机关系最密切的是计算机,在投影机附带的连接线上,VGA信号线和信号控制线都是用来直接和计算机连接的。信号控制线用于连接投影机的主控制端口与计算机的COM端口。

器改变其光学特性,紧随光闸的偏振滤光片,将滤去其他方向的光,而只允许与其光学缝隙方向一致的光通过,这个光与 CRT 信号相复合,投射到屏幕上。因此,DLV 投影机在充分利用 CRT 投影机的高分辨率和可调制性特点的同时,还利用氙灯光源高亮度和色彩还原度高的特点。DLV 投影机不仅是一款分辨率、对比度、色彩饱和度很高的投影机,还是一款亮度很高的投影机。其分辨率普遍达到 1 250 × 1 024 像素,最高可达 2 500 × 2 000 像素,对比度一般在 250:1 以上,色彩数目普遍为 24 位、1670 万种,投影亮度普遍在 2 000 lm ~ 12 000 lm,可以在大型场所中使用。

Panasonic 的 PT-U1X88/PT-U1X68 投影机的设计理念是高性能、易携带,因此,该投影机将所有的功能集于简洁、轻巧的机身中。PT-U1X88/PT-U1X68 通过 DLV 投影技术显著提高了画面质量。

在新型 DLV 投影功能下,投影机能根据房间光线的环境自动设计并调整投影图像的方式。这种技术大大提高了投影机在明亮房间里对色彩的感应度。它把因明亮环境而减弱的颜色分为 6 种基础色(红、绿、蓝、黄、青、紫红),分别加以精确的控制。DLV 投影机将画面质量提高到了一个新水平,使投影画面质量和计算机上显示的画面质量相差无几。PT-U1X88/PT-U1X68 操作快捷简便,快速启动功能在用户按下启动键两秒内即可投影图像。当演示结束时,直接按下关机键即可切断电源离开,内置的电容存储了足够的能量来冷却灯泡。另外,一键自动设置功能和自动梯形校正功能在演示过程中大大节约了用户的时间,简化了操作。

5. 背投影机

其实背投是相对正投而言的,因此只要大家对正投的概念搞清楚了,背投的概念自然也就清楚了。一般来说,投影机的正投就是观众和投影机位于投影屏幕的同一侧,那么从投影机发射出来的光照到投影屏幕上后,观众通过投影屏幕反射回来的光就能看到投影影像了。与此相对的是,背投是观众和投影机位于投影屏幕的两边,将投影机安装在机身内的底部,从投影机射出来的光照到半透明的背投屏幕上时会有部分光透过屏幕,观众是通过透射出来的光来看看到投影的影像的。由一台投影机、一块背投屏幕和反射系统以及结构构成的单机单面投影显示系统,是最简单的背投系统。将计算机中的信息或者其他输入端的信息输出到投影机,再通过背投影机投射到专用背投屏幕上,这种投影方案可以适用于商务演示、多媒体教学、会议室等许多场合。

与正投影机的分类一样,根据投影技术的不同,背投影机也可以分为 CRT 背投、LCD 背投、DLP 背投以及 LCOS 背投等几种类型。不同类型的背投影机所实现的功能和应用的场合可能有所不同。

由于背投是将投影机是放在箱体里的,投射到屏幕上的图像不会受到环境光的影响,因此在较亮的环境下也可以完好地显示图像。此外,倘若遇到在背投影机或者屏幕在安装中,发现投影空间场地有限的时候,大家还可使用背投屏幕专用反射镜系统,也就是说用反射镜把放映光束折返一次或两次,在放映投射距离不足的空间内达到规定的放映距离,以保证画面尺寸。

6. 几种投影技术的比较

几种常用的投影技术的特点如表 9-1 所示。从表中可看出,基于 D-ILA 技术的投影机在成本、分辨率方面有着独有的优势。尽管目前在量产方面仍有技术瓶颈,但随着众多厂商的量产努

力,这一工艺技术瓶颈迟早会被克服。相对于目前投影机厂商长期在 DLP 的产量不足、价格昂贵及专利授权的压力下以及液晶面板供货不足的前提下,D-ILA 技术的成熟和不断发展将是投影机厂商的另一解决方案。

表 9-1 几种常用的投影技术的特点

	LCD 投影机	DLP 投影机	D-ILA 投影机
优势	制程技术较完整,具备量产技术,光学引擎结构较简单	易于轻型化,光利用率高,高分辨率	可利用半导体制程大量生产,从而大幅降低面板生产成本,开口率提高,高分辨率
劣势	光利用率低,开口率低,存在散热问题	制程复杂,优良率低,仅 TI 提供晶片组	相对于 DLP 技术不利于轻型化,较高的光学元件成本,相对于 DLP 技术影像对比度较差

9.7.2 投影机接口的分类及用途

投影机的接口大致可以分为两类:模拟信号接口和数字信号接口,有些机型还提供有音频输入输出端子。

1. 模拟信号接口

模拟信号接口包括复合视频信号输入接口(CVBS)、S-Video 输入接口、欧洲的标准视频接口(SCART)、色差输入接口、RGBHV 输入接口和 VGA-PC 的 15 引脚 VGA 输入接口等。CVBS 主要用于输入电视信号,S-Video 输入接口是随着录像机的出现而发展起来的视频接口,现在被广泛应用于录像机和影碟机中。SCART 接口用于传输 RGB 信号。色差输入接口用于传输亮度信号和色差信号分开的影像,有逐行扫描和隔行扫描之分。RGBHV 接口传输的信号和 VGA 相当,但其 5RCA 的插座通常用于专用显示领域。VGA-PC 的 15 引脚 VGA 输入接口是最常见的接口之一,用于连接计算机的视频输出。

2. 数字信号接口

数字信号接口包括 USB 接口、IEEE-1394 接口、DVI 接口和 HDMI 接口等。USB 接口通常只是用来传送图片文件,而不能传送视频文件。IEEE-1394 接口则可用于传送 MPEG2 数码图像,但由于技术较复杂而且比较昂贵,所以配置 IEEE-1394 接口的投影机也比较少见。DVI 接口可传输高达 2 048 × 1 536 像素、75 Hz 的视频图像,因此被广泛应用于投影机和高清晰度显示器上,市面上大多数中高端投影机均配备有 DVI 数字接口,但最近也有些新产品采用 HDMI 接口代替了 DVI 接口。HDMI 接口是在 DVI 接口的基础上,增加了数字音频输入,从而成为专用的多媒体信息接口,而且支持 1 920 × 1 080 像素的高清晰数字信号,有望成为未来的视频设备的标准接口。

与投影机关系最密切的是计算机,在投影机附带的连接线缆中,VGA 信号线和信号控制线都是用来直接和计算机连接的。信号控制线用于连接投影机的主控制端口与计算机的 COM 端口。

在日常的应用中,为了方便计算机与投影机的屏幕切换,通常使用一个视频分配器进行中转连接,从视频分配器上引出两路VGA信号,其中一路信号传输到投影机中显示,另一路信号传输到显示器中显示,演示者只需通过计算机的显示器操作,而不需要站在投影幕布前面看着投影内容进行讲解了。笔记本计算机一般都提供D-Sub 15引脚VGA输出接口,再加上自身的LCD显示器,因此没必要使用视频分配器连接,但需要注意的是输出模式。笔记本计算机一般提供3种输出模式:仅液晶屏幕输出,仅VGA端口输出,液晶屏幕与VGA端口同时输出。默认状态是仅液晶屏幕输出。因此在连接投影机时,需要更改笔记本的输出模式。

3. 投影机连接的未来方向

随着投影技术的不断成熟,投影机越来越趋向人性化设计,“即插即用”使得投影机的操作显得十分方便。目前绝大多数投影机都能自动识别输入的信号源,同时根据信号源类型的不同,自动进行切换、自动更改设置。投影机的连接技术将向着越来越人性化的方向发展。另外,无线连接的投影机也是投影机未来发展的趋势。

9.7.3 投影机的主要性能指标

投影机的主要性能指标有亮度、对比度、扫描频率和视频带宽、分辨率和灯泡寿命。

1. 亮度

亮度是指投影机输出到屏幕上的光的强度,为指屏幕表面受到光照射发出的光能量与屏幕面积之比。目前,投影机亮度的度量单位是lm(流明)高的亮度可以使投影机投射图像清晰亮丽。当投影机输出的光通量一定时,投射面积越大亮度越低,反之则亮度越高。决定投影机光输出的因素有投影及荧光屏面积、性能及镜头性能,通常荧光屏面积越大,光输出越高。带有液体耦合镜头的投影机镜头性能好,投影机光输出也可相应提高。便携式投影机的亮度指标一般都在800 lm~2 000 lm之间。

2. 对比度

对比度表示的是投影机所呈现的黑色与白色之间的亮度差异。对比度越高,投影机能表现的颜色越细腻。根据芯片成像原理的不同,一般来说LCD便携投影机的对比度可能达到400:1以上,而DLP便携投影机的对比度可达到1 000:1以上。显然在对比度这一指标上DLP投影机要优于LCD,但这并不是说一定要选择DLP投影机,因为事实上一般文档投影只需要400:1的对比度即可。用户可以在选购时试一下投影效果。

3. 扫描频率和视频带宽

(1) 水平扫描频率(行频)

电子束在屏幕上从左至右的运动叫做水平扫描,也叫行扫描。每秒钟扫描次数叫做水平扫描频率(行频)。视频投影机的水平扫描频率是固定的,为15.625 kHz(PAL制式)或15.725 kHz(NTSC制式)。数据和图形投影机的扫描频率不是固定的。频率范围在15 kHz~60 kHz的投影机通常叫做数据投影机,在这个频段内,投影机可自动跟踪输入信号行频,由锁相电路实现与输入信号行频的完全同步。水平扫描频率是区分投影机档次的重要指标。

(2) 垂直扫描频率(场频)

电子束在水平扫描的同时,又从上向下运动,这一过程叫垂直扫描。每扫描一次形成一幅图

像,每秒钟扫描的次数叫做垂直扫描频率,也叫做刷新频率,它表示这幅图像每秒刷新次数。对于CRT投影机来说,垂直扫描频率一般不应低于50 Hz,否则图像会有闪烁感。

(3) 视频带宽

投影机的视频带宽是指在视频信号幅值下降至0.707倍时,对应的信号上限频率。0.707倍对应的增量是-3 dB,因此又叫做-3 dB带宽。

4. 分辨率

在投影机的性能指标中,分辨率是较易混淆的一个概念。SVGA、XGA和SXGA是目前投影机支持的几种主要的分辨率。投影机都有一个真实分辨率(本机分辨率),表示投影机实际上一次能够投影的最大像素数量。换句话说,SVGA投影机一次只能显示48万(800×600)像素。

分辨率不只影响图像的清晰度,还影响投影机对计算机图像的像素数量,使得计算机图像中的一些像素在投影机放映时是通过同一个像素表现出来的。对照片和视频图像来说,这样做并不会带来太大的影响,因为你不会太注意图像的细节。但是对于文本显示质量却大打折扣。

另外,分辨率还有RGB分辨率和视频分辨率。RGB分辨率是指投影机在接RGB视频信号时可达到的最高像素,RGB分辨率与水平扫描频率、垂直扫描频率及视频带宽均有关;视频分辨率是指投影机在显示复合视频时的最高分辨率。

5. 灯泡的使用寿命

除CRT投影机以外,其他类型的投影机都有灯泡。灯泡作为投影机的唯一消耗材料,也是选购投影机时必须考虑的重要因素。一般投影机都预先装好了灯泡,而且厂商也提供了90天或者是使用1 000小时的保修,大部分灯泡的平均使用寿命大于2 000小时。此外,要注意所购投影机所使用的灯泡在国内是否有销售点,因为不同品牌投影机使用的灯泡一般是不能互换使用的,这样可以避免因买到不合适的灯泡造成投影机不能使用。

另外,在选购时还应根据使用的场合和目的来选择适合的接口类型。通常,除了要有模拟信号接口外,目前有必要考虑有数字信号接口,如DVI和HDMI接口。

投影机要有遥控器,可方便地进行投影机的控制。有些遥控器还有激光的功能,甚至有些遥控器还可当无线的鼠标,可以用遥控的方式控制计算机。

9.7.4 便携式投影机

通常认为重量在2.5 kg以内的投影机为便携式投影机。便携式投影机一般可以用一个笔记本电脑包大小的包携带。目前便携式投影机所采用的是LCD和DLP技术。

LCD投影机的核心部件是液晶板,这种技术应用于投影机历史较长,技术比较成熟。随着LCD 0.5英寸液晶板投入使用以及光路设计的改进,LCD投影机产品的重量已经降到了1.6 kg左右,携带更加方便。

采用DLP处理技术的便携式投影机,由于采用DMD芯片,在便携性方面颇具优势,重量只有1 kg左右,故有人戏称为超轻便投影机。

一般来说,应用便携投影机通常不是用来演示复杂的视频和动画,而是用来展示静态的文

档、图表(比如 Word 文档、PowerPoint 演示文稿)这样的内容,主要用于一些具有移动性和临时性的集会、商务演示等,参加的人员一般在几十人之内,投影面积在 60 英寸以内。所以便携式投影机不需要太高的亮度。

对于便携投影机,应主要考虑的是其重量、携带方便和防震措施。

不要只考虑到单机的重量,也要把其他附属器材考虑在内,包括电缆、遥控器以及携带盒等。投影机是一种精密电子产品,它集机械、液晶或 DMD、电子电路技术于一体,震动对它们是致命杀手。以液晶投影机为例,强震会造成液晶片的位移,影响放映时 LCD 的会聚,出现 RGB 颜色不重合的现象;光系统中的透镜,反射镜也会产生变形或损坏,影响图像投影效果;变焦镜头在冲击下会使轨道损坏,造成镜头卡死,甚至镜头破裂无法使用。由于便携式投影机要移动办公所用,这方面一定要加以重视。如果演示文稿有音效,要选购一台配备有较佳内置喇叭的投影机。

习 题

1. 简述多媒体概念。
2. 列举音频卡的主要特点及主要功能部件。
3. 列举音频卡的音频接口的特点及常用标准。
4. 举例说明摄像头的工作原理。
5. 举例说明多功能显示卡的常见功能和应用原理。
6. 简述视频卡的常见功能。
7. 简述视频卡的主要性能指标。
8. 举例说明常见的 VGA 设备和应用原理。
9. 简述投影机的分类及其工作原理。
10. 列举投影机的接口的特点及常用标准。

计算机通信和通信设备

第10章

10.1 计算机通信

计算机通信是计算机技术和通信技术相互融合的产物。计算机与通信的相互融合主要体现在以下两个方面:一方面,通信网络为计算机之间的数据传输和交换提供了必要的手段;另一方面,数字计算技术的发展渗透到通信技术中,又提高了通信网络的各性能。现代术语“Telecommunication”(电信)一词可以说体现了这两方面的技术,计算机通信也从属于电信的范畴。

计算机通信是为了达到协同工作的目的,即在两台或多台计算机之间经由数据通路(包括通信网络)进行的信息交换。上述计算机必须是“自治的”,也可理解为任何一台计算机都是独立运行的,不受其他计算机的控制。

10.1.1 计算机技术与通信技术的融合

用于提供语音、数据、图像和视频传输服务的通信设备对于数字技术和计算机系统依赖正在不断地增强。原先的电话网都是采用模拟技术实现接入和交换的,目前正在不断地引入数字技术和计算机技术。这主要是由于数字技术的众多优越性,如大规模和超大规模集成电路的使用降低了数字设备的成本;采用数字技术的时分复用比基于模拟的频分复用更有效、更便于实现线路共享;监测网络状态和控制网络运行的绝大多数信号本身就是数字信号,很容易并入全数字的网络。电话网中越来越多地使用数字技术,而这些技术都离不开计算机,这是因为计算机能够高效、自动地完成信息的提取、存储和处理。目前电话网中的许多设备是嵌入系统,即基于计算机的系统。例如,数字程控交换机实际上就是一台用于电话交换网的专用计算机,呼叫处理、路由建立和维护、特种业务的处理等等都受到系统中 CPU 的控制。

另外,远程信息处理和通信的发展也越来越多地引入了计算机技术。远程信息处理可以利用公用电话网,把数字数据转换成模拟信号,并以呼叫的形式像语音信号一样

传输,但传输效率很低。为了提高效率,将计算机的数据以数据包(称为分组,packet)的形式发送。分组在网络中传输,在发送分组的用户之间共享路径和资源,因此,网络必须知道如何处理分组。这也是通信呈现计算机化的另一个例子。

在通信设备不断使用计算机技术的同时,计算机也在越来越多地使用通信技术。许多数据处理的事务通常需要一组协同工作的计算机来实现,这称为分布式处理。分布式处理不仅要依靠通信设备将系统中的终端和计算机互连起来,而且要求计算机本身也必须具备通信的能力,并能实现高速及可靠的信息传输和交换。

10.1.2 通信系统的组成

通信系统是一个相当复杂的系统,任何一个通信系统都可以抽象为如图10-1所示的模型。由图10-1可见,通信系统应包括信源、变换器、信道、反变换器和信宿等几个要素。另外,在通信过程中会叠加上由噪声源产生的干扰。这些要素的含义如下所述。

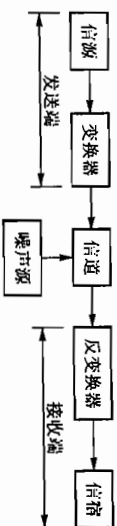


图 10-1 通信系统的一般模型

1. 信源

信源是发出信息的一端,可分为模拟信源和数字信源。模拟信源(如电话机、电视摄像机)输出连续幅度的模拟信号;数字信源(如电传机、计算机等各种数字终端设备)输出离散的数字信号。在人们打电话时,说话者就是信源,通过人的声带振动,会产生声音(为模拟声波信号)。

2. 变换器

变换器的作用是把信源产生的各种原始信号转换成适合在某种具体信道中传送的信号,转换后的信号同样载有原有的信息。例如,电话机的送话器(话筒)就是变换器,它用于将声音(声波信号)转换成适合在电话线上传送的模拟电信号。变换器通常由编码器和调制器组成,编码器的作用是将信源发出的信号转换为适合传输(如便于识别、纠错等)的信号,而调制器的作用是将要传输的信号以适合传输的形式表示(如按频率、幅度、相位表示等)。

3. 信道

信道是指传输信号的通道,可以有线的,也可以是无线的。有线和无线又有多种传输介质,如有线通信可采用铜线、同轴电缆、光缆等,无线通信可采用红外线、不同频段的无线电波等。当信号在信道中传输时,信道也对信号产生各种干扰和噪声。传输介质的固有特性和干扰直接关系通信的质量。

4. 反变换器

反变换器的基本功能是实现变换器的反变换,其任务是从带有干扰的接收信号中正确地恢复原始信号。对于多路复用信号,还具有解除多路复用和实现正确分路的功能。电话机的扬声器

器(听筒)的作用就是将电话线上传来的模拟电信号还原为声音。反变换器通常由解调器和解码器组成,解调器和解码器的作用与调制器和编码器正好相反。接收到的信号通过解调器和解码器后,理论上将获得与信源发出的相同的信号。

5. 信宿

信宿是传输信息的归宿,其作用是将复原的原始信号转换成相应的消息。在人们打电话时,听者就是信宿,通过其耳朵接收声音。

6. 噪声源

噪声源是信道中的噪声以及分散在通信系统其他各处的噪声的集中表示。在打电话时听筒中的杂音就是由于干扰产生的。

根据信道传输的信号的特点可以分为数字信道(以数字离散信号传输数据)和模拟信道(以连续模拟信号传输数据),模拟信号和数字信号如图10-2所示。

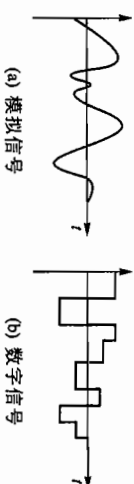


图 10-2 模拟信号和数字信号

10.2 数字通信的基本概念

根据所采用的信道,通信可分为模拟通信和数字通信。与模拟通信相比,数字通信具有如下优点:

- 抗噪声(干扰)能力强。
- 可以控制差错,提高传输质量。
- 便于用计算机进行处理。
- 易于加密,保密性强。
- 可以传输语音、数据、影像,通用而灵活。

10.2.1 数据通信的分类

数据通信的分类可以有多种方式,下面是最常用的两种分类方式。

① 根据数据传输是多位还是一位,数字通信可以分为并行通信和串行通信两种方式,而串行通信又可分为同步通信和异步通信。随着通信技术的发展,目前,有用串行通信取代并行通信的趋势。

② 根据传输模式(通信方式,数据流动的方向)数字通信可以分为:单工(数据只能单向传输)、半双工(数据可以双向传输,但不能在同一时刻双向传输)和全双工(数据可同时双向传

列——7FH,来标识一个帧的开始,适用于任意数据类型类型的帧)两种。
数字通信两个方向的信号的传输可以具有两条物理上独立的传输线路,也可将带宽一分为二,分别用于不同方向的信号传输。

10.2.2 通信速率与数据传输方式

带宽是通信信道的宽度。在模拟通信中是指信道频率上界与下界之差,是介质传输能力的度量,以赫兹(Hz)为单位。在数字通信中一般使用位每秒(b/s或bps)作为带宽的计量单位,实际上就是最高的数据传输速率。

表征信号传输速率的是波特率,单位是波特(Baud)。波特率表示每秒传送的码元数,而在计算机通信中一般常用的是数据传输速率(比特率)。一个信号往往可以携带多个二进制位,所以在固定的信号传输速率下,比特率往往大于波特率。换句话说,一个码元中可以传送多个比特。比特率和波特率的关系如下式所示:

$$\text{比特率 (b/s)} = \text{波特率} \times \log_2 M$$

其中,M是信号的编码级数。例如,当波特率为9600时,若M=16,数据传输速率为38.4 Kb/s;若M=2,则数据传输速率为9600 b/s。

数据的传输可以是以基带(编码后的数字信号直接在信道上发送)或频带(调制成模拟信号后再传送)方式进行传输。在某些有线信道中,若传输距离不太远且通信容量不太大时,可以直接传送数字信号,我们称之为数字信号的基带传输。而在另外一些信道,特别是无线信道和光纤信道中,数字基带信号则必须经过调制将信号频谱搬移到高频处才能在信道中传输,我们把这种传输称为数字信号的频带传输。

10.2.3 数据的同步方式

在通信时,数据的同步是相当重要的,同步目的是使接收端与发送端在时间基准上一致(包括开始时间、位边界、重复频率等)。同步的方式一般有同步(见图10-3)、字符同步(异步传输)(见图10-4)和帧(Frame,数据链路中的传输单位)同步(同步传输)(见图10-5)。位同步是最严格的同步方式。字符同步以字符为边界实现字符的同步接收,也称为起止式或异步制。每个字符的传输需要:1个起始位,5~8个数据位,1.5或2个停止位,计算机的串行通信采用的就是字符同步方式。帧同步需要能识别一个帧的起始和结束,帧同步有面向字符(以同步字符——SYN,16H,来标识一个帧的开始,适用于数据为字符类型的帧)和面向比特(以特殊位序

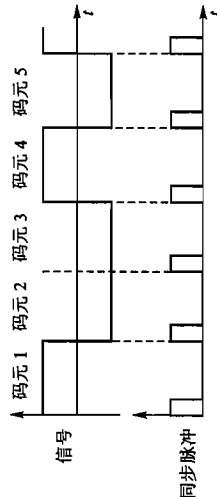


图 10-3 位同步方式

列——7FH,来标识一个帧的开始,适用于任意数据类型类型的帧)两种。

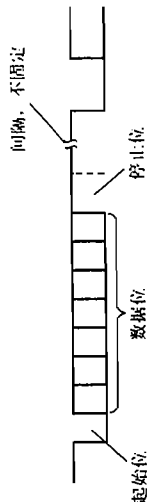


图 10-4 字符同步方式

帧起始	控制信息	数据	校验	帧结束
8 位	m 位	0~n 位	8~32 位	8 位

图 10-5 帧同步方式

10.2.4 编码与调制

在数字通信中,通常需要进行编码与调制。编码是用数字信号来表示数字或模拟数据,而调制是用模拟信号来表示数字或模拟数据。

对模拟数据进行数字编码的目的是在数字信道上发送模拟数据。对模拟数据的数字编码必须满足采样定理:如果模拟信号的最高频率为F,若以2F的采样频率对其采样,则采样得到的离散信号序列就能完整地恢复出原始信号。模拟数据的数字编码一般需经过3个步骤:采样(按一定时间间隔对模拟信号进行采样)、量化(对每个样本舍入到量化级别上)和编码(对每个舍入后的样本进行编码)。图10-6所示是对语音信号进行脉冲调制(Pulse Coded Modulation, PCM)编码的过程。

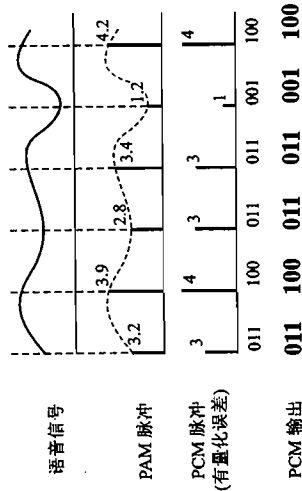


图 10-6 语音信号的 PCM 编码过程

将用数字数据进行模拟调制的目的是为了在模拟信道中传输数字信号。常用的3种调制技术为:幅移键控(Amplitude Shift Keying, ASK)、频移键控(Frequency Shift Keying, FSK)和相移键

控(Phase Shift Keying, PSK),如图10-7所示。若载波为 $S(t) = A\cos(\omega t + \psi)$,其参量包括幅度 A 、频率 ω 、初相位 ψ ,调制就是要使 A 、 ω 或 ψ 随数字基带信号的变化而变化。ASK 用载波的两个不同振幅表示 0 和 1;FSK 用载波的两个不同频率表示 0 和 1;而 PSK 用载波的起始相位的变化表示 0 和 1。

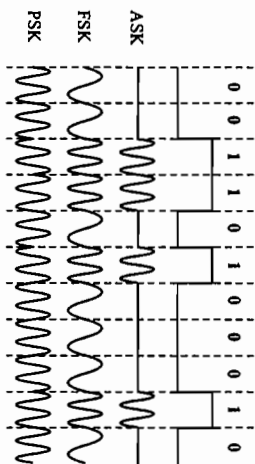


图 10-7 ASK、FSK 和 PSK 的工作原理

10.2.5 多路复用技术

多路复用是为了实现多个信息源共享一个公共信道,以提高线路的利用率。一般当信道的传输能力大于每个信源的平均传输需求时,可考虑采用多路复用。多路复用一般有以下几种方式:频分多路复用(Frequency Division Multiplexing, FDM)、时分多路复用(Time Division Multiplexing, TDM)和波分多路复用(Wave Division Multiplexing, WDM)。

FDM 的原理如图 10-8 所示,整个传输频带被划分为若干个频率通道,每路信号占用一个频率通道进行传输。频率通道之间留有防护频带以防相互干扰。

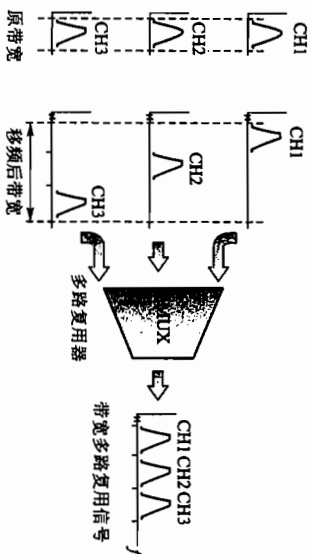


图 10-8 频分多路复用的原理

TDM 的原理如图 10-9 所示,把时间分割成小的时间片,每个时间片又分为若干个时隙,每路数据占用一个时隙进行传输。TDM 的每路数据总是使用一个固定的时隙,所以也称为同步时分多路复用。在 TDM 系统中,若某用户无数据发送,其他用户也不能占用该时隙,将会造成带宽的浪费。

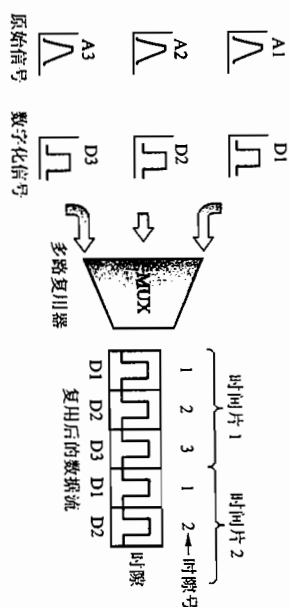


图 10-9 时分多路复用的原理

统计 TDM (Statistical TDM, S-TDM) 改变了时隙的静态分配策略,允许以先来先服务或基于优先权的原则,将时隙动态分配给现有数据传输的信源,如图 10-10 所示,从而解决了带宽浪费的问题。S-TDM 也称为异步时分多路复用。

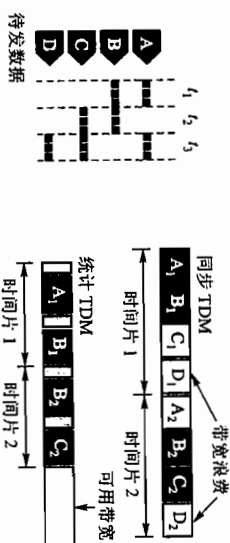


图 10-10 S-TDM 的原理及与 TDM 带宽应用的比较

WDM 的原理如图 10-11 所示,把整个波长频带划分为若干个波长范围,每路信号占用一个波长范围进行传输,与 FDM 的原理相似。

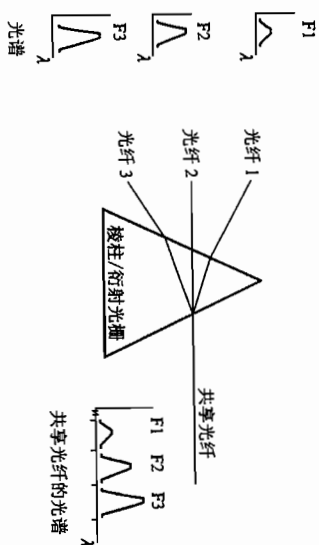


图 10-11 波分多路复用的原理

除上述的多路复用技术外,在无线通信中还有:按需分配多路访问(Demand Assignment Multiple Access, DAMA)、码分多路访问(Code Division Multiple Access, CDMA)等。在宽带技术中,通常会采用逆向复用(也称为线路聚合、line aggregation),其作用正好与多路复用相反,将几个低速信道合并成一个高速信道。

10.2.6 信号带宽、信道带宽与差错控制

信号带宽是信号频谱的宽度,也就是信号的最高频率分量与最低频率分量之差。而信道带宽则限定了允许通过该信道的信号下限频率和上限频率,也就是限定了一个频率通道。如为了无失真地进行数据传输,信号带宽必须在信道带宽之内,即信号的最高频率分量必须小于等于信道的上限频率;同时,信号的最低频率分量必须大于等于信道的下限频率。

在某些特殊情况下,可以不遵守上述原则,即允许存在失真,如电话通信。一般电话信道的频率范围从300 Hz~3 300 Hz,带宽为3 kHz,而语音信号频谱则一般在100 Hz~7 kHz的范围之内。电话信道会使语音信号的高频和低频部分滤掉。由于语音信号的主要能量集中在中心的一些频率分量附近,所以通过电话信道传输的语音信号,虽有失真,但仍能分辨。

在通信过程中,由于干扰的存在必定会引起通信中出现差错。产生差错的原因主要有:信号衰减和热噪声;信道的电气特性引起信号幅度、频率、相位的畸变;信号的反射与串扰;冲击噪声、闪电、大功率电机的启动和停止等。控制差错的基本方法是:接收方进行差错检测,并向发送方应答,告知是否正确接收。差错检测主要有两种方法:奇偶校验(Parity Checking)和循环冗余校验(Cyclic Redundancy Check, CRC)。差错控制的技术主要有:自动请求重传(Automatic Repeat Request, ARQ)、停等 ARQ、Go-back-N ARQ 和选择重传 ARQ。

由于自动纠错编码的开销较大,在一般的应用中不使用自动纠错。

10.3 分布系统与计算机网络

计算机通信的应用方式可分为远程处理和分布处理两种类型。早期的远程处理称为计算中心方式,即通过通信线路将作业(如用户程序)从远程终端输入主机,运算后返回结果。现在已发展到自治的计算机之间的远程信息处理,例如远程信息检索、数据库访问、WWW(World Wide Web)服务、电子商务、视频会议等。分布处理方式是利用并行处理技术,将一个大型作业分散于一组计算机进行处理,以增强各计算机的处理能力。

事实上,分布系统是一个软件系统,建立在网络之上。软件给予网络高度的内聚性和透明性,因此网络与分布系统之间的区别在于软件(特别是操作系统)而不是硬件。另外,网络与分布系统之间有许多重叠的问题。例如,两者都要移动文件,但其区别是由系统还是由用户发起的。

分布系统的高级形态可称为多机系统,多机系统专指同一地点的许多大型主机互连组成的功能更加强大、能高速并行处理的计算机系统。对这种系统互连的要求是高带宽和多样的连通性。由于考虑了远距离传输的特点(如需要差错、流量的控制,运行复杂的路由选择算法等),计算机网络中的信息传输往往开销很大,因而实际上的有效数据传输速率要比通信线路能够提供

的带宽小得多。而在近距离的多机互连系统中这些开销实际上都是不必要的。多机系统对互连的另一个要求是灵活的连通性,即使做不到完全直接相连,也要求能够快速地进行交换互连模式以适应并行计算的要求。由于距离的原因,在计算机网络中的互连是相当有限的,为弥补这一缺陷发展了各种交换技术。

按照计算机网络通信协议的方式处理多机系统中的高速并行计算是不合适的,计算机网络和多机系统中的互连网络在理论和方法上是完全不同的两个学科。

计算机网络是指能够交换信息且独立自治的计算机及通信子网的集合。计算机网络涉及3个方面的问题:

- ① 必须有两台或两台以上自治的计算机互相连接起来才能构成网络。
- ② 需要一条通道(或通信子系统)将这些计算机连接起来。
- ③ 计算机之间要交换信息,彼此之间就需要有某些规定和约定,这就是通信协议(protocol)。

计算机网络由通信子网和资源子网两部分组成,通信子网负责数据的无差错和有序传递,包括通信处理器、集中器、复用器、路由器、网桥等设备。资源子网是计算机网络的本地系统环境,包括主机、终端和应用程序等。资源子网的主要功能是:用户资源配置、数据的处理和管理、软件和硬件共享以及负载均衡等。当然,还必须包括一个合适的接口,以保证主机能无缝地(seamless)接入一个通信子网。

只要有一个合适的接口,并遵守一定的协议,现有的或未来的电信网络,包括公用电话交换网(PSTN)、公用交换数据网(PSDN)、综合业务数字网(ISDN)、宽带综合业务数字网(B-ISDN)、无线蜂窝网、卫星通信网等,都可以用作计算机通信网的通信子网。

网络的组成元素可分为网络结点和通信链路两部分。网络结点又分为端结点和交换结点两种。端结点指通信的源结点和宿结点,例如用户主机和用户终端;交换结点指网络通信过程中起控制、暂存和转发信息作用的结点,例如程控交换机、集中器、路由器、网桥等。端结点是用户数据的发源地和归宿地,而交换结点本身不向用户提供新的数据,只是对用户数据进行传输。通信链路指传输信息的通路,即通信子网中的传输介质,其作用是在机器之间传递比特。因此,通信子网主要由交换结点(包括驻留其中的通信软件)和通信链路组成,而资源子网中主要包括拥有资源的用户主机和请求资源的用户终端,它们都是端结点。但必须明确的是,通信子网和资源子网是从计算机网络的功能上进行划分的,而结点和链路则是从物理组成角度来分类的。应该说,端结点(如用户主机)中所包含的通信子系统是属于通信子网的范畴。

1. 网络拓扑结构

所谓网络的拓扑结构是指通信子网中交换结点的互连模式。按拓扑结构进行划分,网络可分为集中式和分布式两种基本类型。

在集中式网络内,所有的信息流必须经过一个中央处理设备(即交换结点)。通信链路都从中央交换结点向外辐射,这个中央结点的可靠性基本上决定了整个网络的可靠性。集中式网络又称为星状网。若很多个主机或终端较集中配置在某处时,可采用集中器或复用器。集中器具有存储功能,因而其输入链路容量的总和可超过输出链路的容量,而复用器的输入链路的总容量不能超过其输出链路的容量。具有集中器或复用器的集中式网络属于共享信道的广播式网络。

由于只有单一的信道,所以一个时刻只能有一个站(主机或终端)发送信息,因此所有要发送信息的站必须通过某种仲裁协议分时地使用信道。

分布式通信子网又称为网络网,即其中任何一个结点都可以和若干个结点直接相连。在分组交换的情况下,即使某个结点或某条链路发生故障,用户发送的数据流也可以经过不同的路径到达目的地,因此分布式网络的可靠性较高。现在一些网络常把主干网络做成分布式的,而非主干网络则做成集中式的。

2. 传输技术

从广义上讲,有两种类型的传输技术:点到点网络和广播网络。点到点网络在每一对源/宿端之间都有一条传输通路。传输的消息按照某种格式组织成分组或包,为了能从源到目的地,分组必须通过一个或多个中间交换结点。通常可能有多条通路,并且其长度也不一样,因此点到点网络中的路由(routing)算法十分重要。

广播网络仅有一条通信信道,由网络上的所有结点共享。分组可以由任何结点发送并被其他所有的结点接收。为了指明分组应被哪个目的结点接收,使用地址来识别,即每个分组都包含地址字段。一旦收到分组,各个端结点将检查它的地址字段,如果是发送给自己的,则处理该分组,否则将分组丢弃。

广播网络通常也允许在地址字段中使用一段特殊代码,以便将分组发送到所有目标结点。使用此代码的分组发出以后,网络上的每个结点都会接收和处理它,这种方式被称为广播(broadcasting)。某些广播系统还支持向所有结点的一个子集发送分组的功能,称为组播或多点传送(multicasting)。一种常见的方案是保留地址字段的某一位来指示组播方式,而地址字段中的剩余位存放组号。每台主机可以注册到任意的组或所有的组。当某一分组被发送给某个组时,它被发送到所有注册到该组的主机。

3. 网络规模

将计算机网络分类的另一个标准是它的连接距离,即网络规模。表10-1列出了按连接距离分类的各种网络。

表 10-1 按连接距离分类的各种网络

网络规模	处理器之间的距离	多个处理器的位置	应用实例
微型网	0.1 m	同一电路板	数据流机器
小型网	1 m	同一系统	多计算机系统
中型网	10 m	同一房间	局域网
	100 m	同一建筑物	
	1 km	同一园区	
大型网	10 km	同一城市	城域网
	100 km	同一国家	广域网
	1 000 km	同一洲内	互联网
	10 000 km	同一地球上	

真正意义上的计算机网络是指局域网(LAN)、城域网(MAN)和广域网(WAN)。网络由于其规模大小的不同而具有各自的一些特点,它们在使用范围、拓扑结构、传输介质以及介质接入控制等方面均可能有所不同。

(1) 局域网

局域网是处于同一建筑物、同一校园或方圆数公里地域内的专用网络。局域网常被用于连接公司办公室或大学里的个人计算机和工作站,以便共享资源(如打印机)和交换信息。

局域网的覆盖范围比较小,这意味着即使在最坏的情况下其传输时间也是有限的,并且可以预先知道传输的最大时间。这就简化了网络的设计和网路的管理。

局域网中使用的传输技术是用一条电缆连接所有的机器。传统的局域网传输速率为10 Mb/s~100 Mb/s,传输时延较短(数十个毫秒),并且出错率低。新研制的局域网传输速度更高,可达到每秒千兆比特的数量级,如千兆以太网(Gigabit Ethernet)。

局域网的拓扑结构通常采用广播式网络,其中常用的两种网络结构为总线型和环状结构。

(2) 城域网

城域网基本上是局域网的扩展,且通常采用相似的技术。它可能覆盖一组邻近的公司办公室和一个城市,可以提供数据、语音,乃至有线电视业务。城域网中仅使用一条或两条电缆,并且不包含交换结点,即把分组分流到几条可能的引出电缆的设备,这样做可以简化设计。

城域网标准之一是分布式队列双总线(DQDB),即IEEE 802.6。DQDB由两条单向总线组成,所有的计算机都连接在上面。每条总线都有端点,这是一个启动传输活动的设备。目的计算机在发送者右方时使用上面的总线,在左方时则使用下面的总线。

(3) 广域网

广域网覆盖的地域范围较大,包括一个国家或某一大陆。在广域网中,通常将连入通信子网的这些机器称为端系统(end system),有时也称为主机(host)。

在大多数广域网中,网络包括大量的电缆或电话线,每一条都连接一对交换结点。如果两个交换结点之间没有直接相连,则必须通过其他中间结点连接。当分组通过中间结点时,将被完整地接收并缓存下来,在需要的输出链路空闲时,该分组就被转发出去。使用这种原理的通信子网被称为存储-转发(store-and-forward)子网。

4. 网络互连与网络软件

计算机网络往往由许多种不同类型的网络互相连接而成。如果几个计算机网络只是在物理上连接在一起,它们之间并不能进行通信,那么这种“互连”并没有什么实际意义。因此在谈到“互连”时,就已经暗示这些相互连接的计算机是可以进行通信的,也就是说,从功能上和逻辑上看,这些计算机网络已经组成了一个大型的计算机网络,或称为互联网(internet),也可简称为互联网、互连网。世界范围的因特网(Internet)就是互联网的著名例子。

当一个网络上的用户需要和另一个网络上的用户通信时,就需要连接不同的、而且往往是不兼容的网络。有时候使用称为网关(gateway)的专用计算机来完成连接,并提供硬件和软件的转换。互连的网络的集合就称为互联网(internet)。常见的互联网是通过广域网连接起来的局域网的集合。

需要说明的是,互联网(internet)这个词通常只代表一般的网络互联的意思,而Internet(因

特网)是指特定的由 ARPANet 发展起来的互联网,它被广泛地用于连接大学、政府机关、公司和个人用户。请注意,虽然 internet 和 Internet 只是第一个字母的大小写不同,但表示的是两个不同的意义。

与计算机网络有关的软件部分大致可分为 5 类:操作系统核心软件、通信控制用协议软件、管理软件、交换与路由软件以及应用软件。计算机网络通过应用软件为用户提供网络服务,即信息资源的共享和传输。应用软件可分为两大类,一类是由网络软件厂商开发的通用应用软件,诸如远程登录、电子邮件、Web 服务器及相应的浏览、搜索工具等。另一类应用软件则是依赖于不同的用户业务的,例如网络上的电子银行、电信管理、制造厂商的分布式控制与操作。

5. 协议的体系结构

通信协议(简称协议)是指相互通信的各方对如何进行信息交换所一致遵循的一整套规则。这正如人们交谈时要使用相同的语言一样。一个网络有一系列的协议,每一个协议都完成某一特定的任务。总的来说,协议的作用是完成计算机之间有序的信息交换。主要有以下几种通信协议。

(1) 面向应用的协议

为完成某些特定应用而制定的协议。从分层的协议结构来看,包括应用层的专用服务,例如, TCP/IP 协议模型中的 FTP、TELNET、SMTP 等, OSI/RM 模型中的 FTAM、VT、JTM、MHS、DS 等。面向应用的协议也可称为进程到进程的协议。

(2) 系统到系统的协议

支持端系统中特定应用进程之间的数据交换,为应用进程提供网络通信服务,包括应用层的公用服务元素、表示层协议(如完成数据的加密/解密)等。

(3) 端到端的协议

完成端到端的可靠传输,建立、保持及维护端到端连接,包括会话层协议(保持对话的同步等)、TCP 和 UDP 等。

(4) 点到点的协议

实现直接相连的结点之间的数据传输,如高级数据链路控制协议(HDLC)和 Internet 中的 PPP 协议等。

(5) 网络接入协议

如媒体访问控制协议(MAC)等。

(6) 其他网内的协议

包括流控协议、寻址和路由控制等。

(7) 网际互连协议

从网络介质的延伸,不同网络用户的寻址,到异构网络间协议的转换等。

任何计算机(端系统)之间的通信,实际上都是通过在不同层次上的对等通信单元(在分层体系结构中称为“实体”)的相互协调来完成的。对计算机网络的研究都是基于这种功能上的通信概念,因而形成了分层的协议体系结构。

计算机网络的整套协议是一个庞大复杂的体系,为了便于对协议的描述、设计和实现,现在都采用分层的体系结构(architecture)。所谓体系结构,是指计算机网络的分层、各层协议和层间

接口的集合,也就是网络及其部件所应完成的功能的精确定义。因此体系结构是计算机网络的一种抽象的、层次化的功能模型。

分层的协议体系结构将庞大复杂的协议分成不同的层次,每一层的功能是不同的,低层功能主要用来对用户提供通信连接,而高层功能能保证数据以正确的形式进行交换并有序地处理数据。

分层结构的基本思想是把整套协议体系分成一些小块(层),如同结构化程序设计中的子程序模块。下一层对其上一层提供服务,每一层本身的功能与下层提供的服务叠加到一起,从而使最高层能为用户提供完整的一组服务,以便实现通信或分布应用。

分层的基本原则是定义每一层向上一层提供的服务,来保证每层的功能相互独立,但不规定如何完成这些服务,这样就允许每一层改变其实现所规定的服务的方法,但要求提供的服务不变,而且不影响其他层次。相反,若不采用分层式结构,某些细小的变化往往会导致修改整套网络软件。

早期各主要的网络公司相继开发了自己的网络体系结构,如 IBM 的系统网络体系结构(System Network Architecture, SNA)与数字设备公司的数字网络体系结构(Digital Network Architecture, DNA)。1974 年国际标准化组织(International Organization for Standardization, ISO)开发了开放系统互连参考模型(OSI/RM—Reference Model for Open Systems Interconnection, OSI/RM),简称为 OSI,其各层的主要功能如表 10-2 所示。

表 10-2 开放系统互连参考模型各层的主要功能

层 次	功 能
7. 应用层	提供电子邮件、文件传输等服务
6. 表示层	转换数据格式,数据加密和解密
5. 会话层	通信同步,错误恢复和事务操作
4. 传输层	网络决策,实现分组和重新封装
3. 网络层	路由选择,计费信息管理
2. 数据链路层	错误检测和校正,组帧
1. 物理层	数据的物理传输

10.4 虚拟专用网络

虚拟专用网络(Virtual Private Network, VPN)是依靠 ISP 和其他 NSP(网络服务提供商),在公用网络中建立专用的数据通信网络的技术。虚拟专用网能够实现专用网的功能,但它不是真正的专用网络。在虚拟专用网中,任意两个结点之间的连接并没有传统专用网所需的端到端的物理链路,而是利用某种公众网的资源动态组成的。它可以在 Internet 上的位于不同地方的两个或多个企业内部网之间建立一条虚拟的企业内部专线。

虚拟专用网的解决方案有 4 种:最常用的是虚拟专用拨号网络(VPDN),就是基于拨号的

VPN;第二种是虚拟专用路由网络(VPRN),是基于路由的VPN;第三种是虚拟租用线路(VLL),是基于虚拟专线的VPN;最后一种是虚拟专用局域网子网段(VPLS),是基于局域网仿真的VPN。它们之间最根本的区别在于数据包在接口(如桥和路由器)之间不同的转发方式。

目前VPDN的应用比较广泛。用户利用拨号网络访问企业数据中心,并从企业数据中心获得一个私有地址,但用户数据可通过公共数据网络进行传送。现阶段采用宽带上网(ADSL)+软件VPN的方案是解决中小企业实现网络互联的经济实用的好办法,如图10-12所示。

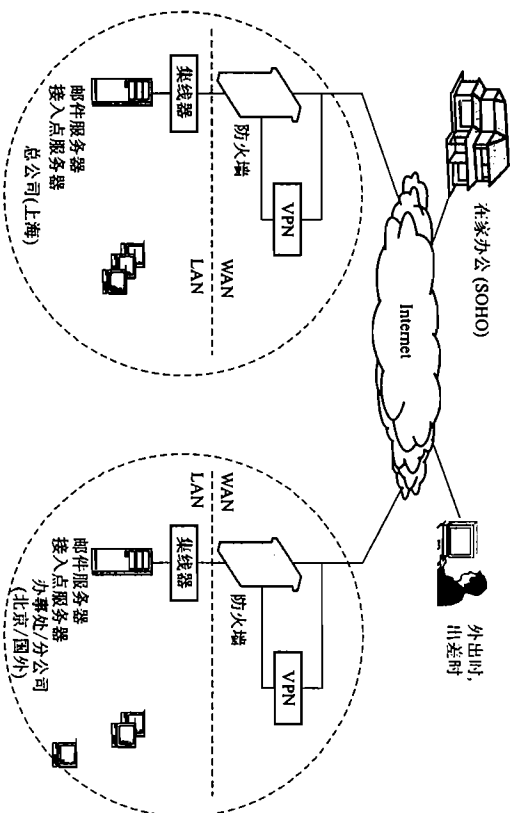


图 10-12 VPN 实现的基本原理

以下简要介绍纯软件VPN产品——Sinfor 网网通。使用Sinfor 网网通能够让企业总部与分支机构的局域网轻松实现安全互联,本地局域网通过Internet得到无限的延伸,实现真正意义上移动办公。原本只能用于单一局域网内的系统或者功能,能够扩展到多个局域网中,实现公司数据和信息的同步和统一,增加企业的效率,提高竞争力。实施软件VPN方案的优势和特点:

- ① 纯软件VPN产品,保护原有投资,无需增加任何硬件设施。
- ② 无需改动原有应用系统和网络结构,保证企业正常的运作方式不受影响。
- ③ 支持各种上网方式(采用ADSL、宽带上网方式效果更佳),不需要再拉DDN专线或申请固定IP,支持两端动态IP寻址,为企业节省大量的通讯费用。
- ④ 网络平台的适应性强,扩展性好,支持任何C/S、B/S结构软件。
- ⑤ 经济实效、使用简单方便,无须专业人员维护。
- ⑥ 采用先进的加密算法防止数据被窃听和篡改。
- ⑦ 采用密码身份认证和硬件捆绑接入认证,除了密码正确外,还必须是总部授权的硬件设备才能接入。防止非法用户或知道密码的辞职员工接入公司网络。

10.5 调制解调器

调制解调器(Modem)由调制器(MODULATOR)和解调器(DEMODULATOR)合并而成。调制器的作用是将数字信号转换为模拟信号,在模拟信道上进行传输;解调器的作用是将模拟信号转换为数字信号。

10.5.1 模拟调制解调器

模拟Modem是利用电话网模拟交换线路实现远距离数据传输的传统技术。从传输速率为300 bps的Bell 103 Modem到33.6 Kbps的V.34bis Modem,经过了数十年的发展历程。近年来随着Internet的迅猛发展,57.6 Kbps的V.90/V.92 Modem(俗称56 Kbps Modem,又称PCM Modem)应运而生。

56 Kbps Modem与传统的Modem在应用上有较大的区别,仅在用户电话与程控交换机间使用一对Modem,交换机与ISP之间为数字连接。56 Kbps Modem有两个关键技术:一是多电平映射调制技术,二是频谱成型技术。多电平映射调制是采用一组PAM调制,从A律(或μ律)PCM编码256个电平中选择部分电平进行调制星座映射,调制符号率为8 kHz。使用频谱成型技术,目的是抑制发送信号中的直流分量,减少混合线圈中的非线性失真。

早期的56 Kbps Modem主要有两大工业标准:一个是X2标准,另一个是K56flex标准,两者互不兼容。ITU-T第16研究组(SG16)1998年9月全会正式通过了V.90建议:“用于PSTN(Public Switched Telephone Network,公用电话交换网)上的,上行速率为33.6 Kbps,下行速率为56 Kbps的数字/模拟调制解调器”。已投入使用的X2或K56flex Modem均可以通过软件升级的方法实现与V.90 Modem的兼容,V.92是V.90的升级版,将上行速率提升为48 Kbps。

图10-13所示为33.6 Kbps和56 Kbps Modem在用户和ISP间的不同的连接关系,其主要区别在于前者只是在交换机间采用了数字信道,而后者只是在用户和交换机间存在模拟信道。

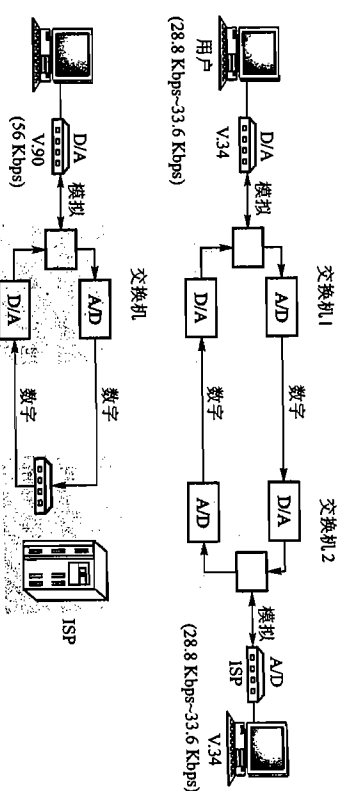


图 10-13 33.6 Kbps 和 56 Kbps Modem 的不同连接方式

模拟调制解调器一般可分为内置式和外置式两种。外置式的接口一般有三种。

① RS-232 接口,图 10-14 所示的即为 RS-232 接口的外置式 56 Kbps Modem。

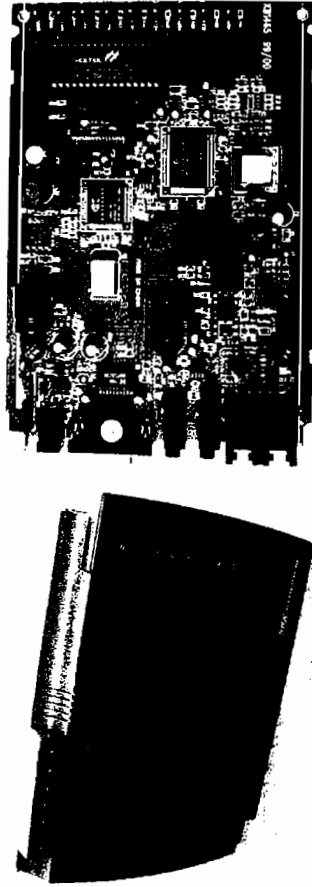


图 10-14 RS232 接口的外置式 modem

② PC 接口 (PCMCIA, 笔记本电脑使用的一种标准接口),图 10-15 所示的为 10 Mbps/100 Mbps 网卡和 56 Kbps Modem PC 卡。

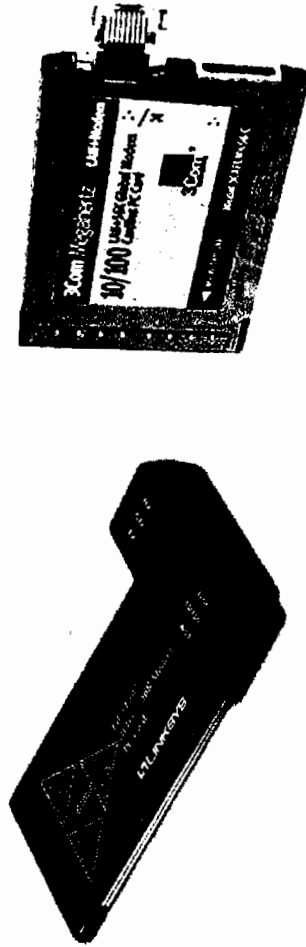


图 10-15 10 Mbps/100 Mbps 网卡和 Modem 三合一 PC 卡

③ USB 接口,如图 10-16 所示。

内置式目前都为 PCI 总线接口,早期的 ISA 总线接口已经淘汰。图 10-17 所示的为 PCI 总线接口的内置式 56 Kbps Modem。

10.5.2 DSL 技术

固定电话通过网络接口设备 (Network Interface Device, NID) 连至最近的电话公司交换中心或中心局 (Central Office, CO),这一段线路被称为本地环。它们一般是由铜双绞线构成的,因此也被称为铜环,通常也称为最后一英里。

数字用户线路 (Digital Subscriber Line, DSL) 技术利用现有的本地环实现数据的高速传输。

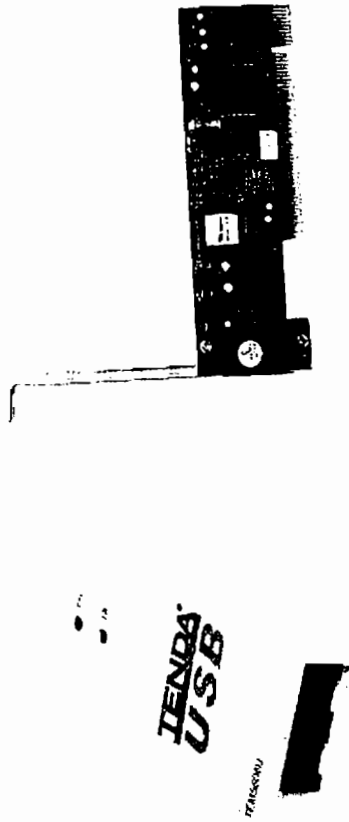


图 10-16 USB 接口的外置式 Modem

图 10-17 PCI 接口的内置式 Modem

DSL 与距离紧密相关,而且依赖于不断变化的环路质量。由于 DSL 采用现有的电缆,因此它能够为本电话公司带来经济效益。

由于本地环的连接是非共享的,因此 DSL 提供了专用的带宽,能够提供更加稳定的传输速率。另外,专用连接非常安全,DSL 的专用本地环不容易受到窃听的攻击。

将电话线同时用于普通的电话服务和高速 Internet 访问是 DSL 的另一优点。当使用调制解调器访问 Internet 时,信息将通过 PSTN 进行传输,电话呼叫也需要经过 PSTN,两种功能无法同时实现。而 DSL 将语音呼叫路由连接至 PSTN,而将 Internet 访问路由连接至一个独立的网络,这是因为 DSL 的数据传输频率与普通电话服务所使用的频率不同。因此,可同时完成上述两种功能。

某些 DSL 提供商提供一个或多个永久的 Internet 协议 (IP) 地址,以及一个长期开放的 DSL 连接。这种连接称为专线接入,即用户拥有固定的静态 IP 地址,24 小时在线。有些提供商使用临时的 IP 地址,必须使用附加的软件才能获得临时的 IP 地址并建立会话。该软件通常使用 Microsoft 拨号网络 (Dial-Up Networking, DUN) 和点对点协议 (Point-to-Point Protocol over Ethernet, PPPoE) 建立会话。这种连接称为虚拟拨号入网,并非真正的电话拨号,需要用户输入账号、密码,通过身份验证后获得一个动态的 IP 地址,因此用户可以掌握上网的主动性。

1. DSL 技术的分类

DSL 技术包含几种不同的类型,通常称为 xDSL。其中 x 为标识性字母。例如,ADSL (Asymmetric Digital Subscriber Line, 非对称数字用户线路)、SDSL (Symmetric Digital Subscriber Line, 对称数字用户线路)、HDSL (High bit-rate Digital Subscriber Line, 高比特率数字用户线路)、VDSL (Very high bit-rate Digital Subscriber Line, 极高比特率数字用户线路)、SHDSL (Super high bit-rate Digital Subscriber Line, 超高比特率数字用户线路)、RADSL (Rate Adapted Digital Subscriber Line, 速率自适应数字用户线路) 和 IDSL (Integrated Digital Subscriber Line, 综合数字用户线路)。另外,还出现了 ADSL 的变种: ADSL Lite、G.Lite 或 UADSL (Universal ADSL),以及 HDSL 的变种: HDSL2。

每种类型的 DSL 能够满足不同的需求,它们具有各自的特性:ADSL、ADSL Line 和 RADSL 是非对称的,即上行链路和下行链路的传输速率不相同,而有些是对称的,如 SDSL 和 HDSL。对称 DSL 较适合于企业,因有如文件传输、视频会议等收发数据量大致相同的工作;有些 DSL 能够与模拟语音共享电话线,而有些则不行。大多数 DSL 只需要与普通电话线相同的一对铜线,而 HDSL 需要两对。

(1) HDSL

HDSL 是一种对称的高速数字用户线路技术。通过两至四对双绞铜线提供全双工 1.544 Mbps/2.048 Mbps(T1/E1)的数据信息传输能力。HDSL 技术最早由 Bellcore 提出,近年来国内外已有众多厂家推出了 HDSL 产品并得到应用。HDSL 接入技术的标准化已取得突破性进展,1998 年 10 月 ITU-T 第 15 研究组(SCG15)全体会议通过了关于 HDSL 的新建议 G.991.1,对本地金属线路上的高速数字用户线路系统做了详细规范,采用 2B1Q 或 CAP 两种线路编码方式,可以支持 604 Kbps、1 168 Kbps 和 2 320 Kbps 三种线路传输速率,但不支持简易老式电话业务(Plain Old Telephone Service, POTS)、模拟数据和 ISDN。HDSL 的无中继传输距离约为 4 km~7 km。

(2) ADSL

ADSL 允许在一对双绞铜线上,在不影响现有电话业务的情况下,进行非对称高速数据传输。ADSL 上行速率为 224 Kbps~640 Kbps,下行传输速率为 1.544 Mbps~9.2 Mbps;传输距离为 2.7 km~5.5 km。ADSL 是在用户电话线上采用分离器(splitter)的办法将模拟语音通道与数字调制解调器分开,即使在 ADSL 连接失败时也不影响语音服务。

ITU-T SCG15 1998 年 10 月原则通过了关于 ADSL 的 G.992.1 和 G.992.1 建议草案,1999 年 6 月 SCG15 全会通过。G.992.1 规范了带分离器的 ADSL 系统,利用该系统可在同一对金属双绞线对上传输高速数据和模拟信号,采用的线路编码为 DMT,下行速率为 6.144 Mbps,上行速率为 640 Kbps。G.992.2 规范了不带分离器(splitterless)的简化的 ADSL(ADSL Lite),具有成本低、安装简便的优点,也采用 DMT 线路编码,下行速率为 1.536 Mbps,上行速率为 512 Kbps。

(3) SDSL

SDSL 使用一对铜双绞线对在上、下行方向上实现 E1/T1 传输速率的技术,也称为单线对数字用户线路(Single pair Digital Subscriber Line, SDSL),是 HDSL 的一个分支,有时也称为中等比特率数字用户线路(Middle bit rate Digital Subscriber Line, MDSL)。SDSL 采用 2B1Q 线路编码,上行与下行速率相同,传输速率由几百 Kbps 到 2 Mbps,传输距离可达 3 km 左右。SDSL 的发展趋势主要有两个:一是开发在单线对上同时传送语音和数据的 HDSL,用于小型办公室/家庭办公室(SOHO);二是开发具有更高传输速率的单线对数字用户线路技术。ITU-T 已就起草新的建议 G.shdsl“单线对高比特率数字用户线路(VDSL)”开展研究。

(4) RADSL

RADSL 是在 ADSL 基础上发展起来的新一代接入技术,其下行速率从 384 Kbps 到 9.2 Mbps,上行速率从 128 Kbps 到 768 Kbps,传输距离可达 5.5 km 左右。RADSL 能够自动地、动态地根据所要求的线路质量调整传输速率,为远距离用户提供质量可靠的数据网络接入手段。正是 RADSL 的这些特点使 RADSL 成为用于网上高速冲浪、视频点播(VoD)、远程局域网(LAN)访问的理想技术,因为在这些应用中用户下载的信息往往比上传的信息(发送指令)多得多。

(5) VDSL

VDSL 是 ADSL 的发展方向,是目前最先进的数字用户线路技术。VDSL 通常采用 DMT 调制方式,在一对铜双绞线上实现双向数字传输,其下行速率可达 13 Mbps~52 Mbps,上行速率可达 1.5 Mbps~7 Mbps,传输距离约为 300 m~1.3 km。

(6) SHDSL

SHDSL 是目前仍处在实验研究中的传输速率高达 155 Mbps 的数字用户线路技术。其传输距离只能达到几十米,实用前景不大。

(7) HDSL2

就 DSL 技术本身而言, HDSL2 算是后起之秀,它的出现使非对称数字用户线(ADSL)和对称数字用户线(SDSL)的缺陷,如上行方向的传输速率仅几百个 Kbps、需要两条双绞线、作用距离近等问题得到较好的解决。HDSL2 是第一代标准化的数字业务。它之所以像 ADSL 一样成为 DSL 技术发展的一种主流趋势,是因为提供 DSL 服务的厂商认识到标准化的重要性。工程技术人员经过两年的共同努力,解决了 DSL 有保证的数据传输速率、有保证的公共服务区有效半径、与其他 DSL 兼容、所有业务在一条双绞线上传送等技术难题,使标准化业务的性能远远超过其他的 DSL 业务。

2. ADSL 设备

ADSL 的非对称特性使得它非常适于 Internet 通信量,因为 Internet 通信量对下行带宽的需求通常远远超出对上行带宽的需求。ADSL 已是国际标准,产品具有互操作性。ADSL 采用离散多音频(Discrete Multi Tone, DMT)或 FDM 技术实现宽带通信。其频谱分布如图 10-18 所示。音频(Tone)也可称为基频或载频(Carrier Frequency)。

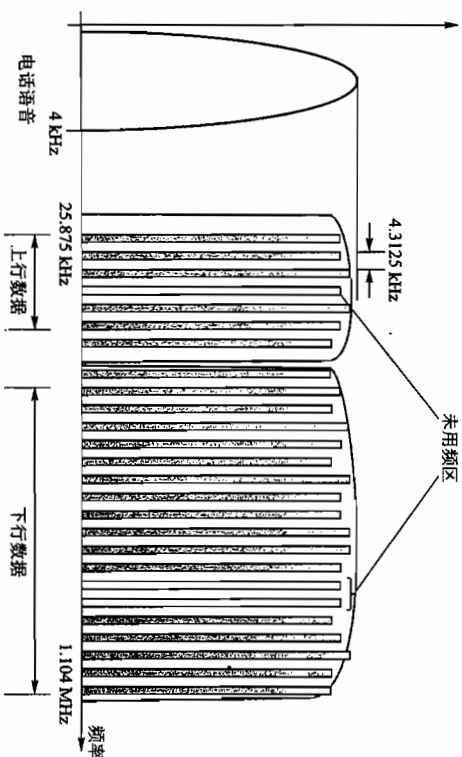


图 10-18 ADSL 频谱分布图

ADSL 的接入模型如图 10-19 所示,主要由中央交换局端模块和远端模块组成。中央交换局端模块包括在中心位置的 ADSL Modem 和接入多路复用器,处于中心位置的 ADSL Modem 被称为 ATU-C (ADSL Transmission Unit-Central)。接入多路复用器中心 Modem 通常被组合成一个接入结点,也称为 DSL 接入复用器 (DSL Access Multiplexer, DSLAM)。远端模块由用户 ADSL Modem 和分离器组成,用户端 ADSL Modem 通常被称为 ATU-R (ADSL Transmission Unit-Remote)。

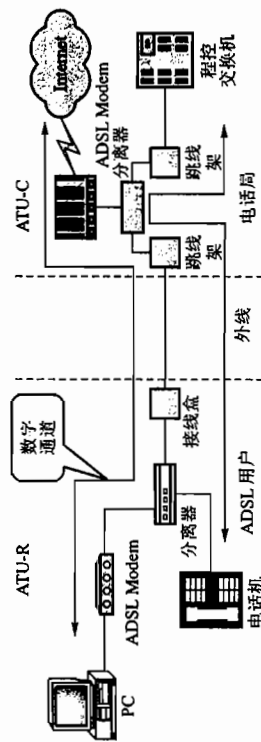


图 10-19 ADSL 的接入模型

ADSL 的用户终端设备主要有四种类型。

- ① ADSL 路由器: ADSL 路由器可为用户提供若干个 10 Mbps/100 Mbps IEEE 802.3 以太网接口, 且具有路由器功能, 支持 RIP、DHCP、PAP、CHAP、NAT 等协议, 用户的计算机通过以太网接口可以构成局域网。
- ② ADSL LAN Modem: 提供一个 10 Mbps/100 Mbps 以太网接口, 用户的计算机通过以太网接口连接。
- ③ ADSL USB Modem: 提供 USB 接口, 用户的计算机通过 USB 接口连接。
- ④ ADSL PCI Modem: 在 PCI 总线扩展槽中插网卡, 接口符合 PCI V2.2, 即插即用。

3. ADSL 设备的安装

ADSL 安装包括局端线路调整和用户端设备安装。在局端方面,由服务商在用户原有的电话线中串接 ADSL 局端设备,只需 2~3 分钟;用户端的 ADSL 安装也非常简易方便,只要将电话线线上分离器,分离器与 ADSL Modem 之间用一条两芯电话线上,ADSL Modem 与计算机的网卡之间用一条交叉网线连通即可完成硬件安装(目前网卡接口的外置式 ADSL Modem 较为普遍),再将 TCP/IP 协议中的 IP、DNS 和网关参数项设置好,便完成了安装工作,如图 10-20 所示。局域网用户的 ADSL 安装与单机用户没有很大区别,只需再加多一个集线器,用直连网线将集线器与 ADSL MODEM 连起来就可以了,如图 10-21 所示。

图 10-22 是以太网接口的外置式 ADSL Modem。

图 10-23 是 USB 接口的外置式 ADSL Modem。

图 10-24 是 PCI 总线接口的内置式 ADSL Modem。

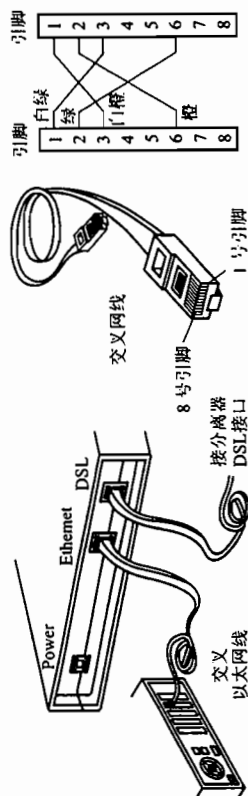


图 10-20 ADSL 用户端设备安装

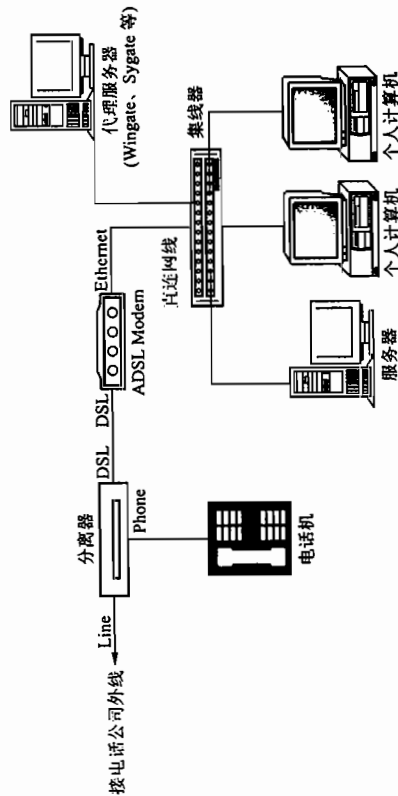


图 10-21 ADSL 局域网用户端设备安装

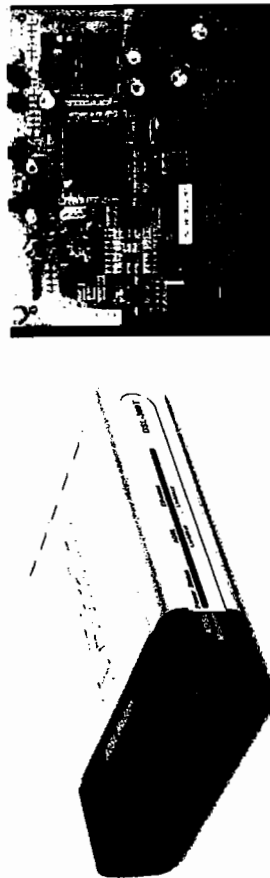


图 10-22 以太网接口的外置式 ADSL Modem



图 10-23 USB 接口的外置式 ADSL Modem

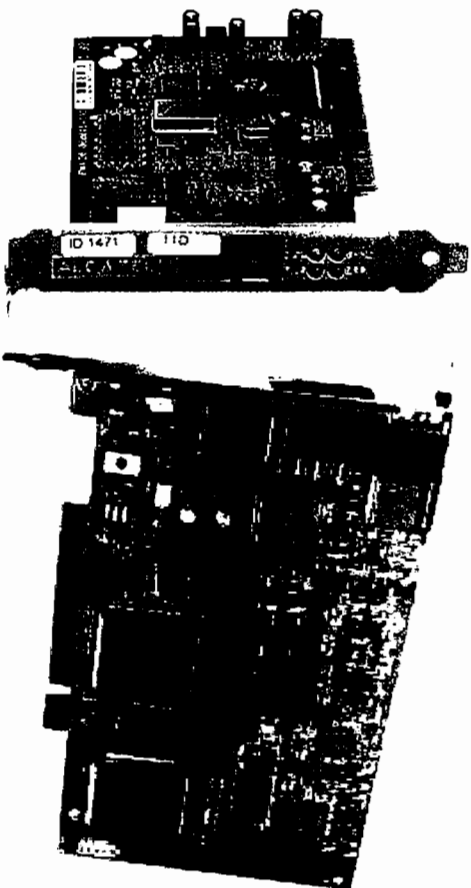


图 10-24 PCI 接口的内置式 ADSL Modem

4. ADSL 无线路由器

现在的家庭中都有多台计算机。一般大学的宿舍中也住着 4~6 人,随着电脑的价格的不断下降,很多的学生都购置了计算机。如果使用宽带(如 ADSL),大家共享接入,那么组建一个小型的局域网是必不可缺少的。这样不仅仅解决了共享一条宽带接入的问题,而且还可以共享资源,组建方案如图 10-25 所示。

图 10-26 为 DCWL-3604AR ADSL 无线路由器。DCWL-3604AR 结合“有线 ADSL”通用接入技术和“无线 WLAN”时尚移动接入,为用户提供前所未有的宽带无线服务。其提供 4 个 LAN 端口连接有线用户,提供 802.11b 无线局域网接入功能,使用户可以移动办公;提供 1 个打印服务器端口,将用户的普通打印升级成网络打印服务器。支持 NAT 和 DHCP,内置 PPPoE 拨号协议,可使多个局域网用户共享同一个 ADSL 线路访问 Internet, DHCP 功能动态分配 IP 地址,

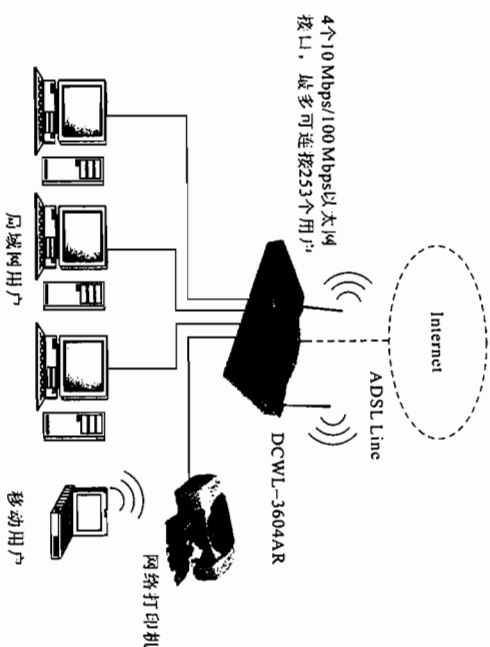


图 10-25 ADSL 无线路由器的应用

大大减少了网络的管理维护。提供防火墙功能可以做到 MAC 地址过滤、协议过滤、DMZ 等局域网安全措施。支持通过 SNMP 方式进行大规模网络的统一管理,可以应用于电信的热点地区覆盖、小型 SOHO 办公室、高档家庭住宅用户等用户。

D-Link 的 ADSL 无线路由器 DSL-604+, 如图 10-27 所示,支持 G.dmt 和 G.lite 两种 ADSL 标准,可自动适应不同标准的 ADSL 局端设备,方便地实现产品之间的互联互通。在连接方面,除了提供 1 个 ADSL 端口和 4 个 10 Mbps/100 Mbps 快速以太网端口的端口模式之外,还提供 1 个支持 802.11b + 协议、最高可达 22 Mbps 传输速率的无线接入点 (Access Point, AP),使无线网络在室内最大可以支持 100 m 覆盖范围,而在室外最大可覆盖 400 m。D-Link 的 DSL-604+ ADSL 无线路由器非常适合家庭组网,功能强大易于维护,无线、有线任意选择。

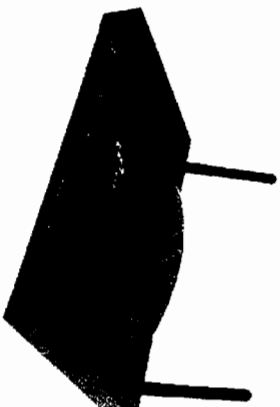


图 10-26 DCWL-3604AR

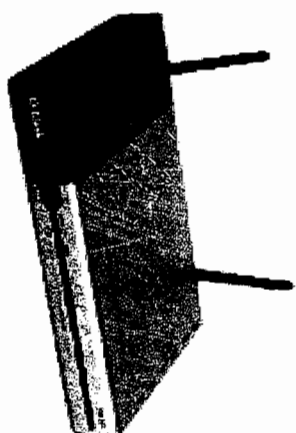


图 10-27 DSL-604+

10.5.3 电缆调制解调器

1. HFC和电缆调制解调器的基本概念

1998年3月,ITU-T第9研究组批准了一批新的J系列建议,其中J.93(有线电视系统中数字电视二次分配的条件接入要求)、J.112(交互有线电视业务的传输系统)和J.113(通过PSTN/ISDN的数字视频广播交互通道)等建议规范了HFC(Hybrid Fiber-Coaxial Network,混合光纤同轴网络,如CATV网络)的接入方式,为HFC交互业务的大规模商业应用铺平了道路,将有力推动HFC网向多功能应用的方向发展。依据网络的信号流向,将HFC网络分为单向HFC和双向HFC两种。

在有线网络中心前端、分前端或光结点上放置电缆调制解调器终端系统(Cable Modem Termination System, CMTS),用户可通过电缆调制解调器(Cable Modem)方式接入Internet。由于数据信号与电视信号所处的频段不同,因此用户同时还可通过传统的模拟电视机接收有线台前端传来的电视节目。

Cable Modem是利用HFC网络传输数据信息的设备,采用频带分段实现在5 MHz~855 MHz频率上同时传输模拟电视信号和数字信号。它使用与电视频道类似的带宽,以NTSC制式为例,下行在50 MHz~855 MHz频段中占用一个6 MHz的带宽,通常采用64/256 QAM(Quadrature Amplitude Modulation,正交调幅)调制,可提供高达36 Mbps的传输速率;上行在5 MHz~42 MHz频段中选择干扰较低的位置占用1 MHz~2 MHz带宽,通常采用正交相移键控(Quadrature Phase Shift Keying, QPSK)调制,可提供500 Kbps~10 Mbps的传输速率。通过上行和下行通道,Cable Modem可以实现用户和局端的双向数字通信。

Cable Modem集Modem、调制器、加密/解密设备、桥接器、网络接口卡、虚拟专用网代理和以太网集线器的功能于一身。它无须拨号上网,不占用电话线,服务商的设备同用户的Modem之间建立了一个虚拟专用网连接,可提供随时在线的永久连接。

在Cable Modem系统中,采用了双向非对称技术,在下行方向有6 MHz的模拟带宽供系统中的用户共享。Cable Modem不同于线路交换的电话网定向呼叫连接,用户在连接时并不占用一个固定带宽,而是与其他活动用户共享,仅在发送、接收数据的瞬间使用网络资源。在毫秒级甚至微秒级的时间内,抓住一切利用带宽的机会下载数据包。如果在网络使用的高峰期中有拥塞,可以通过灵活地分配附加带宽来解决。只需简单分配一个6 MHz频段,就能倍增下行速率。另一种方法是在用户端重新划分物理网络,按照访问频度给用户合理分配带宽,速率可与专线媲美。

但是,利用Cable Modem和HFC进行组网时,在稳定性、可靠性、供电以及运行维护体制上还存在一些问题。

① 尽管理论速率达到了51 Mbps,由于Cable线路带宽是共享的,在用户达到一定规模后实际上无法提供宽带数据业务,用户分享到的带宽是非常有限的。

② 由于所有用户的信号都是在同一根同轴电缆上进行传送的,因此有被搭线窃听的危险。

③ CATV的树状网络特点也容易造成单点故障,如电缆的损坏、放大器故障、传送器故障都会造成整个结点上的用户服务的中断。

④ 虽然前期用户一定可以享受到非常优质的服务,因为在用户数量很少的情况下线路的带宽以及频段都是非常充裕的。但是,每一个用户的加入都会增加噪声、占用频道、减少可靠性以及影响线路上已有的用户服务质量。

⑤ 另外,从组网成本上考虑,需要对HFC改造完成后才能够应用。目前我国大部分HFC只能满足450 MHz的频段要求,而利用HFC提供双向业务至少需要750 MHz的带宽。这显然需要更换所有不符合要求的同轴电缆。同时,要实现双向的HFC需要更换目前有有线电视网上使用的单向放大器。

2. 电缆调制解调器的连接

Cable Modem的连接非常简便,一般可分为内置式(见图10-28)、网卡接口和USB接口的外置式(见图10-29)。

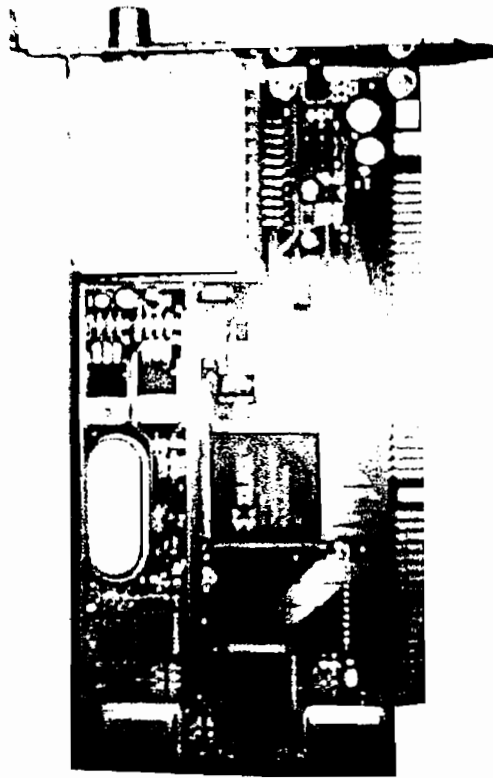


图 10-28 内置式电缆调制解调器

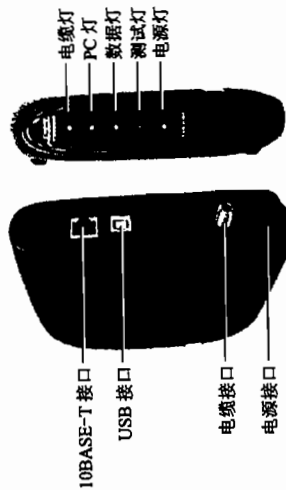


图 10-29 网卡接口和USB接口的外置式电缆调制解调器

图 10-30(a) 所示是 Motorola 的 SBC900, 这是一款具有无线接入和网关功能的 Cable Modem, 最多可支持 252 个用户。图 10-30b 所示是具有 Cable Phone 功能的 Cable Modem, 可提供两路 Cable Phone。Cable Phone 功能需运营商提供支持。

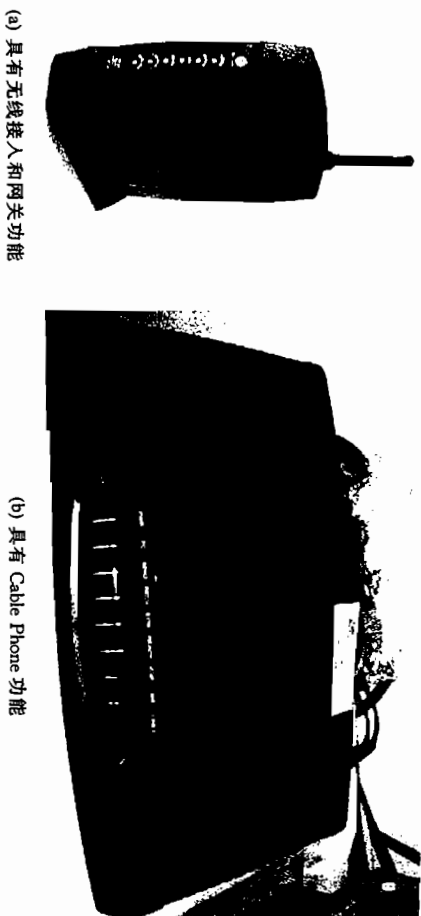


图 10-30 具有无线接入和网关功能或 Cable Phone 功能的电缆调制解调器

10.5.4 电力线调制解调器

1. PLC 和电力调制解调器的基本概念

电力线通信是指利用电力线通信 (Power Line Communication, PLC, 或 Power Line Telecommunications, PLT) 技术在电力线上传输数据和语音信号的一种通信方式。利用电力线上网的用户只需要添加一个特制的 PLC Modem, 这个 Modem 的数据线顶端是一个与常规电源插头规格相同的插头。只需将计算机与 Modem 连接, 再将 Modem 插上电源插座, 计算机即可上网, 而且这种方法传送信息的速度比 ISDN 要快 30 多倍。

PLC 的通信规范由 HomePlug Power Alliance (HomePlug 电力联盟) 制定。HomePlug AV 是继 HomePlug 1.0 之后的标准, HomePlug AV 标准能满足高清晰的 TV 或家庭中其他数字娱乐的需求。HomePlug 1.0 的规范覆盖 4 MHz ~ 21 MHz 的通信频段, 在这个频段内划分了 84 个 OFDM 通信信道。

PLC Modem 常采用正交频分多路复用 (Orthogonal Frequency Division Multiplexing, OFDM) 技术, 在严重电磁干扰的环境下可保证稳定完整地传输数据。OFDM 的原理是几个通信信道按 90 度的相位进行频分, 使得当某一个信道波形过零点时相邻信道的波形恰好是幅值最大值, 这样就保证了信道间的波形不会因外来的干扰而交叠、串扰。在 224 Mbps PLC Modem 中, 将 2 MHz ~ 30 MHz 分为 7 个独立的子频段, 每个子频段有 128 个载频, 每个载频采用独立的功率输出等级和 OFDM 技术。

由于 PLC Modem 是基于民用家庭或办公室的电力线, 因此无法为每一个端站划分独立的物

理信道, 从而电力线通信只能基于共享方式, 这样冲突就不可避免。而电力线的拓扑环境比较复杂, 随意的冲突检测难以进行, 因此 PLC Modem 通常采用具有避免冲突的载波侦听多路访问 (CSMA/CA) 协议。该协议的目的不是为了检测冲突, 而是通过划分时间段的方法来避免冲突。PLC Modem 的传输速率一般为 14 Mbps、45 Mbps 和 85 Mbps, 如今已出现了 224 Mbps 的产品。

电磁兼容性 (ElectroMagnetic Compatibility, EMC) 是指电器、电子产品能在规定的电磁环境中正常工作, 并不对该环境中其他产品产生过量的电磁干扰 (ElectroMagnetic Interference, EMI)。这里包含着两个方面的要求: 其一是要求产品对外界的电磁干扰具有一定的承受能力; 其二是要求产品在正常运行过程中, 该产品对周围环境产生的电磁干扰不能超过一定的限度。

2. 电力线调制解调器的应用

PLC Modem 的应用方法如图 10-31 所示, 它利用已敷设的电力线实现高速通信, 不必另外敷设电缆术。

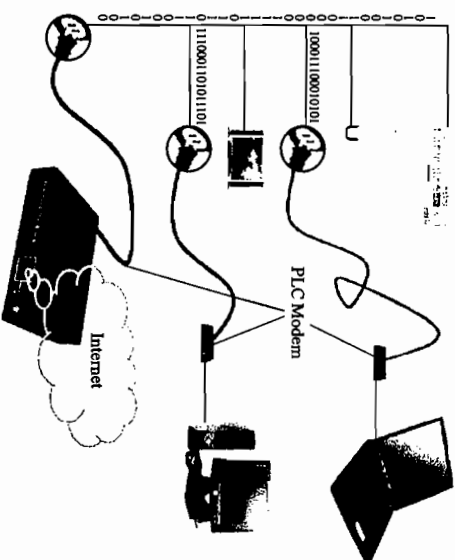


图 10-31 PLC Modem 的应用方法

PLC 系统的基本构成如图 10-32 所示。图中上方的通信设备称为前端, 可配备有高速以太网路接口和 PLC 接口, 是电力线通信网的对外接口, 可与 Internet 连接。用户把客户前提设备 (Customer Premises Equipment, CPE) 即 PLC Modem 接在电源插座上, 再与计算机连接即可享受视频流和 IP 电话等宽带网络服务。如果 CPE 设备和前端相隔较远, 则可使用中继器 (repeater) 等信号放大器。在这个区域无需布线, 就形成一个电力线的局域网, 实现电力线上网。

图 10-33 所示的分别是前端、中继器和 CPE 设备。

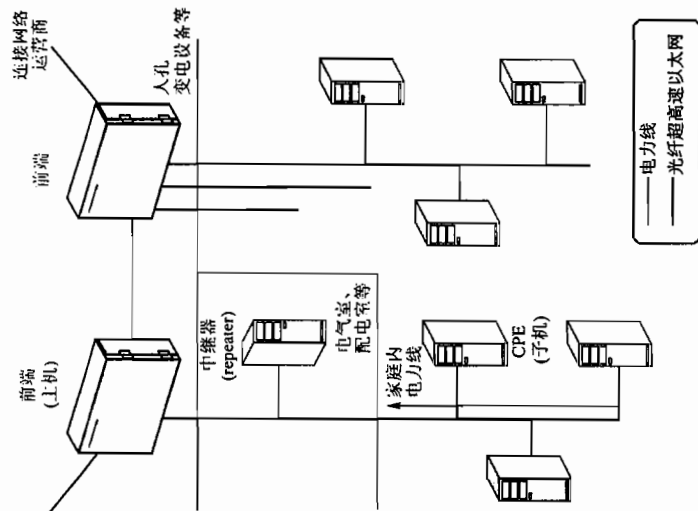


图 10-32 PLC 系统的基本构成

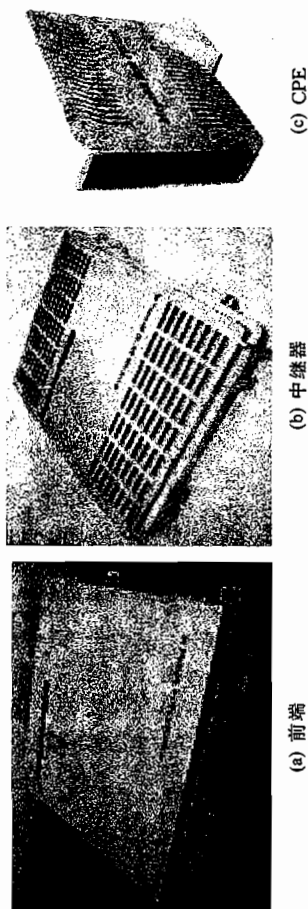
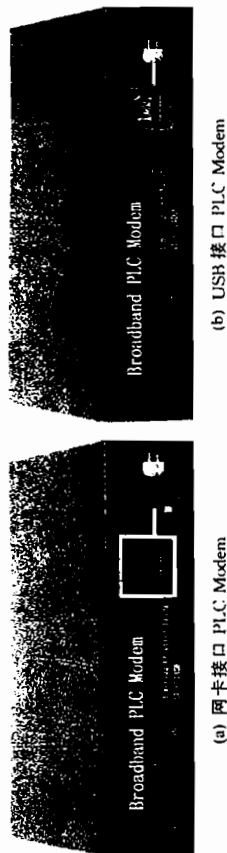


图 10-33 前端、中继器和 CPE 设备

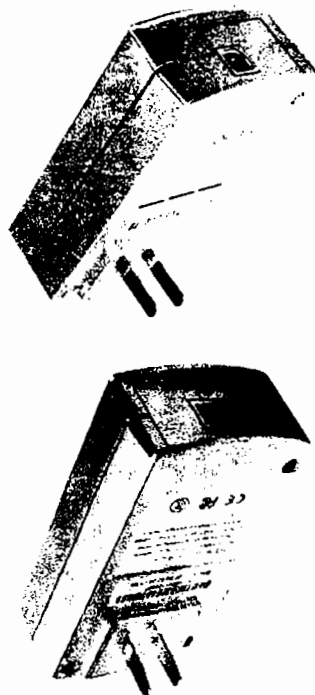


(a) 网卡接口 PLC Modem

(b) USB 接口 PLC Modem

图 10-34 ATL60140E 和 ATL60140U 台式外置 PLC Modem

图 10-35 所示的是支持 IEEE 802.3 协议的墙上外置 PLC Modem, 其中一台采用网卡接口, 另一台采用 USB 接口。



(a) 网卡接口 PLC Modem

(b) USB 接口 PLC Modem

图 10-35 ATL60142E 和 ATL60142U 墙上外置 PLC Modem

图 10-36 所示的是支持 IEEE 802.3 协议的嵌入式 PLC Modem, 其采用 USB 接口。图 10-37 所示是嵌入式电力线监控设备, 也采用 USB 接口。

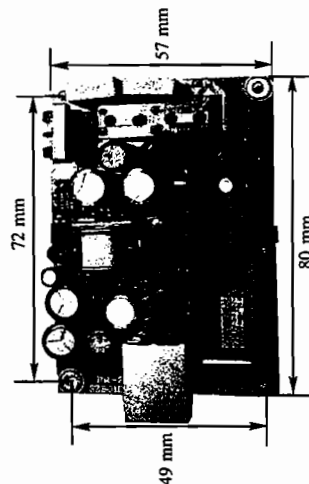


图 10-36 ATL60143PLC Modem

图 10-37 ATL60142M 嵌入式电力线监控设备

图 10-34 所示的是支持 IEEE 802.3 协议的台式外置 PLC Modem, 其中一台采用网卡接口, 另一台采用 USB 接口。

图 10-38 所示是支持 IEEE 802.3 协议的交换机。

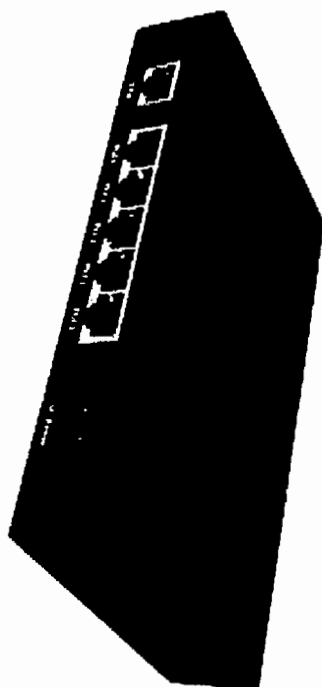


图 10-38 ATL60149 交换机

10.5.5 ISDN 和 ISDN 调制解调器

ISDN (Integrated Services Digital Network), 即综合业务数字网, 是由综合数据网 (Integrated Data Network, IDN) 发展而来, 是数字交换和数字传输相结合的产物, 是一个全数字的网络。在 POTS/PSTN 中, 用户到电话局之间采用的是模拟传输, 如图 10-39 所示, 显然由于沿途噪声的积累容易引起失真。而对于 ISDN 来说, 实现了用户线路的数字化, 如图 10-40 所示, 提供端到端的数字连接, 传输质量自然大大提高。

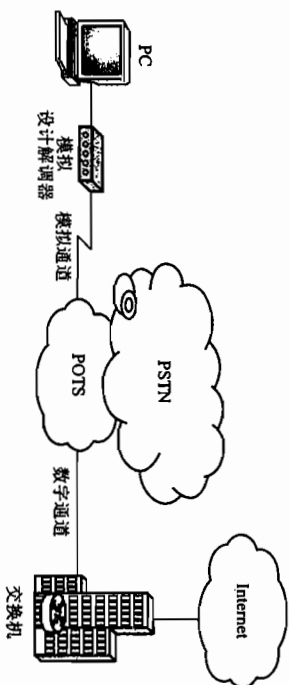


图 10-39 模拟调制解调器的接入示意图

ISDN 的基本速率接口 (BRI) 提供 2 个 B 信道和 1 个 D 信道 (2B + D)。BRI 的 B 信道传输速率为 64 Kbps, 用于传输用户数据; D 信道的传输速率为 16 Kbps, 用于传输控制信号。

在北美和日本, ISDN 的主速率接口 (PRI) 提供 23 个 B 信道和 1 个 D 信道, 总速率可达 1.544 Mbps, 其中 D 信道速率为 64 Kbps。而在欧洲、澳大利亚等国家, ISDN 的 PRI 提供 30 个 B 信道和 1 个 64 Kbps 的 D 信道, 总速率可达 2.048 Mbps。我国电话局所提供 ISDN 的 PRI 为 30B + D。目前广泛使用的是窄带 ISDN, 即 N-ISDN, 根本原因在于 N-ISDN 是以数字电话网为基础。

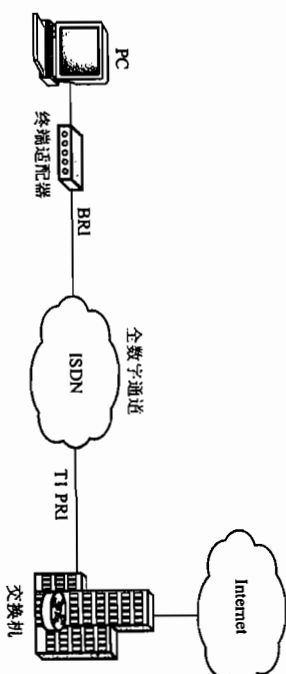


图 10-40 ISDN Modem 的接入示意图

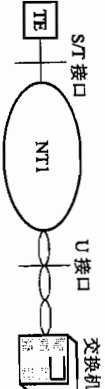
数字电话网的主要业务是 64 Kbps 速率的电路交换, 这在根本上限制了 ISDN 技术的发展及其支持新技术和业务的能力。早在 20 世纪 90 年代初期, ITU 就为宽带 ISDN (即 B-ISDN) 提出了相应的标准和建议, 确定了目前的新技术——异步传输模式 (ATM)。其传输速率可达 155 Mbps, 若传输介质采用光纤, 传输速率可达到数 Gbps。

ISDN 的组成部件包括用户终端 (TE)、终端适配器 (TA)、网络终端 (NT) 和 ISDN 路由器等设备。用户终端主要分为两种类型: 类型 1 和类型 2。其中类型 1 终端设备 (TE1) 是 ISDN 的标准终端设备, 通过 4 芯的双绞线数字链路 (数字电话机和 4 类传真机等); 类型 2 终端设备 (TE2) 是非 ISDN 标准终端设备, 必须通过终端适配器才能与 ISDN 连接。如果 TE2 是独立的设备, 则它与终端适配器的连接必须经过标准的物理接口, 如 RS-232C、V.24 和 V.35 等。

ISDN Modem 并不是专门为计算机所设计的, 它不仅可以作为 Modem 使用, 而且可用来自 ISDN 的线路转换成两路普通的模拟线路, 是 ISDN 线路的一个终端设备, 命名为 ISDN 终端适配器, 简称 TA。TA 可以自动选择一个空闲的 B 信道来进行通信。比如, 当仅使用一个 B 信道上网的时候, 如果有外线电话打进来, 那么电话机就会振铃, 可以正常接听电话; 如果需要往外打电话, 只要提起电话机, TA 会自动选择空闲的 B 信道来进行通话。当然如果同时使用了两个 B 信道, 外面呼叫你的号码时将会听见忙音。TA 上有一些指示灯, 用来指示当前 B 信道的使用状况。

网络终端 (NT) 是用户与网络连接的第一道接口设备, 包括 NT1 和 NT2 两种。NT1 (一类网络终端) 是 ISDN 应用中的重要设备, 如图 10-41 所示, NT1 实现了 U 接口和 S/T 接口间的物理层转换功能, 它有两个接口, 其中 U 接口与交换机线路终端相连, S/T 接口则可接数字话机、终端适配器 (TA)、网络适配器等终端设备 (TE)。我国使用欧洲标准, 图 10-41 ISDN 网络终端 NT1 提供的接口。一般来说, NT1 设备由本地电话公司 (也称局端) 在安装 ISDN 线路时提供。

图 10-42 和图 10-43 所示的是上海贝尔的网络终端 NT1 和智能网络终端 NT1+。NT1+ 提供两个 S/T 接口和两个模拟接口, 用户可以不通过 TA 就使用模拟电话等模拟设备, 并可支持多



种 ISDN 的补充业务,如呼叫等待、呼叫保持、三方通话等。在 NT1+ 的 S/T 接口接入 ISDN 适配卡,便可以 128 Kbps 的速率访问 Internet。

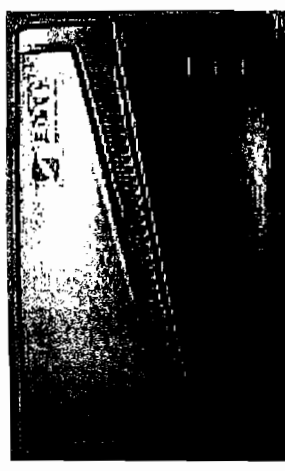


图 10-42 网络终端 NT1



图 10-43 智能网络终端 NT1 +

大多数终端适配器使用 T 接口。ISDN Modem 可以使用 U 接口(即有内部 NT1)或使用 S/T 接口。越来越多的 ISDN Modem 加进了 NT1,可以直接连接到 U 接口。其缺点是只能使用 ISDN Modem,且不能在一条 ISDN 线上连接更多的终端适配器。

图 10-44 所示是 PCI 总线接口的内置式 ISDN Modem,提供有 4 个 S/T 接口。

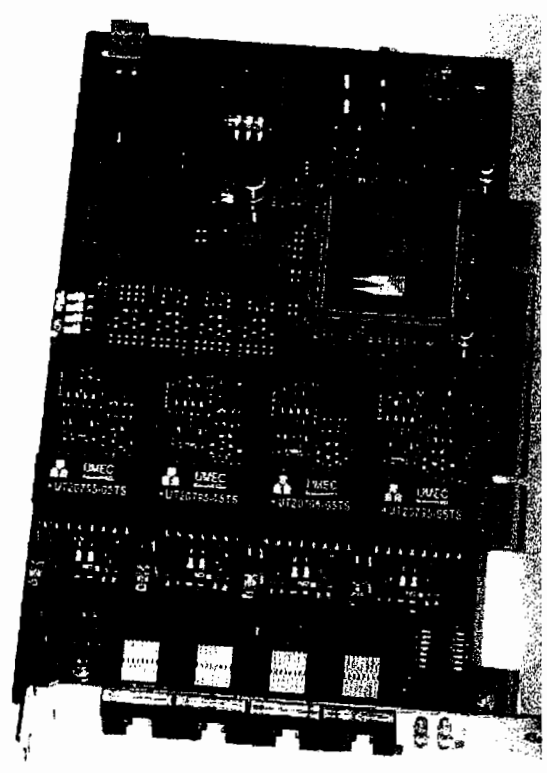


图 10-44 PCI 总线接口的内置式 ISDN Modem(4 个 S/T 接口)

图 10-45 所示是 PCI 总线接口的内置式 ISDN Modem,提供一个 E1(PRI)接口。
图 10-46 所示是 USB 接口的外置式 ISDN Modem,同时还提供了 14.4 Kbps 的传真功能。
图 10-47 所示是 Cisco SOHO 803 ISDN 接入路由器,提供一个 ISDN BRI S/T 接口、4 路

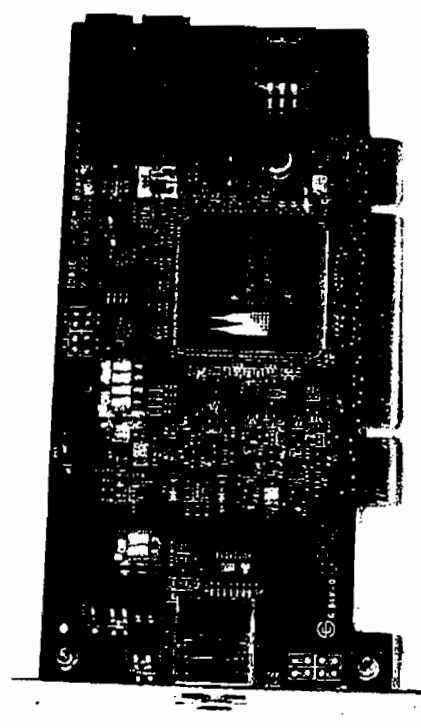


图 10-45 PCI 总线接口的内置式 ISDN Modem(一个 E1 接口)



图 10-46 USB 接口的 ISDN Modem



图 10-47 ISDN 接入路由器
10BaseT(RJ-45)、两个模拟电话端口和一个控制端口(RJ-45)。

10.5.6 CDMA 和 GSM/GPRS 调制解调器

CDMA (Code Division Multiple Access, 码分多路访问) 和 GSM (Global System of Mobile communication, 全球移动通信系统)/GPRS (General Packet Radio Service, 通用分组无线服务, 是一种基于 GSM 系统的无线分组交换技术, 提供端到端的、广域的无线 IP 连接; 它利用 GSM 网络中未使用的 TDMA 信道, 提供中速的数据传输; 目前, GPRS 移动通信网的传输速率可达 115 Kbps) 网络均提供收发短信、进行数据传输和上网功能, CDMA 和 GSM/GPRS Modem 肯定支持前两项功能, 只要使用简单的 AT 命令即可, 十分方便。有些 CDMA 和 GSM/GPRS Modem 还提供语音传输功能。图 10-48 所示的是外置式 GSM Modem。目前, 大多数 CDMA 和 GSM/GPRS Modem 还支持 TCP/IP、UDP 以及 PPP 协议, 因此, 也将此类 Modem 称为无线网卡。

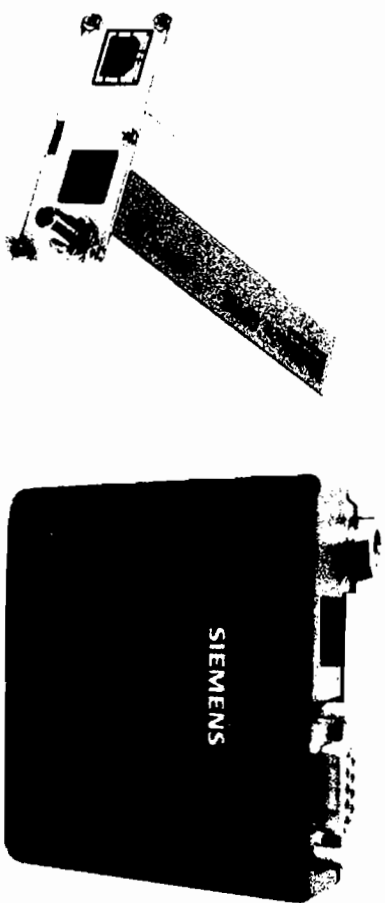


图 10-48 外置式 GSM Modem

10.6 网络设备

要将网络互相连接起来, 必须使用一些中间设备(或中间系统), 按 ISO 的术语称为中继 (relay) 系统。根据中继系统所在的层次, 可以有以下 5 种中继系统:

- ① 物理层(第一层, 层 L1)中继系统, 即中继器 (repeater), 又称转发器。
 - ② 数据链路层(第二层, 层 L2)中继系统, 即网桥或桥接器 (bridge); 以太网第二层交换机就是基于多端口网桥发展而来。
 - ③ 网络层(第三层, 层 L3)中继系统, 即路由器 (router)。
 - ④ 网桥和路由器的混合体——桥路器 (brouter), 兼有网桥和路由器的功能。
 - ⑤ 在网络层以上的中继系统, 即网关 (gateway)。
- 组建计算机网络时, 正确选择网络设备是重要的任务之一。通常涉及的网络设备有网络适

配器、中继器、集线器、交换机、路由器、网桥和网关等。其他的还包括服务器和网络打印机等。

10.6.1 网络适配器

网络适配器也叫网络接口卡 (Network Interface Card, NIC), 简称网卡。网卡是局域网中最基本的部件之一, 它是连接计算机与网络的硬件设备。无论是双绞线、同轴电缆、光纤还是无线连接, 都必须借助于网卡才能实现数据通信。

网卡的主要工作原理是整理计算机发往网线上的数据, 并将数据分解为适当大小的数据包之后向网络上发送出去。对于网卡而言, 每块网卡都有一个唯一的网络结点地址, 它是网卡生产厂家在生产时烧入 ROM 中的, 我们把它叫做 MAC 地址 (物理地址), 且保证绝对不会重复。

目前大都使用以太网网卡。台式机通常使用的有线网卡, 按其传输速率可分为 10 Mbps 网卡、10 Mbps/100 Mbps 自适应网卡以及 1 000 Mbps 网卡。无线网卡主要用于笔记本电脑等移动设备。根据网络类型, 还有其他的网卡, 如令牌环网卡、ARCnet 网卡等。

影响网卡性价比的因素有下述四种。

(1) 媒体访问方法

媒体访问方法对网卡性能有重要的影响。IEEE 802.3 CSMA/CD 和 IEEE 802.5 令牌环在访问媒体的确定性上有极大差别: 前者没有确定性; 后者虽有确定性, 但在用户不多、负载较轻的使用环境下不如 CSMA/CD 简便和快速。

(2) 最高比特率

最高比特率是给定媒体可能获得的最大比特率。实际的有效比特是要少得多, 其原因是有一定的协议额外开销。

(3) 网卡上处理器

有的网卡备有自身的处理器, 有的则没有。在无处理器的情况下, 所有处理都必须由主机完成。目前多数网卡都备有自身的处理器。

(4) 网卡与主机的数据传输方式

网卡与主机的接口可使用下述 3 种方法或它们的任何组合: 共享存储器法、DMA 法以及 I/O 端口法。网卡与主机之间的传输速率还取决于总线接口的数据宽度。

对用户而言, 选择网卡通常应注意网卡的下述特性。

(1) 媒体连接方式

媒体连接方式通常都提供 BNC、AUI 或 RJ-45 中的两种或全部。BNC 用于连接细缆, AUI 用于连接粗缆, RJ-45 用于连接双绞线。目前通常采用双绞线连接, 因此, 只需带 RJ-45 口就够了。

(2) I/O 地址和中断

当前的网卡都支持即插即用 (PnP), 可在安装网卡时避免 I/O 地址和中断与其他硬件冲突。

(3) 状态指示种类

状态指示用来指示网卡的工作状态, 便于了解其工作状态和诊断故障。一般用指示灯来指示工作状态, 通常配置有电源指示、发送指示 (Tx)、接收指示 (Rx)、链路状态指示 (Link)、超长指示 (Jabber) 和碰撞指示 (Collision) 等指示灯。

(4) 与 SNMP 的兼容性

与简单网络管理协议(SNMP)兼容,表明具备网络管理功能,通过网络管理站可对此网卡进行管理。在网络范围较大(网点多分布范围广)的网络,通常应将对SNMP的支持作为一种选择的重要依据。

(5) 自举 ROM 能力
支持自举 ROM 意味着能使无盘工作站连接到网络上,通常应有两种网络操作系统的自举 ROM。这种能力常作为一种选项。

图 10-49 所示是 PCI 总线接口的内置式网卡,网络接口为 RJ-45。

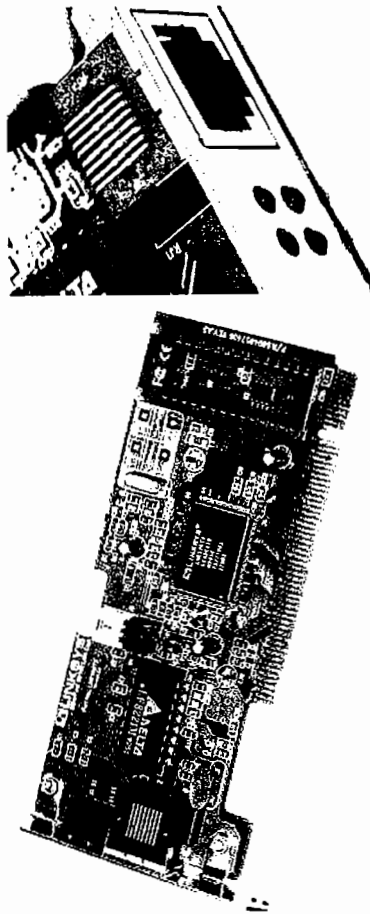


图 10-49 PCI 总线接口的内置式网卡

图 10-50 所示是 USB 接口的外置式网卡,网络接口为 RJ45。

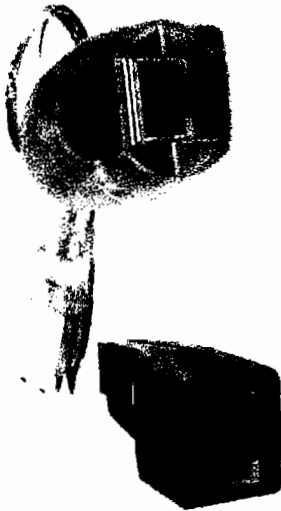


图 10-50 USB 接口的外置式网卡

图 10-51 所示为阿尔法 AFL-T16 千兆网卡,在最好的状态下传输速率可达到近 800 Mbps。

双绞线按其是否外加金属网丝套的屏蔽层而区分为屏蔽双绞线(Shielded Twisted Pair, STP)和非屏蔽双绞线(Unshielded Twisted Pair, UTP)。大多数局域网使用 UTP 作为传输介质。UTP 由 8 根不同颜色的线分成 4 对绞合在一起,成对扭绞的作用是尽可能减少电磁辐射与外部电磁

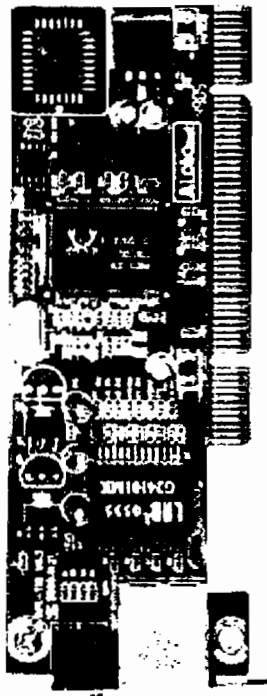


图 10-51 AFL-T16 千兆网卡

干扰的影响。目前最常用的标准是 5 类线和超 5 类线,传输频率为 100 MHz 和 125 MHz/200 MHz,主要用于 100Base-T 和 10Base-T 网络。

EIA/TIA-568 标准描述了网络布线的性能和安装规范,既包括了电缆本身,也包括了电缆如何与模块插头和插座的连接。

RJ-45 插头(俗称水晶头)由金属片和塑料构成,接头前端有 8 个凹槽,简称“SE”。凹槽内的金属触点共有 8 个,简称“8C”(C 代表 Contact,触点)。特别需要注意的是 RJ-45 插头引脚序号,当金属片面对我们的时候从左至右引脚序号是 1~8。EIA/TIA-568 标准对于电缆与模块插头和插座的连接有两个方案: T568A 和 T568B,如图 10-52 所示,这些标准定义了电缆中的色线与 RJ-45 模块插头和插座时对应关系。这两种都可采用,但不能互换。考虑到向后兼容性,一般选用 T568B。这两种颜色代码之间的唯一区别就是橙色和绿色线对的互换。相同标准的为直通线,不同标准的为交叉线。一般情况下,同层设备使用交叉线;不同层设备使用直通线。有些集线器具有自动检测功能,可以任意使用直通线或交叉线。

光纤接口类型很多。用户连接器(Subscriber Connector, SC)光纤接口主要用于局网交换环境,在一些高性能千兆交换机和路由设备上提供了这种接口。它与 RJ-45 接口看上去很相似,不过 SC 接口显得更扁些,其明显区别还是里面的触点。如果是 8 条细的铜触点,则是 RJ-45 接口,如果是一根铜柱则是 SC 光纤接口。其他的光纤接口有 ST、LC 等多种类型,并分为单工和双工。

图 10-53 所示的是 3Com 公司的 3C996-SX 网卡,其主要特性如下:

- 介质: 1000Base-SX。
- 连接器: SC。
- 总线: 32/64 位 PCI(PCI 2.2)/PCI-X(PCI-X 1.0)。
- 工作距离: 1000Base-SX。850 nm 多模全双工光纤最长 500 m; 50/125 nm 多模光缆最长 500 m; 62.5/125 nm 多模光缆最长 220 m。
- 兼容性: 兼容 IEEE 802.3, 802.3u, 802.3z, 802.3x, 802.1Q, 802.3ad。

10.6.2 无线网卡

无线网卡目前支持的协议主要有常用的 IEEE 802.11x、CDMA 和 GPRS。使用支持 IEEE

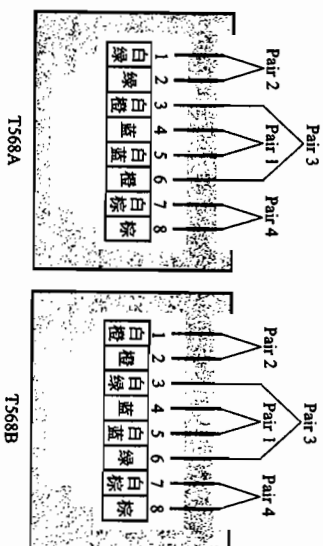
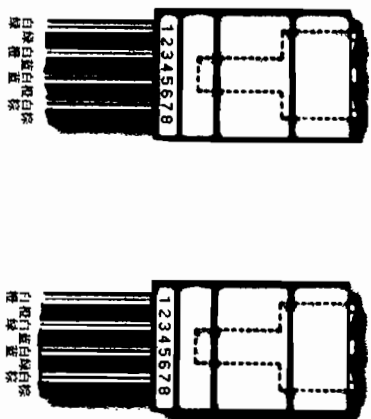


图 10-52 TS68A 和 TS68B

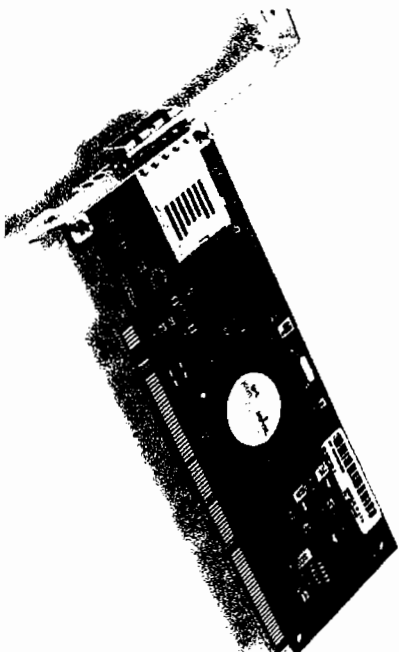


图 10-53 3COM 的 3C996-SX 光纤网卡

802.11x 协议的无线网卡上网时,需要通过具有无线接入点功能的设备接入,而使用支持 CDMA 或 GPRS 的无线网卡(也称为 CDMA Modem, GSM/GPRS Modem)上网时,需要有 CDMA 网络或 GPRS 网络的支持。无线网卡的接口主要有 PCI 总线接口、PCMCIA 接口、CF 接口和 USB 接口等。图 10-54 所示是 PCI 总线接口的无线网卡,图 10-55 所示是 PCMCIA 接口的无线网卡。

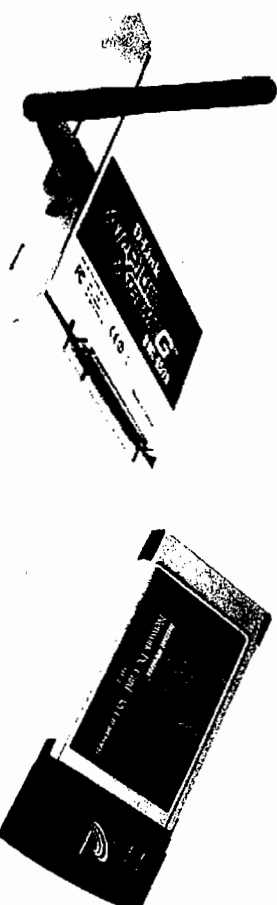


图 10-54 PCI 总线接口的无线网卡

图 10-55 PCMCIA 接口的无线网卡

图 10-56 所示是 PCMCIA 接口的 CDMA 无线网卡,图 10-57 所示是 PCMCIA 接口的 GPRS 无线网卡。



图 10-56 PCMCIA 接口的 CDMA 无线网卡

图 10-57 PCMCIA 接口的 GPRS 无线网卡

爱国者的无线通,创造性地集 GPRS、蓝牙、闪存三大功能于一体,可为用户提供较 GPRS 网卡更为便捷、自如的多种数据传输及存储服务。其主要特性如下:

- ① 115.2 Kbps GPRS 无线上网,速度更快,不受地点限制,让你无线上网更自由。
- ② 蓝牙无线传输技术,性能更稳定,功能更完善,无线组网更方便,数据传输更自由。
- ③ 超稳定 A 级闪存芯片,随时存储数据信息,保证网络断线不丢数据,数据存储及交流更自由。
- ④ 标准 USB 接口设计,适用所有带 USB 接口的笔记本及台式机计算机,产品应用更自由。

10.7 中继器

传输介质超过了网段长度后,可采用中继器(Repeater,也称集中器——Concentrator)延伸网络的距离,对弱信号予以再生、放大。IEEE 802 标准规定最多允许 4 个中继器接 5 个网段。中继器工作在物理层,不提供网段隔离功能。当中继系统是转发器时,一般不称之为网段互连,因为这仅仅是把一个网段扩大了,而这仍然是一个网段。在局域网中,目前很少使用单一功能的中继器,一般使用集线器或交换机。

10.8 网桥

网桥(Bridge)也称桥接器,是连接两个局域网的存储转发设备,用它可以连接具有相同或相似体系结构的网络系统。网桥的功能包括:对所接收的信息帧只做少量的封装,而不做任何修改;可以采用另外一种协议来转发信息;有足够的缓存空间,以满足高峰期的要求;必须具有寻址和路径选择的能力。

图 10-58 所示的是 SMC 专为商务和家庭用户所设计的一款支持最新 WPA 安全标准 IEEE 802.11g 的产品。这是一款多功能产品,集成了无线网桥、无线 AP、无线中继器和无线网络适配器等功能。

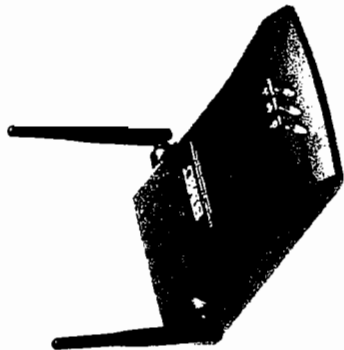


图 10-58 SMC2870W 无线网桥

10.9 集线器

集线器(Hub)的主要功能是对接收到的信号进行再生、整形和放大,以扩大网络的传输距离,同时把所有结点集中在以它为中心的结点上。它工作于 OSI 参考模型第一层,即物理层。集线器是中继器的一种,其区别仅在于集线器能够提供更多的端口服务,所以集线器又叫多口中继器。集线器主要以优化网络布线结构、简化网络管理为目标而设计。集线器是对网络进行集中管理的最小单元,像树的主干一样,它是各分支的汇集点。

集线器是一种以星状拓扑结构将通信线路集中在一起的设备,相当于总线,是局域网中应用最广的连接设备,按配置形式可分为独立型集线器、模块化集线器和堆叠式集线器三种。智能型集线器改进了一般集线器的缺点,增加了桥接能力,可滤掉不属于自己网段的帧,增大网段的频宽,且具有网管能力和自动检测端口所连接的 PC 网卡速度的能力。市场上常见有 10 Mbps、100 Mbps 等速率的集线器。

以集线器为中心结点的优点是:当网络系统中某条线路或某结点出现故障时,不会影响其他

结点的正常工作,这就是集线器刚推出时与传统的总线网络的最大的区别和优点,因为它提供了多通道通信,大大提高了网络通信速度。集线器的主要不足体现在如下几个方面:用户带宽共享,带宽受限;广播方式,易造成网络风暴;非双工传输,网络通信效率低。

如今由于交换机的价格越来越低,基本接近集线器的价格,所以集线器将逐渐退出历史舞台,为交换机所取代。不过由于历史原因,大家还是称组建小型局域网的设备为集线器,但这里的集线器实际上是交换机,只是一个习惯叫法而已。

图 10-59 所示为几款典型的集线器产品。

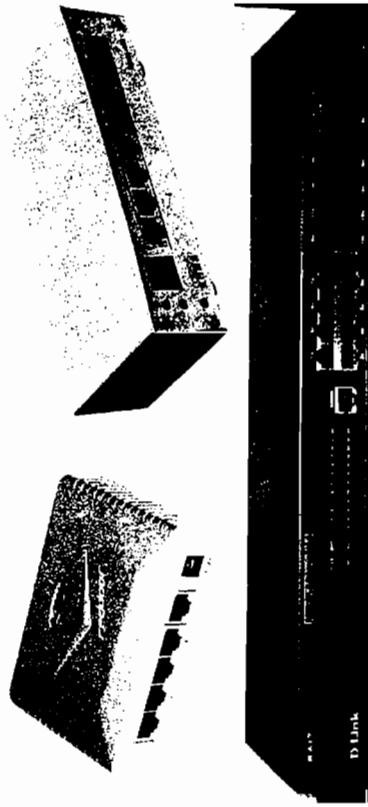


图 10-59 几款集线器产品

10.10 交换机

交换机(Switch)是一种基于 MAC 地址(网卡的硬件地址)识别,能完成数据包封装、转发功能的网络设备。交换机以前是高端用户使用的设备,交换机内部有多个总线,成本要比集线器高。不过由于近几年网络技术的飞速发展,交换机的价格已经降低到了接近集线器的水平。性能更好而价格相差不多这一不争的事实,决定了交换机将彻底替代集线器。

交换机的分类不是很严格,由于升级迭代的原因,某制造商的某一特定型号可能属于多个类别,而此制造商的两个不同型号可能属于同一类别。一般可分为低端固定交换机、低端可变换交换机、中端交换机和高端交换机。高端交换机还可以集成路由模块(称为第三层交换)以及其他功能(如负载均衡和防火墙)。

从理论上讲,交换机是根据以太网数据包的目的地址将以太网数据包从原端口送至目的端口,可同时对不同目的端口的数据包发送以太网数据包,可以同时传输多个数据包,这个特点决定了通过交换机组建的网络传输速率更快。交换机可以同时建立多个传输路径,所以应用在连接多台服务器的网段上可以收到明显的效果。

交换机主要用于连接集线器、服务器或分散式主干网。交换机所采用的技术可分为以下几种。

(1) 直通交换(cut through)

一旦收到数据包中的目标地址,在收到全帧之前便开始转发。适用于同速率端口和碰撞误码率低的环境。直通方式的以太网交换机可以理解为在各端口间是纵截交叉的线路矩阵电话交换机。它在输入端口检测到一个数据包时,检查该数据包的包头,获取数据包的目的地址,启动内部的动态查找表转换成相应的输出端口,在输入与输出交叉处接通,把数据包直通到相应的端口,实现交换功能。由于不需要存储,延迟非常小,交换非常快,这是它的优点。它的缺点是,因为数据包内容并没有被以太网交换机保存下来,所以无法检查所传送的数据包是否有误,不能提供错误检测能力。由于没有缓存,不能将具有不同速率的输入/输出端口直接接通,而且容易丢失数据包。

(2) 存储转发(store and forward)

确认收到帧后先处理坏帧再转发。适用于不同速率的端口和碰撞、误码率高的环境。存储转发方式是计算机网络领域应用最为广泛的方式。它把输入端口的数据包先存储起来,然后执行循环冗余校验(CRC),在对错误数据包处理后才取出数据包的目的地址,通过查找表转换成输出端口地址并转发数据包。正因如此,存储转发方式在数据处理时延时较大,这是它的不足,但是它可以对进入交换机的数据包进行错误检测,有效地改善网络性能。尤其重要的是它可以支持不同速率的端口间的转换,保持高速端口与低速端口间的协同工作。

(3) 碎片隔离(Fragment Free)

这是介于上述两种方案之间的一种解决方案。它检查数据包的长度是否够64个字节,如果小于64字节,说明是假包,则丢弃该包;如果大于64字节,则发送该包。这种方式也不提供数据校验。它的数据处理速率比存储转发方式快,但比直通式慢。

图10-60所示是具有光纤接口的交换机,提供1个SC接口,TFM5116Ei-M型为多模光纤,TFM5116Ei-S型为单模光纤。

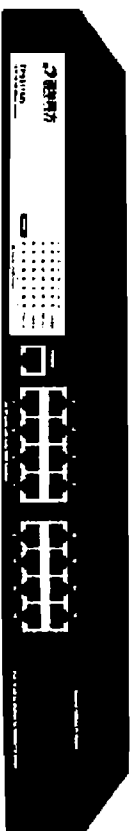
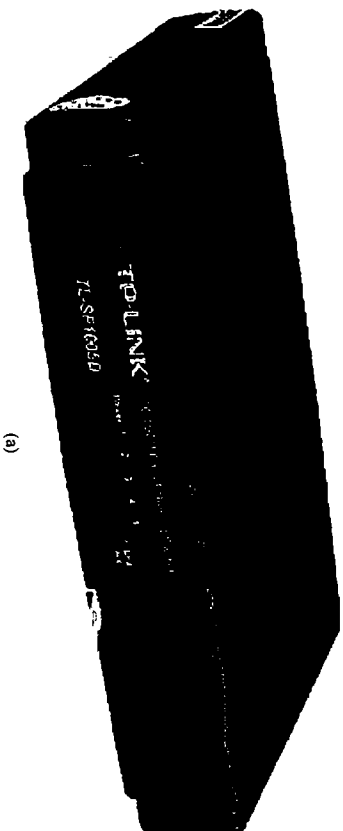
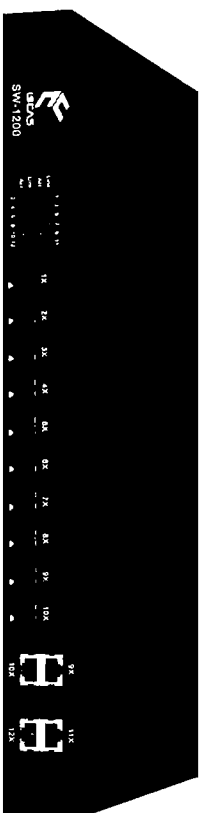


图 10-60 TFM5116Ei, 具有光纤接口的交换机

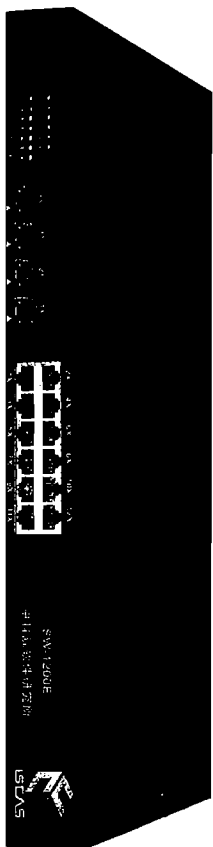
图 10-61 所示是其他一些品牌的交换机。



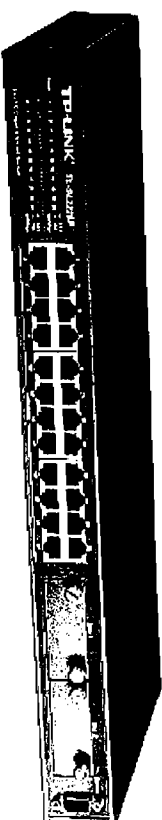
(a)



(b)



(c)



(d)

图 10-61 几款交换机产品

10.11 路由器

路由器是网络层上的设备,用于网络与网络之间的连接。路径的选择是路由器的主要任务。路径选择包括两种基本的活动:一是最佳路径的判定;二是网间数据包的传送。用路由器连接的网络可以使用在数据链路层和物理层完全不同的协议。路由器是在多个网络和介质之间实现网络互联的一种设备,是一种比网桥更复杂的网络互连设备。

路由器一般面向企业,一个低端的华为 2621 路由器就需要近 9 000 元。普通消费者是不需要路由器的。路由器主要提供网络数据包的路由功能,当然随着功能的加强还会在网络地址转换、访问控制列表、路由协议等方面有更好的表现。路由器的每个端口所对应的价格非常高,这也是为什么目前所有路由器设备的端口数都不会很多的原因。路由器各个端口处在不同的冲突域。

使用路由器后,形形色色的通信子网融为一体,形成了一个更大范围的网络。从宏观的角度

上看,可以认为通信子网实际上是由路由器组成的网络,路由器之间的通信则通过各种通信子网的通信能力予以实现。

10.12 宽带路由器

这里需要明确一个概念,家用路由器并不是真正的路由器,只是一种专门用于共享网络接入的设备,我们常常称之为宽带路由器。所以在价格上与真正的路由器没法比,目前市场宽带路由器的价格从100元到几百元不等。

宽带路由器只拥有简单的路由功能,再强大一点的可能会有网络地址转换功能,但总体来讲其功能与真正的路由器功能相差甚远,倒是在很多功能上和交换机有相似之处,例如提供数据包的高速转发,每个端口一个冲突域,多个端口共享一个广播域等。所以,如果给宽带路由器下一个明确定义的话,应该是这样的:宽带路由器=普通交换机+路由功能。

集线器不能为网络中的用户提供网络共享功能,所以在家庭组建网络时不能使用集线器代替宽带路由器来共享网络。但是,可以使用代理服务器+集线器的形式来共享上网。

10.13 集线器、交换机与路由器的区别

集线器(hub)的作用可以简单地理解为将一些机器连接起来组成一个局域网。而交换机(又名交换式集线器)作用与集线器大体相同。但是两者在性能上有区别:集线器采用共享带宽的工作方式,而交换机是独享带宽。这样在机器很多或数据量很大时,两者将会有比较明显的差异。而路由器与以上两者有明显区别,它的作用在于连接不同的网段并且找到网络中数据传输的最佳路径,一般情况下个人用户不用着路由器。路由器是产生于交换机之后,就像交换机产生于集线器之后,所以路由器与交换机也有一定联系,并不是完全独立的两种设备。路由器主要克服了交换机不能转发数据包的不不足。

总的来说,路由器与交换机的主要区别体现在以下几个方面。

(1) 工作层次不同

最初的交换机工作在OSI/RM模型的数据链路层(第二层),所以它的工作原理比较简单。而路由器一开始就设计工作在OSI/RM模型的网络层(第三层),可以得到更多的协议信息,路由器可以做出更加智能的转发决策。

(2) 数据转发所依据的对象不同

交换机是利用物理地址或MAC地址来确定转发数据的目的地址。而路由器则是利用不同IP地址来确定数据转发的地址。IP地址是在软件中实现的,描述的是设备所在的网络,有时这些第三层的地址也称为协议地址或者网络地址。MAC地址通常是硬件自带的,是由网卡生产商分配的,而且已经固化到了网卡中,一般来说是不可更改的。而IP地址则通常由网络管理员或系统自动分配。

(3) 传统的交换机只能分隔冲突域,不能分隔广播域,而路由器可以分隔广播域。

由交换机连接的网段仍属于同一个广播域,广播数据包会在交换机连接的所有网段上传播,在某些情况下会导致通信拥塞和安全漏洞。连接到路由器上的网段会被分隔成不同的广播域,广播数据不会穿过路由器。虽然第三层以上交换机具有VLAN功能,也可以分隔广播域,但是各子广播域之间是不能通信的,它们之间的通信仍然需要路由器。

(4) 路由器提供了防火墙的服务

路由器仅仅转发特定地址的数据包,不传送不支持路由协议的数据包和未知目标的网络数据包,从而可以防止广播风暴。

交换机一般用于局域网与广域网之间的连接,交换机归于网桥,是数据链路层的设备,有些交换机也可实现第三层的交换。路由器用于广域网之间的连接,可以在异构网络之间转发分组,工作于网络层,但只是从一条线路上接受输入分组,然后向另一条线路转发。这两条线路可能属于不同的网络,并采用不同协议。相比较而言,路由器的功能较交换机要强大,但速度相对也慢,价格昂贵。第三层交换机既有交换机线速转发报文的功能,又有路由器良好的控制功能,因此得以广泛应用。

“交换”是当今网络领域出现频率最高的一个词。从桥接到路由到ATM直至电话系统,无论何种场合都可将其套用,搞不清到底什么才是真正的交换。其实交换一词最早出现于电话系统,特指实现两个不同电话机之间语音信号的交换,完成该工作的设备就是电话交换机。所以从本意上来讲,交换只是一种技术概念,即完成信号由设备入口到出口的转发。因此,只要是和符合该定义的设备都可被称为交换设备。由此可见,“交换”是一个涵义广泛的词语,当它被用来描述数据网络第二层的设备时,实际上是指桥接设备;而当它被用来描述数据网络第三层的设备时,指的是路由设备。近几年,交换机的性能有了许多改进,其中最突出的改进是虚拟网络和第三层交换。

对于不同规模的网络,路由器的作用的侧重点有所不同。在主干网上,路由器的主要作用是路由选择。主干网上的路由器,必须知道到达所有下层网络的路径。这需要维护庞大的路由表,并对连接状态的变化做出尽可能迅速的反应。路由器的故障将会导致严重的信息传输问题。在城域网中,路由器的主要作用是网络连接和路由选择,即连接下层各个基层网络单位——园区网,同时负责下层网络之间的数据转发。在园区网内部,路由器的主要作用是分隔子网。早期的互联网基层单位是局域网(LAN),其中所有主机处于同一逻辑网络中。随着网络规模的不断扩大,局域网演变成以高速主干和路由器连接的多个子网所组成的园区网。在其中,各个子网在逻辑上独立,而路由器就是唯一能够分隔它们的设备,它负责子网间的报文转发和广播数据隔离,在边界上的路由则负责与上层网络的连接。

传统交换机和路由器的性能和功能不可兼得。交换机交换速度快,但控制功能弱;路由器控制功能强,但报文转发速度慢。解决这个矛盾的技术是第三层交换。第三层交换机既有高速转发报文的能力,又有良好的控制功能。但第三层交换机无法取代路由器却是不争的事实,这是因为它有各自的用武之地。第三层交换机非常适应局域网环境,但无法适应网络拓扑各异、传输协议不同的广域网环境。而路由器则广泛用于广域网中。

10.14 网关

网关(或协议转换器)是工作在 OSI/RM 模型网络层之上的具有协议转换功能的设施。所以称为设施,是因为网关不一定是一台设备,有可能在一台主机中实现网关功能。

网关用来互连完全不同的网络,把一种协议转换成另一种协议,把一种数据格式转换成另一种数据格式,把一种速率转换成另一种速率,以求两者的统一。它根据用户通信用的计算机的 IP 地址,界定是否将用户发出的信息送出本地网络;同时,它还将接收外界发送给本地网络计算机的信息。

网关用于以下几种场合:

- ① 异构型局域网的互连,如互连专用交换网与遵循 IEEE 802 标准的局域网。
- ② 局域网与广域网的互连。
- ③ 广域网与广域网的互连。
- ④ 局域网与主机的互连(当主机的操作系统与网络操作系统不兼容时,可以通过网关连接)。

网关主要有以下几类:

- ① 协议网关:协议网关通常用于使用不同协议的网络区域间的协议转换。
- ② 应用网关:应用网关用于不同数据格式间转换数据格式。
- ③ 安全网关:安全网关是各种技术的融合,具有重要且独特的保护作用,其范围从协议级过滤到十分复杂的应用级过滤。

图 10-62 所示为 SonicWALL PRO 5060C 网关,主要参数见表 10-3。目前的参考价格为 520 000 元。

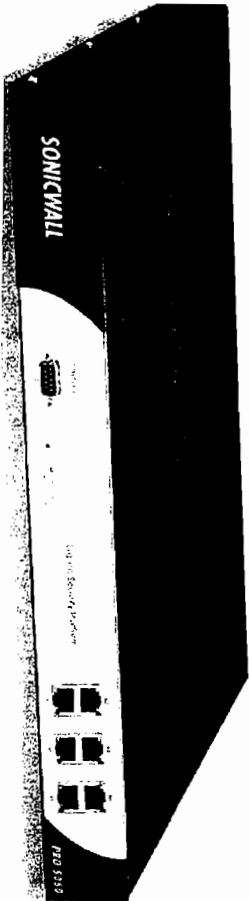


图 10-62 SonicWALL PRO 5060C

表 10-3 SonicWALL PRO 5060C 主要参数	
处理器	带 Cavium Nitrox 加密处理器的 Intel 处理器
端口	6 个 10 Mbps/100 Mbps/1000 Mbps 自动检测以太网铜缆端口 1
防火墙性能	全状态吞吐量:1 + Gbps,连接数:750 000,策略数:15 000,攻击防护:22 种 DoS,DDoS 及扫描攻击
用户数限制	无限制
最大并发连接数	750 000
VPN 处理能力	IPSec VPN,与其他遵循 IPSec 的 VPN 网关兼容,捆绑 2 000 个用于远程用户的 VPN 客户对话(最多 6 000 个);支持多达 4 000 个 VPN 的点点对点 VPN 策略;3DfS 及 AES 加密性能:500 Mbps
支持标准	TCP/IP、UDP、ICMP、HTTP、HTTPS、RADIUS、IPSec、ISAKMP/IKE、SNMP、L2TP、DHCP、PPPoE 及 PPTP
产品电源	100 V ~ 240 V AC
产品功率	120 W
网关重量	7.05 kg
网关尺寸	1U 可机架安装(43.18 cm x 41.23 cm x 4.45 cm)
其他性能	支持网络地址转换(NAT)协议、动态主机配置协议(DHCP)、以太网点对点协议(PPPoE)、点对点隧道协议(PPTP)
其他特点	1 年保修,负责维修和更换由于厂商原因而造成的任何故障产品,90 天支持与固件更新,访问 SonicWALL 电子支持工具

图 10-63 所示为 SonicWALL PRO 100 网关,主要参数见表 10-4。目前的参考价格为 29 132 元。

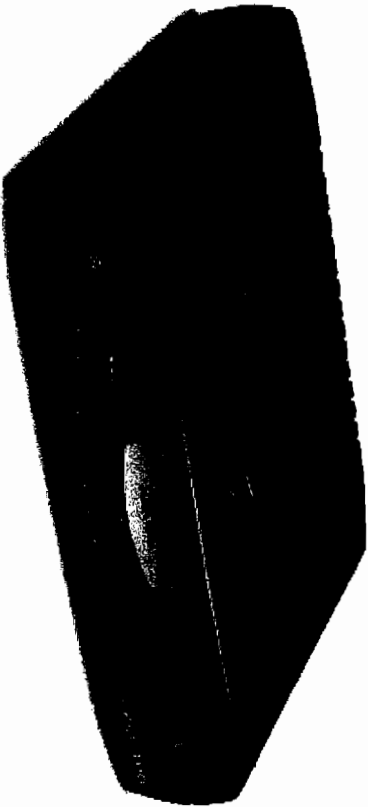


图 10-63 SonicWALL PRO 100

表 10-4 SonicWALL PRO 100 主要参数

并发连接数	6 000
网络吞吐量	75 Mbps
安全过滤带宽	20 Mbps
用户数限制	无用户数限制
入侵检测	DoS,DDoS
安全标准	UL,CE,GS Mark,VCCI,FCC Part 15,TUV,C-Tick,FCC,UL,BSMI,VCCI,CSA,ICSA Firewall,ICSA IPsec VPN
控制端口	RS-232
管理	SNMP

10.15 网络打印机

10.15.1 网络打印机的基本概念

网络打印机已不再是计算机的外设,而是作为一个独立的网络结点实现高速打印输出。网络打印机是通过网络实现打印的,也就是说,把这台打印机接入网络并安装相应的软件就能实现打印,而不用连接在任何一台计算机上。网络打印在目前主要有两种方式:一种是通过外置式的打印服务器来实现网络打印;另一种是通过内置的打印服务器(俗称“网卡”)来实现网络打印。

实际上,通过外置式打印服务器来实现网络打印的打印机不能算是真正的网络打印机,因为这种打印机本身不是通过网络接口与网络相连的,实现普通接口与网线接口转换功能的是外置式打印服务器,而非打印机本身。严格来讲,网络打印机是带内置式打印服务器的打印机。

不能说任何一台打印机只要接上打印服务器就是网络打印机了。网络打印机的主板上具有专门针对网络环境的设计,网络打印机必须具备接收功能和队列功能,一般还应具备网络管理和自动诊断功能及自动报错功能。

网络打印机应具有高速、高效的特点。因此,打印部分应具有较高的打印速度。虽然曾经有极少数喷墨网络打印机面世,但占据主流的网络打印机几乎全部是激光打印机。

10.15.2 网络打印机的应用

网络打印机的硬件组成部分为打印部分和网络部分。打印部分的功能和指标与普通打印机的类似,主要的是打印质量、打印速度和打印幅面及纸张存储量。网络部分的主要功能有以下几种。

(1) 支持的网络类型

网络打印机的网卡一般是自适应的 10 Mbps/100 Mbps 网卡,可以直接连接到企业内部局域网

网上,包括以太网(Ethernet)和令牌环网(Token Ring),并且支持 AppleTalk、IPX/SPX、TCP/IP 等网络协议。兼容多种网络系统平台,包括 Windows、Macintosh、UNIX 等操作系统。常见的网络连接为 RJ-45 双绞线接口和 BNC 同轴电缆接口。

(2) 管理软件

网络打印机与其他普通打印机的一个主要区别就在于网络打印机不仅需要打印机的驱动程序,而且还需要一个安装在服务器上的网络打印机管理软件来管理网络打印机。随着网络技术的发展,网络打印机的管理软件在管理方式上也得到了质的飞跃,如专业的打印机制造商 HP 公司就把网络打印机的管理软件从本地计算机搬到了 Web 上。网络打印机的管理软件要符合实际应用的需求。

(3) 硬件结构

对于网络打印机来说,打印服务器的内置方式是相当重要的。有些产品将打印服务器是与打印机主板集成在一起的,但如果打印服务器出现故障,可能需更换整个主板,这种结构也不便于硬件升级。

(4) 其他的辅助功能

各个厂商都有自己的独创的技术,如 Epson 的“作业平衡”,HP 的“ColorSmart II”等。

HP LaserJet 1022n 为内置式网络黑白激光打印机,分辨率为 1 200 × 1 200 dpi。打印接口为 USB 接口,还有一个用于实现网络打印的以太网接口。此款打印机支持 A4 幅面的打印,能打印的介质类型有普通纸、激光打印纸、相纸、信封、投影胶片、标签、卡片、明信片。打印速度为每分钟 18 页,月打印负荷 8 000 页。使用的耗材为鼓粉一体式设计,硒鼓的使用寿命为 2 000 页(5%覆盖率)。还有一个 250 页输入纸盒和一个 10 页优先输入插槽。

HP LaserJet 1022n 提供 ProRes 1200、FastRes 1200 以及 FastRes 600 三种打印质量。其中 ProRes 1200 打印质量能够提供分辨率为 1 200 × 1 200 dpi 的高质量打印输出。HP 1000 系列能够提供从个人到网络、到无线网络的全面的解决方案,HP LaserJet 1022nw 具备无线网络功能的网络打印机。

Brother HL-2070N 为内置式网络黑白激光打印机,最高分辨率为 2 400 × 600 dpi。打印接口为具有 USB 2.0 接口和 IEEE 1284 并行接口,还有一个用于实现网络打印的 10/100BASE-TX 以太网接口。此款打印机支持 A4 幅面的打印,能打印的介质类型有普通纸、铜版纸、再生纸、信封、标签和透明胶片等。打印速度为每分钟 20 页,月打印负荷 8 000 页。使用鼓粉分离式耗材:硒鼓最多可打印 12 000 页;墨盒最多可打印 2 500 页(以 5%的打印覆盖率打印 A4 或信纸)。还有一个 250 页输入纸盒和一个 10 页优先输入插槽。

Brother HL-2070N 支持内置自动双面打印单元和局域网内部共享打印功能。

Epson AcuLaser C3000N 为 A4 幅面的彩色网络激光打印机,可以实现每分钟输出 24 页黑白或彩色文档。C3000N 采用了 Epson 特有的“迅彩”图形加速芯片(ASIC 控制芯片)通过硬件技术实现图形图像快速而且准确的彩色输出,使得它的首页输出时间仅为 13.5 秒,彩色和黑白打印速度同为每分钟 24 页。月打印量最大可达 60 000 页,完全能满足一般企业的打印需要。由于采用了 AcuBrite 聚酯碳粉技术,能使打印效果更加细腻、均匀,打印输出分辨率高达 2 400 dpi。

10.16 VoIP

10.16.1 VoIP 的基本概念

VoIP (Voice over IP) 就是俗话说的 IP 电话 (Internet Protocol Phone), 是利用 IP 网络实现语音通信的一种通信手段。IP 电话主要有 3 种应用模式: “计算机-计算机”、“计算机-电话”和“电话-电话”。

1. 传统语音技术与 VoIP 技术

PSTN 通过标准的信令控制网络 (SS7), 成为覆盖全球的大型电话网络。这个网络和 Internet 有着惊人的相似, 它们都采用级联结构, 具有易于扩展的特性。

通过 PSTN 传送语音业务有良好的服务质量 (QoS) 保证, 且延迟极低 (小于 2 ms)。但是, 基于电路交换的结构使得 PSTN 电路数量庞大, 结构复杂, 成本高, 难以增加新的业务 (SS7 的应用已使得 PSTN 向智能网发展, 但硬件的改造仍很复杂), 计费复杂, 交换机间互通性问题严重。

由于 PSTN 网络中线路资源固定分配方案的低效率和 IP 网络的蓬勃发展, 人们早就考虑利用分组网络进行电话业务, 这就是 IP 电话。最初的应用是计算机用户间通过 IP 网络进行电话或传真通信; 随后是计算机用户与 PSTN/PBX (Private Branch Exchange) 用户之间通过 IP 网络进行电话或传真通信; 最后是 PSTN/PBX 用户通过 IP 网络进行电话或传真通信。

2. VoIP 技术标准

IP 技术发展到现在, 目前主要有 H.323、SIP、MGCP 和 RSVP 等几种标准。目前 VoIP 的体系结构只有两种, 一种是基于 H.323 的, 而另一种是基于 SIP 的。

(1) H.323

H.323 遵循了传统运营商的管理体系, 与 PSTN 比较一致, 即采用了公用号码、静态组网的方式, 在很多方面符合运营商和最终用户的习惯。H.323 是最早被采纳的 VoIP 标准。

H.323 协议体系规定了基于分组网进行两点/多点实时媒体通信的系统逻辑组件、消息定义和通信过程。H.323 的逻辑组件包括网守 (Gatekeeper, 又称网关控制器——Gateway Controller)、多点控制单元 (MCU)、网关 (Gateway, 位于网络边界)、代理 (Proxy)。其中的网关用作 PSTN 和 IP 网间的桥梁, 网守可以集中控制各个网关。从 PSTN 来的呼叫被 H.323 网关受理, 把其中的话流转换成 IP 数据包经过 IP 网络传送到目的网关下的指定电话终端; 同时受理 IP 网络来的呼叫 (可以来自计算机终端或另一个 PSTN 网络) 转向 PSTN 目的终端。H.323 的核心协议如图 10-64 所示。

基于网守的概念使得 H.323 网络是集中控制的, 便于统一维护和管理。但因此呼叫处理时延大, 同时网络规模也因此受到极大的限制。

(2) 会话发起协议 (SIP)

会话发起协议 (Session Initiation Protocol, SIP) 在组网方面明显体现了互联网的思想, 它不是采用类似 PSTN 的电话号码, 其号码可以转换为 Internet 域名, 因此是一种动态组网方式, 具有很

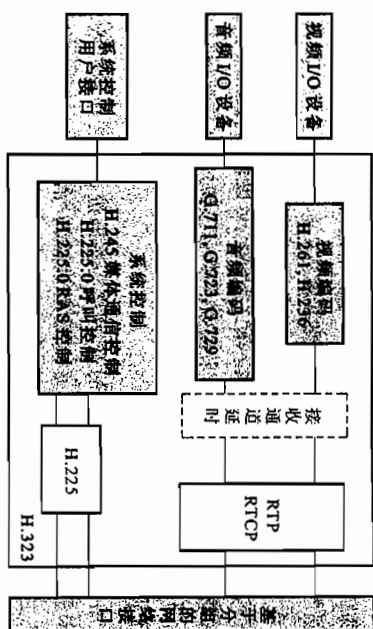


图 10-64 H.323 的核心协议

好的移动性。用户在打电话时, 类似于寻找某个网址。如果运营商建设的是一个大规模、电信级网络, 那么当前 SIP 在体制、运营、管理等诸多方面还不是很成熟。它的优点可能在某些接入的应用环境中体现出来, 例如灵活性、终端智能化、多业务等。如果一个全国范围内的大网采用 SIP, 那么其成熟度和与现行资费、运营体制的矛盾便会充分暴露出来。例如对呼叫属于国内长途还是国际长途就不好判断。

与 H.323 相比, SIP 需要相对智能的终端, 即终端需要包含用户代理客户机构和用户代理服务两部分, 由这两部分实现呼叫请求、呼叫应答和一些用户的特定需要。正是因为 SIP 系统有了相对智能的终端系统, 所以它才有可能实现用户个性化的需要。SIP 的普遍使用有待于互联网用户的进一步发展和网上计算机的数量与电话机上电话机的数量具有可比性。

(3) 媒体网关控制协议 (MGCP)

该协议体系是由 Level3、Bellcore 和 Cisco 发起的业务提供商/厂商共同体提出并由 Internet 工程任务组 (Internet Engineering Task, IETF) 开发形成的。简单网关控制协议 (Simple Gateway Control Protocol, SGCP) 最初是为了开发一个比 H.323 简单且易于扩展的 VoIP 协议而开发, 后来 SGCP 和 IP 数据广播 (Internet Protocol Datacasting, IPDC) 协议的结合形成了媒体网关控制协议 (Media Gateway Control Protocol, MGCP)。MGCP 与 H.223 和 SIP 不同, H.323 和 SIP 提出两套 IP 电话体系结构, 二者完全独立, 不能互相兼容, 只能互通。MGCP 不涉及 IP 电话的体系结构, 只涉及网关分解问题, 因而不仅可用于 H.323 VoIP 系统, 也可用于 SIP VoIP 系统。

MGCP 的特点是将协议体系的功能彻底分解, 每一组功能由相应的逻辑组件完成, 而且严格按照客户-服务器 (Client/Server) 模式设计。网关可分解成媒体网关 (MG) 和媒体网关控制器 (MGC), MG 在 MGC (或软交换) 的控制下, 实现跨网媒体业务。从逻辑上看似乎很简单, 网关由 MG 和 MGC 组成。从物理上看就不那么简单了, 到目前为止, 网关的分解还没有确定的方式, 可根据不同的需求进行分解。

(4) 资源预留协议 (RSVP)

资源预留协议 (Resource Reservation Protocol, RSVP) 是一组通信规则, 它允许为视频及其他高带宽消息组播的传送预留信道或路径。RSVP 是 Internet 综合服务 (IIS) 模型的一部分, 它保证了尽最大努力的服务、实时服务以及受控的链路共享。

Internet 上基本的路由选择是“尽最大努力”, 对于大多数用户已足够, 但仍不能满足在 Internet 上传送音视频节目所需的连续的流传输。有了 RSVP, 可以预留 Internet 的带宽, 并能以此比从前更高的数据传输速率和更可靠的数据流来接收。RSVP 同时也支持单播 (一个源到一个目的地) 及多播 (一个目的地的传输)。IPv4 和 IPv6 都包括了 RSVP。

3. VoIP 的优点

由于 VoIP 采用了基于统计时分复用的 IP 网络为基础进行电话业务传送, 并采用了先进的数字信号处理技术进行语音编码、语音压缩、语音静默检测/消除等, 比固定传输速率、固定电路分配的 PSTN 网络更具优势, 主要表现在以下几点。

- (1) 符合三网合一的发展方向
在高效率传输数据业务的 IP 网络上传输语音、传真甚至视频业务已成必然之势, IP 网络已遍布全世界, 这无疑是三网合一的可靠途径。
- (2) 充分利用网络资源
先进数字信号处理技术的采用, 使得一路语音传送所需的带宽从原来的 64 Kbps 减小到

8 Kbps 甚至更低, 从而使得在同一条线路上可以增加 7~10 倍的呼叫; 同一呼叫的 IP 包可以通过不同的路由到达目的地, 一方面可以充分利用网络资源, 一方面提高了可靠性。

- (3) 价格低廉的服务

正因为数字信号处理技术和统计时分复用的结合使得通过 VoIP 进行电话业务传送的成本远远低于传统 PSTN, 这也是它的最大优势和最大的好处。单纯从带宽的节约上, VoIP 的成本就只有传统 PSTN 的 1/10; 且没有长途概念, 所有业务均按时间或流量计费, 所以在长途业务上更具竞争力。

10.16.2 VoIP 的实现

大多数 VoIP 终端都需要同时支持 1~2 路语音通道, 这确保能满足双方和三方会议通话的需求。每路语音通道由用于编码的 ITU G. 7xx 算法、语音质量 (G. 168 - 2000) 和电话信令 (SIP 或 H. 323) 组成。

VoIP 设备必须支持多种编解码类型, 具体使用哪种类型取决于网络的要求。常用的编解码标准有 G. 729AB、G. 723.1、G. 711 和 iLBC。语音质量组件包括声学回声消除 (G. 167 AEC)、语音活动检测 (VAD) 和舒适噪声产生 (CNG) 以及抖动缓冲。回声消除用于消除语音的回声, VAD 用来编码和压缩语句间的静音, 而 CNG 用来在电路或环境导致的噪声情况下提供语音检测。抖动缓冲可以确保不管经过网络的延时有多大, 语音包都能以正确的顺序得到处理, 同时使包的丢失最少。其他组件还有像用于语音邮件和寻呼的双音多频 (DTMF) 信号系统拨号这样的电话算法、基于 SIP 或 H. 323 的通话建立控制。实际上, 上述标准目前已经很普遍, 但任何 VoIP 应用都需要具备软件可升级性, 语音编解码和通话控制标准还在不断发展, 因此现有的设计能够支持可能出现的新的语音算法或标准是非常重要的。VoIP 的实现成本、功耗和功能之间的平衡对设计

提出了重大的挑战。RISC/DSP 可满足 VoIP 产品的功耗、成本和性能目标。

过去这些算法的处理通常都是在双处理器系统上进行的。系统中的 RISC 处理器用于控制和管理, DSP 处理器用于语音编码解码和高质量的算法处理。如今整合了 RISC 和 DSP 功能的嵌入式处理器足以作为全双工通道提供执行这些运算所需的带宽。

每种架构实现 DSP 功能的方式不同, 因此处理语音算法的效率也有高有低, 但是设计师可采用一种通用规则, 即一个全双工通道使用约 100 MHz 的带宽或 100 MCPS (每秒百万周期)。在 DSP 性能得到增强的 RISC 核心上, 这个带宽一般占总的可用带宽的一半以下, 因此能给出大多数 VoIP 设备预留出足够的应用余量。

通用的嵌入式 RISC 处理器在 20 世纪 90 年代后期开始集成数字信号处理 (DSP) 功能, 如 ARM9E 处理器, 它支持快速和灵活的 16 x 16 MAC 与饱和算法。这种创新设计非常适合语音和音频处理, 因为这种处理需要组合的数字信号处理和饱和算法。在某些情况下, 新的 DSP 指令专门针对处理 16 位语音数据所需的常见的循环 (loop)。后续的几代处理器 (如 ARM11) 集成了更复杂的信号处理逻辑和指令, 可进一步提升那些需要中等或更高复杂度的信号处理应用的性能。在几种嵌入式处理器中还集成了单指令多数据 (SIMD) 引擎, 配合新的指令可以充分发挥视频和多媒体处理算法的固有并行处理优势。这些架构还改善了语音处理性能, 与 DSP 增强的核

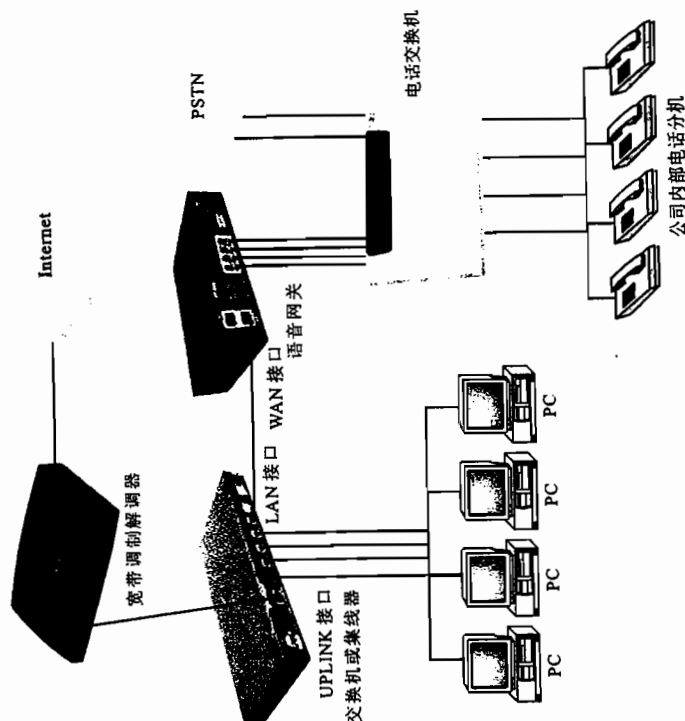


图 10-65 通过语音网关实现 VoIP

心相比,速率提升20%~30%。

10.16.3 VoIP 的应用方案

中小型企业可以通过 VoIP 终端,实现与外地办事处或机构、客户的通话,以节省大量的长途费用。中小型企业的特点是:一般已建立了局域网,多台计算机通过局域网内的某一计算机或小型路由实现共享上网,需要装备 VoIP 终端,以满足通话的需要。

1. 应用方案 1

图 10-65 所示是 VoIP 应用的典型方案,可保持原有的网络布局不变。将一台语音网关连接到网络交换机上,并将语音网关与公司的电话交换机相连。通过 ADSL 或 Cable Modem 虚拟拨号接入,使得公司内部所有分机都可以拨打 IP 电话。同时,也可以通过公司的电话交换机和连接公司电话外线,拨打普通电话。在此方案中,只需购买一部语音网关就可以实现公司内部所有分机都可拨打 IP 电话的功能。

2. 应用方案 2

将 IP 电话机直接连接在公司局域网内就可以满足公司特定人员拨打 IP 电话的需要,如图 10-66 所示,使用 IP 电话的人员可以拥有一个唯一的 IP 电话号码。

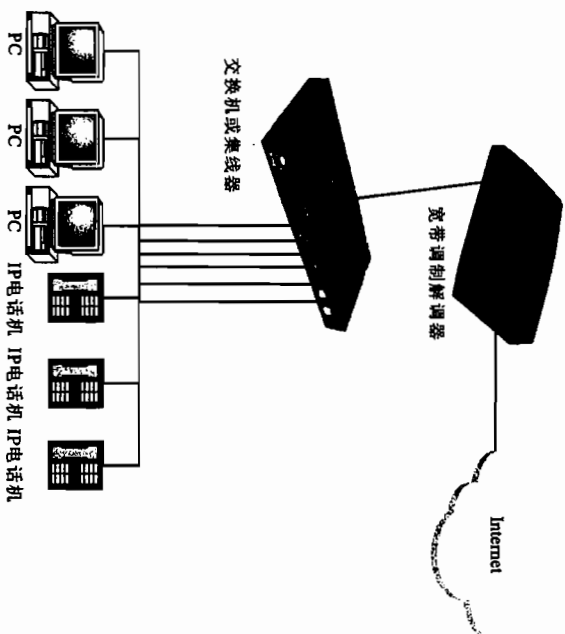


图 10-66 通过 IP 电话机实现 VoIP

习题

1. 简述通信系统的组成。
2. 简述模拟通信和数字通信的主要区别。
3. 简述计算机网络的组成和分类。
4. 简述 OSI/RM 的特点。
5. 简述几种 Modem 的作用和应用。
6. 目前家庭宽带上网的主要方法有哪些? 简述各自的特点。
7. 网络设备主要有哪些? 其工作特性有哪些区别?
8. 目前个人无线上网有哪些方法? 简述所需硬件设备的特点。
9. 集线器和交换机的主要区别是什么?
10. 交换机和路由器的主要区别是什么?
11. 网关和网桥的主要区别是什么?
12. 网络打印机与普通打印机的主要区别是什么? 网络打印机的主要特点有哪些?
13. 简述 VoIP 的基本工作原理。
14. 目前与 VoIP 技术有关的标准有哪些?
15. 简述 VoIP 语音网关应用方式。
16. 什么 IP 电话机? 如何通过 IP 电话机实现 VoIP 的功能?

算设备、电视机顶盒、手机、数字电视、多媒体、汽车、微波炉、数码相机、家庭自动化系统、电梯、空调、安全系统、自动售货机、工业自动化仪表与医疗仪器等。嵌入式计算机系统同通用型计算机系统相比具有以下特点。

① 嵌入式计算机系统通常采用面向特定应用的嵌入式 CPU, 与通用型 CPU 的最大不同是嵌入式 CPU 大多用于为特定用户群设计的系统, 通常具有低功耗、体积小、集成度高等特点, 能够把通用 CPU 中许多由板卡完成的任务集成在芯片内部, 从而有利于嵌入式系统设计趋于小型化, 移动能力大大增强, 跟网络的耦合也越来越紧密。

② 嵌入式系统是将先进的计算机技术、半导体技术和电子技术与各个行业的具体应用相结合的产物。这一点决定了它必然是一个技术密集、资金密集、高度分散、不断创新的知识集成系统。

③ 嵌入式系统的硬件和软件都必须高效的, 其设计需量体裁衣、去除冗余, 力争在同样的硅片面积上实现更高的性能, 这样才能更具有竞争力。

④ 嵌入式系统和具体应用有机地结合在一起, 它的升级换代往往和具体产品同步进行, 因此嵌入式系统产品一旦进入市场, 具有较长的生命周期。

⑤ 为了提高执行速度和系统可靠性, 嵌入式系统中的软件一般都固化在存储器芯片或单片机本身中, 而不是存储在磁盘等载体中。

⑥ 嵌入式系统本身不具备自举开发能力, 即设计完成以后用户通常不能对其中的程序进行修改, 必须有一套相应的开发工具和环境才能进行开发。

11.1.2 嵌入式系统的结构及分类

嵌入式系统一般指非 PC 系统, 它包括硬件和软件两部分。硬件包括处理器/微处理器、存储器、外设器件、I/O 接口及图形控制器等。软件部分包括操作系统 (OS) 和应用程序, 其中操作系统必须是实时的、多任务的操作系统。有时设计人员把这两种软件组合在一起。应用程序控制着系统的运作和行为; 而操作系统控制着应用程序编程与硬件的交互作用。

嵌入式系统将微处理器或控制器的系统电路与其有软件相结合, 以期达到最高的性价比。嵌入式系统的体系结构如图 11-1 所示, 其基本硬件组成如图 11-2 所示。

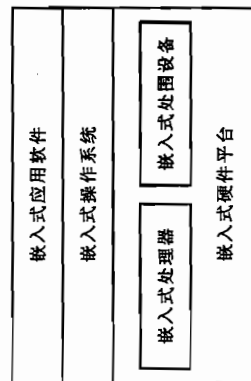


图 11-1 嵌入式系统的体系结构示意图

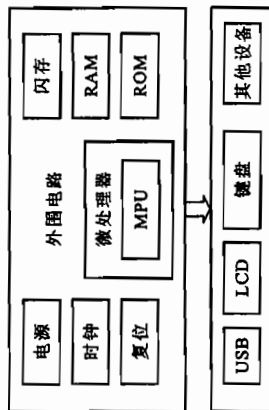


图 11-2 嵌入式系统基本硬件组成示意图

第 11 章 专用计算机与工业控制系统

计算机的出现将人类社会带入信息化时代, 同时也给传统工业注入新的活力, 表现为计算智能的引入, 出现了智能仪表、程控通信设备、现代工业控制系统等。反过来, 与应用领域的紧密结合, 也促进了一类专用计算机或处理器的系统。这些专用计算机或处理器包括嵌入式系统、数字信号处理器、片上可编程系统、单片微型机等。与一般通用计算机不同, 它们往往作为专用系统的核心部件出现, 在能耗、体积、实时性和功能专一等方面常常有特殊要求。本章主要介绍嵌入式系统、数字信号处理器、片上可编程系统、单片微型机的概念、功能、特点和结构。作为计算机应用的重要实例, 本章的最后一节对分布式和现场总线式工业控制系统的组成、工作原理和标准进行了介绍, 以期对专用计算机或处理器以及计算机在工业控制领域中的应用有一初步的了解。

11.1 嵌入式系统

11.1.1 嵌入式系统的概念及特点

嵌入式系统 (Embedded System) 是以应用为中心, 以计算机技术为基础, 并且软硬件可裁剪, 适用于应用系统对功能、可靠性、成本、体积、功耗的严格要求的专用计算机系统。它一般由嵌入式微处理器、外围硬件设备、嵌入式操作系统以及用户的应用程序等 4 部分组成, 用于实现对其他设备的控制、监视或管理等功能。

根据 (美国) 电气和电子工程师协会 (IEEE) 的定义, 嵌入式系统是“用于控制、监视或者辅助操作机器或设备的装置” (devices used to control, monitor, or assist the operation of equipment, machinery or plants)。可以看出, 此定义是从应用的角度来考虑的。嵌入式系统是软件和硬件的综合体, 还可以涵盖附属机电装置等。其更一般的定义为: “以应用为中心, 以计算机技术为基础, 软件与硬件可裁剪, 功能、可靠性、成本、体积与功耗有严格要求的专用计算机系统。”

嵌入式系统几乎包括了生活中的所有电器设备, 如个人数字助理 (PDA)、移动计

1. 嵌入式微处理器

嵌入式系统的核心是嵌入式微处理器。嵌入式微处理器一般具备以下4个特点。

- ① 对实时多任务有很强的支持能力,能完成多任务并且有较短的中断响应时间,从而使内部的代码和实时内核的执行时间减少到最低限度。
- ② 具有功能很强的存储区保护功能。由于嵌入式系统的软件模块化结构,为了避免在软件模块之间出现错误的交叉作用,需要设计强大的存储区保护功能,同时也有利于软件诊断。
- ③ 可扩展的处理器结构,以便迅速开发出满足应用需求的高性能嵌入式微处理器。
- ④ 嵌入式微处理器必须具有很低的功耗,尤其是用于移动计算的和靠电池供电的嵌入式系统更是如此,其功耗一般在毫瓦甚至微瓦级。

嵌入式处理器大体可分为以下几类。

(1) 嵌入式微处理器

嵌入式微处理器可谓是通用CPU的微缩版。相对于通用CPU,嵌入式微处理器具有体积小、功耗少、成本低的优点,速度上会慢一些。嵌入式微处理器通常可以运行嵌入式操作系统,应用于比较高档的领域。典型的产品如32位的ARM核心、64位的MIPS。

(2) 嵌入式微控制器

嵌入式微控制器的最大特点是单片化,常称为单片机。顾名思义,单片机就是将众多的外围设备(简称外设,如A/D、I/O等)集成到一块芯片中,从而大幅度降低了成本。单片机非常适合工业控制领域,典型的产品如著名的51系列。

(3) 专用微处理器

相对于上述比较通用的处理器,专用微处理器是针对某一特定领域专门设计的微处理器,如昂贵的视频游戏机微处理器等。DSP处理器本质上属于专用微处理器。DSP处理器的系统结构和指令进行了优化设计,使其适合于执行数字信号处理算法(如FFT、FIR等)。DSP处理器运行速度非常快,在数字信号处理的方方面面大显身手。由于越来越广泛的领域需要高速数字信号处理,DSP处理器也有越来越通用化的倾向,常常可以把DSP处理器单独列成一类。我们将在下一节进一步介绍DSP处理器。

2. 嵌入式外围设备

在嵌入式系统硬件系统中,除了中心控制部件(MCU、DSP、EMPU、SOC)以外,用于完成存储、通信、调试、显示等辅助功能的其他部件,都可以算作嵌入式外围设备。目前常用的嵌入式外围设备按功能可以分为存储设备、通信设备和显示设备三类。

存储设备主要用于各类数据的存储,常用的有静态易失型存储器(RAM、SRAM)、动态存储器(DRAM)和非易失型存储器(ROM、EPROM、EEPROM、FLASH)三种,其中FLASH凭借其可擦写次数多、存储速度快、存储容量大、价格便宜等优点,在嵌入式领域内得到了广泛应用。

目前存在的绝大多数通信设备都可以直接在嵌入式系统中应用,包括RS-232接口(串行通信接口)、SPI(串行外围设备接口)、IrDA(红外线接口)、I2C(现场总线)、USB(通用串行总线接口)、Ethernet(以太网接口)等。

由于嵌入式应用场合的特殊性,通常使用的是阴极射线管(CRT)、液晶显示器(LCD)和触摸屏(Touch Panel)等外围显示设备。

3. 嵌入式操作系统

为方便、快捷开发嵌入式系统,需要有专门负责管理存储器分配、中断处理、任务调度等功能软件模块,这就是嵌入式操作系统。嵌入式操作系统是用来支持嵌入式应用的系统软件,是嵌入式系统极为重要的组成部分,通常包括与硬件相关的底层驱动程序、系统内核、设备驱动程序与内核的接口、通信协议、图形用户界面(GUI)等。

嵌入式操作系统具有通用操作系统的基本特点,如能够有效管理复杂的系统资源,能够对硬件进行抽象,能够提供库函数、驱动程序、开发工具箱等。但与通用操作系统相比较,嵌入式操作系统在系统实时性、硬件依赖性、软件固化性以及应用专用性等方面,具有更加鲜明的特点。

嵌入式操作系统根据应用场合可以分为两大类:一类是面向消费电子产品的非实时系统,这类设备包括个人数字助理(PDA)、移动电话、机顶盒(STB)等;另一类则是面向控制、通信、医疗等领域的实时操作系统,如WindRiver公司的VxWorks、QNX系统软件公司的QNX等。实时系统(Real Time System)是一种能够在指定或者确定时间内完成系统功能,并且对外部和内部事件在同步或者异步时间内能做出及时响应的系统。在实时系统中,操作的正确性不仅依赖于逻辑设计的正确程度,而且与这些操作进行的时间有关,也就是说,实时系统对逻辑和时序的要求非常严格,如果逻辑和时序控制出现偏差将会产生严重后果。实时系统主要通过三个性能指标来衡量系统的实时性:它们是响应时间(Response Time)、生存时间(Survival Time)和吞吐量(Throughput)。响应时间是实时系统从识别出一个外部事件到做出响应的时间内,生存时间是数据的有效等待时间,数据只有在这段时间内才是有效的;吞吐量是在给定的时间内系统能够处理的事件总数,吞吐量通常比平均响应时间的倒数要小一点。实时系统根据响应时间可以分为弱实时系统、一般实时系统和强实时系统三种。弱实时系统的设计宗旨是使各个任务运行得越快越好,但没有严格限定某一任务必须在多长时间内完成。弱实时系统更多关注的是程序运行结果的正确与否以及系统安全性能等,对任务执行时间的要求相对来讲较为宽松,一般响应时间可以是数十秒或者更长。一般实时系统是弱实时系统和强实时系统的一种折中,它的响应时间可以在秒的数量级上,广泛应用于消费电子设备中。强实时系统则要求各个任务不仅要保证执行过程和结果的正确性,同时还要保证在限定的时间内完成任务,响应时间通常要求在毫秒甚至微秒的数量级上,这对涉及医疗、安全、军事的软硬件系统来说是至关重要的。时限(Deadline)是实时系统中的一个重要概念,指的是对任务截止时间的要求,根据时限对系统性能的影响程度,实时系统又可以分为软实时系统(soft real-time-system)和硬实时系统(hard real-time-system)。软实时指的是虽然对系统响应时间有所限定,但如果系统响应时间不能满足要求,并不会导致系统产生致命的错误或者崩溃;硬实时则指的是对系统响应时间有严格的限定,如果系统响应时间不能满足要求,就会引起系统产生致命的错误或者崩溃。如果一个任务在时限到达之时尚未完成,对软实时系统来说还是可以容忍的,最多只会降低系统性能,但对硬实时系统来说则是无法接受的,因为这样带来的后果根本无法预测,甚至可能是灾难性的。在目前实际运用的实时系统中,通常允许软硬件两种实时性同时存在,其中一些事件没有时限要求,另外一些事件的时限要求是软实时的,而对系统产生关键影响的那些事件的时限要求则是硬实时的。

4. 嵌入式应用软件

嵌入式应用软件是针对特定应用领域,基于某一固定的硬件平台,用来达到用户预期目标的

计算机软件。由于用户任务可能有时问和精度上的要求,因此有些嵌入式应用软件需要特定嵌入式操作系统的支持。嵌入式应用软件和普通应用软件有一定的区别,它不仅要求其准确性、安全性和稳定性等方面能够满足实际应用的需要,而且还要尽可能地优化,以减少对系统资源的消耗,降低硬件成本。

总而言之,嵌入式系统一般是专用系统,而通用计算机是通用计算平台;嵌入式系统的资源比通用计算机少得多;嵌入式系统软件故障带来的后果比通用计算机大得多;嵌入式系统一般采用实时操作系统;嵌入式系统大都有成本、功耗的要求;嵌入式系统得到多种微处理体系的支持;嵌入式系统需要专用的开发工具。

11.2 DSP

11.2.1 DSP 的概念

信息化的基础是数字化。而数字化的核心技术之一是数字信号处理。数字信号处理的任务在很大程度上需要由 DSP 器件来完成。

DSP 既是 Digital Signal Processing(数字信号处理)的缩写,也是 Digital Signal Processor(数字信号处理器)的缩写。数字信号处理是一种将现实世界中的真实信号(专业术语称之为连续信号)转换为计算机能够处理的信息的过程。比如人们说话的声音,就是一个连续信号。除此之外,现实生活中还有很多这样的信号,比如光、压力、温度等。这些信号通过一个模拟-数字(A/D)的转换过程,变成数字信号送给处理器,进行数字处理,处理结束后,再把结果通过数字-模拟(D/A)的转换过程重新变成连续信号。用一般的通用微处理器可以完成这些工作,但是面临的问题是要满足如此高的计算速度,就很难保证耗电量很低,更难保证价格足够便宜。因此,产生了另一种微处理器:数字信号处理器,简称 DSP。

DSP 是微处理器的一种,这种微处理器具有极高的处理速度,因为这类处理器的应用场合要求具有很强的实时性。比如通过移动电话进行通话时,如果处理速度不够快,就只能等待对方停止说话,这一方才能说话,出现语音严重滞后或抖动。

DSP 典型的特征是:每个处理周期能够处理多次乘法和加法运算,具有实时运算能力、实时仿真能力和实时模拟能力,具有很强的通用性,具有很高的可靠性和性价比。

在当今的数字化时代背景下,DSP 已成为通信、计算机、消费类电子产品等领域的基础器件,被誉为信息社会革新的旗手。业内人士预言,DSP 将是未来集成电路中发展最快的电子产品,并成为电子产品更新换代的关键因素,它将彻底变革人们的工作、学习和生活方式。

11.2.2 DSP 的发展历史

DSP 发展历史大致分为三个阶段:20 世纪 70 年代理论先行,20 世纪 80 年代产品普及,20 世纪 90 年代后突飞猛进。

在 DSP 出现之前,数字信号处理只能依靠微处理器(MPU)来完成。但 MPU 较低的处理速度无法满足高速实时的要求。直到 20 世纪 70 年代,虽然提出了 DSP 的理论和算法基础,但由于

半导体器件的局限,那时的 DSP 大都停留在教科书上,即便研制出来的 DSP 系统也是由分立元件组成的大型系统,其应用领域仅限于军事、航空航天部门。

随着大规模集成电路技术的发展,1982 年世界上诞生了首款 DSP 芯片。这种 DSP 器件采用微米工艺 NMOS 技术制作,虽功耗和尺寸稍大,但运算速度却比 MPU 快了几十倍,尤其在语音合成和编码解码器中得到了广泛应用。DSP 芯片的问世是个里程碑,它标志着 DSP 应用系统由大型系统向小型化迈进了一大步。

20 世纪 80 年代中期,随着 CMOS 工艺技术的进步与发展,第二代基于 CMOS 工艺的 DSP 芯片应运而生,其存储容量和运算速度都得到成倍提高,成为语音处理、图像硬件处理技术的基础器件。

20 世纪 80 年代后期,第三代 DSP 芯片问世,运算速度进一步提高,其应用于范围逐步扩大到通信、计算机领域。

20 世纪 90 年代后 DSP 发展最快,相继出现了第四代和第五代 DSP 芯片。现在的 DSP 属于第五代产品,它与第四代相比,系统集成度更高,将 DSP 核心及外围元件集成在单一芯片上。这种集成度极高的 DSP 芯片不仅在通信、计算机领域大显身手,而且逐渐渗透到人们日常消费领域。

11.2.3 DSP 的特性

经过 30 多年的发展,DSP 产品的应用已扩大到人们的学习、工作和生活的各个方面,并逐渐成为电子产品更新换代的决定因素。举一个例子说明,考虑数字领域里最通常的功能:滤波。简单地讲,滤波就是对信号进行处理,从信号里清除噪声或干扰,改善其信噪比从而改善其特性。

下面通过与采用传统的模拟器件对信号进行滤波的方法相比较,来看看采用 DSP 进行数字滤波的优越性。

模拟滤波器(或模拟电路)的性能受温度等环境因素的影响较大,而数字滤波器则基本上不受环境的影响。

模拟滤波器一旦制造出来,其特性(例如通带频率范围)不易改变。使用微处理器实现数字滤波器,可以通过对其重新编程来改变滤波的特性。

数字滤波器易于在非常小的宽度内进行复制,即适合模块化的结构,实现大规模电路集成。

模拟信号处理与数字信号处理方法的详细比较如表 11-1 所示。

表 11-1 模拟信号处理与数字信号处理的比较

比较因素	模拟方式	数字方式
修改设计的灵活性	修改硬件设计,或调整硬件参数	改变软件设置
精度	由元器件精度确定	由 A/D 的位数和计算机字长、算法确定
可靠性和可重复性	受环境温度、湿度、噪声、电磁场等的干扰和影响大	不受这些因素的影响

比较因素	模拟方式	数字方式	续表
大规模集成	尽管已有一些模拟集成电路,但品种较少,集成度不高、价格较高	便于大规模集成,且 DSP 器件体积小、功能强、功耗小、一致性好、使用方便、性能/价格比高	
实时性	除电路引入的延时外,处理实时	由计算机的处理速度决定	
高频信号的处理	可以处理包括微波毫米波乃至光波信号	按照奈奎斯特准则的要求,受取样/保持(S/H)、A/D 转换和处理速度的限制	

与通用处理器相比,DSP 采用了不同的存储器结构,并设计了特殊的专用指令集,从而更适用于数字信号处理的实时计算。考虑一个数字信号处理的实例,比如有限冲击响应滤波器(FIR)。用数学语言来说,FIR 滤波器实际上是做一系列的点积。将一个输入量和一个系数向量,在系数和输入样本的滑动窗口内做乘法运算,然后将所有的乘积加起来,形成一个输出值。类似的运算在数字信号处理过程中大量地重复发生,为此有必要设计一种提供专门支持的器件,于是促成了 DSP 处理器与通用处理器(GPP)的分流。两者性能的比较如下。

(1) 对密集乘法运算的支持

GPP 不是设计用来做密集乘法运算的,即使是一些现代的 GPP,也要求多个指令周期来完成一次乘法。而 DSP 处理器使用专门的硬件来实现单周期乘法。DSP 处理器还增加了累加器寄存器来处理多个乘积的和。累加器寄存器通常比其他寄存器位宽,并增加额外的位来避免累加溢出。同时,为了充分体现专门的乘法—累加硬件的好处,几乎所有的 DSP 指令集都包含有显式的 MAC(乘累加)指令。

(2) 存储器结构

传统的 GPP 使用冯·诺依曼存储器结构。这种结构中,只有一个存储器空间通过一组总线(一对地址总线和数据总线)连接到处理器核心。通常,做一次乘法运算会发生 4 次存储器访问,用掉至少 4 个指令周期。大多数 DSP 采用了哈佛结构,将存储器空间划分成两个,分别存储程序和数据。它们有两组总线连接到处理器,允许同时对它们进行访问。这种安排将处理器存储器的带宽加倍,更重要的是同时为处理器提供数据与指令。在这种布局下,DSP 得以实现单周期的 MAC 指令。现在典型的高性能 GPP 实际上已包含两个片内高速缓存,一个存储数据,一个存储指令,它们直接连接到处理器核心,以加快运行时的访问速度。从物理上说,这种片内的双存储器和总线的结构几乎与哈佛结构的一样了。然而从逻辑上说,两者还是有重要的区别。

GPP 使用控制逻辑来决定哪些数据和指令字存储在片内的高速缓存里,程序员并不加以指定(也可能根本不知道)。与此相反,DSP 使用多个片内存储器和多组总线来保证每个指令周期内多次访问存储器。在使用 DSP 时,程序员要明确地控制哪些数据和指令要存储在片内存储器中。程序员在写程序时,必须保证处理器能够有效地使用其双总线。

此外,DSP 处理器几乎都不具备数据高速缓存。这是因为 DSP 的典型数据是数据流。也就是说,DSP 处理器对每个数据样本做计算后,就丢弃了,几乎不再重复使用。

(3) 零开销循环

所谓零开销循环是指 DSP 处理器在执行循环时,不用花时间去检查循环计数器的值、条件转移到循环的顶部,而是通过专门的硬件将循环计数器减 1。

与此相反,GPP 的循环一般使用软件来实现。某些高性能的 GPP 使用转移预报硬件,几乎达到与硬件支持的零开销循环同样的效果。

(4) 定点计算

大多数 DSP 使用定点计算,而不是浮点计算。虽然 DSP 的应用必须十分注意数字的精确,用浮点计算应该比较合理,但是对 DSP 来说,廉价和快速是非常重要的。定点处理器比浮点处理器更便宜,而且更快。为了不使用浮点处理器而又保证数字的精度,DSP 处理器在指令集和硬件方面都支持饱和计算、舍入和移位。

(5) 专门的寻址方式

DSP 处理器往往都支持专门的寻址方式,这对通常的信号处理操作和算法是很有用的,例如,模块(循环)寻址(对实现数字滤波器的时延线很有用)、位倒序寻址(对 FFT 很有用)。这些非常专门的寻址方式在 GPP 中是不常使用的,只能用软件来实现。

(6) 执行时间的预测

大多数的 DSP 应用(如蜂窝电话和调制解调器)都是严格的实时应用,所有的处理必须在指定的时间内完成。这就要求程序员准确地确定处理每个样本需要多少时间,或者,至少要知道,在最坏的情况下,需要多少时间。

如果用低成本 GPP 去完成实时信号处理的任任务,执行时间的预测一般会成为什么问题,因为低成本 GPP 具有相对直接的结构,执行时间比较容易预测。然而,低成本 GPP 不能支持大多数实时 DSP 应用所要求的处理能力。DSP 处理器对高性能 GPP 的优势在于,即便使用了高速缓存的 DSP,哪些指令会放进去也是由程序员(而不是处理器)来决定的,因此很容易判断指令是从高速缓存还是从存储器中读取。DSP 处理器一般不使用动态特性,如转移预测和推理执行等。因此,给定一段代码的执行时间是可预测的,从而使程序员得以确定芯片的性能限制。

(7) 定点 DSP 指令集

定点 DSP 指令集是按两个目标来设计的:使处理器能够在每个指令周期内完成多个操作,从而提高每个指令周期的计算效率。将存储 DSP 程序的存储器空间减到最小(存储器对整个系统的成本影响甚大,在对成本敏感的 DSP 应用中这点尤为重要)。

为了实现这些目标,DSP 处理器的指令集通常允许程序员在一个指令内说明若干个并行的操作。例如,一条 MAC 指令同时包含了一次乘法和累加操作,在典型的 DSP 处理中,一条指令就包含了计算 FIR 滤波器中一延迟节所需要的所有操作。与 GPP 的指令集相比,DSP 的指令集效率高,但直观性差,且不容易使用。GPP 的程序一般使用 C 或 C++ 等高级语言,通常不必在意处理器的指令集是否容易使用。DSP 应用程序通常都是用汇编语言写的(至少部分是汇编语言优化的),因为大多数广泛使用的高级语言,例如 C,并不适合于描述典型的 DSP 算法。同时,DSP 结构的复杂性,如多存储器空间、多总线、不规则的指令集、高度专门化的硬件等,使得难于为其编写高效率的编译器。即便用编译器将 C 源代码编译成为 DSP 的汇编代码,优化的任务仍然很重。典型的 DSP 应用都具有大量计算的需求,并有严格的开销限制,使得程序的优化仍然必不可少(至少是对程序的最关键部分是如此)。

(8) 开发工具的要求

由于 DSP 应用要求高度优化的代码,大多数 DSP 厂商都会提供一些开发工具,以帮助程序员完成其优化工作。例如,大多数厂商都提供处理器的仿真工具,以准确地仿真每个指令周期内处理器的活动。无论对于确保实时操作还是代码的优化,这些都是很有用的工具。

GPP 厂商通常并不提供这样的工具,主要是因为 GPP 程序员通常并不需要详细到这一层的信息。GPP 缺乏精确到指令周期的仿真工具,由于几乎不可能预测高性能 GPP 对于给定任务所需要的周期数,因而无法说明如何去改善代码的性能。

11.3 SOPC 及其技术

微电子技术的近期发展成果,为 SOC(System On Chip)的实现提供了多种途径。对于经过验证而又具有批量的系统芯片,可以做成专用集成电路 ASIC 而大量生产。而对于一些仅为小批量应用或处于开发阶段的 SOC,若马上投入流片生产,需要投入较多的资金,承担较大的试制风险。最近发展起来的 SOPC(System On Programmable Chip)技术则提供了另一种有效的解决方案,即用大规模可编程器件的 FPGA(Field Programmable Gate Array)来实现 SOC 的功能。可编程逻辑器件产生于 20 世纪 70 年代。其出现的最初目的是为了用较少的 PLD(Programmable Logic Device)品种替代种类繁多的各式中小规模逻辑电路。经 30 多年的发展过程,PLD 的结构、工艺、功耗、集成规模和工作速度等都得到了重大进步。尤其是在 20 世纪 90 年代,出现了大规模集成度的 FPGA,单片的集成度由原来的数千门,发展到数十万甚至数百万门。芯片的 I/O 接口也由数十个发展至上千个。有的制造商还推出了含有硬核嵌入式系统的 IP。

所谓 IP 核是指知识产权(Intellectual Property, IP)。美国 Dataquest 咨询公司将半导体产业的 IP 定义为用于 ASIC、ASSP 和 PLD 等当中,并且是预先设计好的电路模块。IP 核模块有行为(Behavior)、结构(Structure)和物理(Physical)三级不同层次的设计,对应描述功能行为的不同分为三类,即硬核(Hard IP Core)、完成结构描述的固核(Firm IP Core)和基于物理描述并经过工艺验证的硬核(Hard IP Core)。IP 硬核通常是用硬件描述语言(HDL)文本形式提交给用户,它经过寄存器传输层(RTL)级设计优化和功能验证,但其中不含有任何具体的物理信息。据此,用户可以综合出正确的门电路级设计网表,并可以进行后续的结构设计,具有很大的灵活性,借助于电子设计自动化(EDA)综合工具可以很容易地与其他外部逻辑电路合成一体,根据各种不同半导体的工艺,设计成具有不同性能的器件。IP 硬核也称为虚拟组件(Virtual Component, VC)。在功能验证时,硬核都可以直接仿真,扫描链也由自己插入,硬核与工艺无关。IP 硬核是基于半导体工艺的物理设计,已有固定的拓扑结构和具体工艺,并已经过工艺验证,具有可保证的性能。其提供给用户的形式是电路物理结构掩模版图和全套工艺文件,是可以拿来就用的全套技术。IP 硬核的设计程度则是介于硬核和硬核之间,除了完成硬核所有的设计外,还完成了门级电路综合和时序仿真等设计环节,以门级电路网表的形式提供给用户。

因此,已完全可能将一个电子系统集成到一片 FPGA 中,即 SOPC,为 SOC 的实现提供了一种新的简单易行而又成本低廉的手段,极大地促进了 SOC 的发展。SOPC 技术最早于 2000 年由美国 Altera 公司提出,并推出了相应的开发软件 Quartus II。SOPC 是基于 FPGA 解决方案的 SOC。与 ASIC 的 SOC 解决方案相比, SOPC 系统及其开发技术具有更多的特色,构成 SOPC 的方

案也有如下多种途径。

1. 基于 FPGA 嵌入 IP 硬核的 SOPC 系统

即在 FPGA 中预先植入嵌入式系统处理器。目前最为常用的嵌入式系统大多采用了含有 ARM 的 32 位 IP 核的器件。尽管由这些器件构成的嵌入式系统有很强的功能,但为了使系统更为灵活完备,功能更为强大,对完成更多任务具有更好的适应性,通常处理器必须配置许多接口器件才能构成一个完整的应用系统。如配置常规的 SRAM、DRAM、Flash 外,还必须配置网络通信接口、串行通信接口、USB 接口、VGA 接口、PS/2 接口或其他专用接口等。这样会增加整个系统的体积、功耗,从而降低系统的可靠性。但是如果将 ARM 或其他 IP 核以硬核方式植入 FPGA 中的同时,利用 FPGA 中的可编程逻辑资源、IP 软核及逻辑宏单元来构建该嵌入式系统处理器的接口功能模块,就能一并解决这些问题。为此,Altera 和 Xilinx 公司都相继推出了这方面的器件。例如,Altera 的 Excilibur 系列 FPGA 中就植入了 ARM922T 嵌入式系统处理器;Xilinx 的 Virtex-II Pro 系列 FPGA 中则植入了 IBM PowerPC 405 处理器。这样就能使得 FPGA 灵活的硬件设计和硬件实现与处理器的强大软件功能有机地结合,高效地实现 SOPC 系统。

2. 基于 FPGA 嵌入 IP 硬核的 SOPC 系统

将 IP 硬核直接植入 FPGA 的解决方案尚不够完美。

① 由于此类硬核多来自第三方公司, FPGA 厂商通常无法直接控制其知识产权费用,从而导致 FPGA 器件价格相对偏高。

② 由于硬核是预先植入的,设计者无法根据实际需要改变处理器的结构,如总线规模、接口方式,乃至指令形式,更不可能将 FPGA 逻辑资源构成的硬件模块以指令的形式形成内置嵌入式系统的硬件加速模块(如 DSP 模块),以适应更多的电路功能要求。

③ 无法根据实际需求需求在同一 FPGA 中使用多个处理器核。

④ 无法裁减处理器资源以降低 FPGA 成本。

⑤ 只能在特定的 FPGA 中使用硬核嵌入式系统,如只能使用 Excilibur 系列 FPGA 中的 ARM 核, Virtex-II Pro 系列中的 PowerPC 核。

如果利用硬核嵌入式系统处理器就能有效地克服上述不利因素。目前最有代表性的硬核嵌入式处理器分别是 Altera 的 Nios 和 Nios II 核及 Xilinx 的 MicroBlaze 核。特别是前者,即 Nios 处理器,使上述 5 方面的问题得到很好的解决。Altera 的 Nios 核是用户可随意配置和构建 32 位/16 位总线(用户可选的)指令集和数据通道的嵌入式系统微处理器 IP 核,采用 Avalon 总线结构通信接口,带有增强的内存、调试和软件功能(C 或汇编程序优化开发功能);含有 First Silicon Solutions(FS2)开发的基于 JTAG 的片内设备(OCI)核(这为开发者提供了强大的软件调试实时代码,OCI 调试功能可根据 FPGA JTAG 接口上接收的指令,直接监视和控制片内处理器的运行情况)。此外,基于 Quartus II 平台的用户可编辑的 Nios 核含有许多可配置的接口模块,包括可配置高速缓存模块——包括片内嵌入式系统块(Embedded System Block, ESB)、外部 SRAM 或 SDRAM,具有 100 MB 以上单周期访问速率——可配置 RS-232 接口、SDRAM 控制器、标准以太网协议接口、DMA、定时器、协处理器等。在植入(配置进)FPGA 前,用户可根据设计要求,利用 Quartus II 和 SOPC Builder,对 Nios 及其外围系统进行构建,使该嵌入式系统在硬件结构、功能特点、资源占用等方面全面满足用户系统设计的要求。Nios 核在同一 FPGA 中被植入的数量

没有限制,只要 FPGA 的资源允许。就成本而言,由于 Nios 是由 Altera 直接推出而非第三方产品,故用户通常无需支付知识产权费用,Nios 的使用费仅仅是其占用的 FPGA 逻辑资源费。因此,选用的 FPGA 越便宜,则 Nios 的使用费就越便宜。特别值得一提的是,用户可以使用 Matlab 和 DSP Builder,或直接使用 VHDL 等硬件描述语言,为 Nios 嵌入式处理器设计各类加速器,并以指令的形式加入 Nios 的指令系统,从而使其成为 Nios 系统的一个接口设备,与整个片内嵌入式系统融为一体。例如,用户可以根据设计项目的具体要求,自主地构建自己的 DSP 处理器系统,而不必拘泥于其他已上市的有限款式的 DSP 处理器。

3. 基于 HardCopy 技术的 SOPC 系统

通过强化 SOPC 工具的设计能力,在保持 FPGA 开发优势的前提下,引入 ASIC 的开发流程,从而向 ASIC 转化。这就是 Altera 推出的 HardCopy 技术。

HardCopy 利用原有的 FPGA 开发工具,将成功实现于 FPGA 器件上的 SOPC 系统通过特定的技术直接向 ASIC 转化,从而克服了传统 ASIC 设计中普遍存在的问题。其中包括开发周期长、产品上市慢、一次性成功率低、最少的投片量要求、设计软件工具繁多且昂贵、开发流程复杂等。例如,一类 ASIC 开发,首先要求有高的技术人员队伍、高达数十万美元的开发软件费用和高昂的掩模费用,且整个设计周期可能长达一年。ASIC 设计的高成本和一次性的低成功率很大部分是由于需要设计和掩模的层数太多(多达十几层)造成,如果利用 HardCopy 技术设计 ASIC,开发软件费用仅 2000 美元(Quantus II),SOC 级规模的设计周期不超过 20 周,转化的 ASIC 与用户设计习惯的掩模层只有两层,且一次性投片的成功率近乎 100%,即实现了所谓的 FPGA 向 ASIC 的无缝转化。而且用 ASIC 实现后的系统性能将比原来在 HardCopy FPGA 上验证的模型又提高近 50%,而功耗则降低 40%。一次性成功率的大幅提高即意味着设计成本的大幅降低和产品上市速度的大幅提高,三种 SOC 方案的比较如表 11-2 所示。

表 11-2 三种 SOC 方案的比较

项目	基于 ASIC 的 SOC	基于 FPGA 的 SOC(SOPC)	基于 HardCopy 的 SOC
单片成本	低	较高	较低
开发周期	长(超过 50 周)	短(少于 10 周)	较短(少于 20 周)
开发成本	设计工程成本高,掩模成本高,软件工具成本高(超过 30 万美元)	设计工程成本低,无掩模成本,软件工具成本低(低于 2000 美元)	设计工程成本低,掩模成本低,软件工具成本低(低于 2000 美元)
集成技术	0.25 μm ~ 65 nm	0.25 μm ~ 90 nm	0.25 μm ~ 90 nm
可重构性	不可重构	可重构	不可重构

HardCopy 技术作为一种全新的 SOC 级 ASIC 设计解决方案,将专用的硅片设计和 FPGA 至 HardCopy 自动迁移过程结合在一起,首先利用 Quantus II 将系统模型成功实现于 HardCopy FPGA 上,然后帮助设计者把可编程解决方案无缝地迁移到低成本的 ASIC 上。这样,HardCopy 器件就把大容量 FPGA 的灵活性和 ASIC 的市场优势结合起来,从而避免了直接设计 ASIC 的困难。从原型设计提升至产品制造,可采用 FPGA 的专有迁移技术,十分容易地将 FPGA 的设计移植到 HardCopy 器件上,达到降低成本、避免了 ASIC 的风险、加快面市周期的目的。HardCopy 器件

(如 HardCopy Stralix 系列、Excalibur 系列的 FPGA)本质上是 FPGA 的精确复制,剔除了可编程性,专用配置和采用金属互连的走线,器件的硅片面积就更小,成本更低,并且还改善了时序特性。

11.4 微控制器

11.4.1 什么是微控制器

微控制器(Micro Controller Unit, MCU),又称单片微型计算机(Single Chip Microcomputer),是指随着大规模集成电路的出现及其发展,将计算机的 CPU、RAM、ROM、定时数器和多种 I/O 接口集成在一片芯片上,形成芯片级的计算机,为不同的应用场合执行不同组合的控制。

微控制器的存储器可分为掩模 ROM、一次性可编程 ROM、快闪 ROM 等类型。掩模 ROM 的微控制器价格便宜,但程序在出厂时已经固化,适合程序固定不变的应用场合;快闪 ROM 的微控制器程序可以反复擦写,灵活性很强,但价格较高,适合对价格不敏感的应用场合或做开发用途;一次性可编程 ROM 的微控制器价格介于前两者之间,因具有一次性可编程能力,适合既要求一定灵活性,又要求低成本的应用场合,尤其是功能不断翻新、需要迅速量产的电子产品。

微控制器经过多年不断地研究、发展,历经 4 位、8 位,到现在的 16 位及 32 位,甚至 64 位。产品的成熟度,投入厂商之多,应用范围之广,真可谓之空前。由于制造程序的改进,8 位微控制器与 4 位微控制器价差相去无几,8 位微控制器已渐成为市场主流,目前 4 位微控制器大部分应用用在计算器、车用仪表、车用防盗装置、呼叫器、无线电话、CD 播放器、LCD 驱动控制器、LCD 游戏机、儿童玩具、磅秤、充电器、胎压计、温湿度计、遥控器及傻瓜相机等;8 位微控制器大部分应用用在电表、马达控制器、电动玩具机、变频式冷气机、呼叫器、传真机、来电辨识器(CallID)、电话录音机、CRT 显示器、键盘及 USB 等;16 位微控制器大部分应用在移动电话、数码相机及录像机等;32 位微控制器大部分应用在 Modem、GPS、PDA、HPC、机顶盒(Set-Top Box)、集线器、网桥、路由器、工作站、ISDN 电话、激光打印机与彩色传真机;64 位微控制器大部分应用用在高级工作站、多媒体互动系统、高级电视游乐器(如 SEGA 的 Dreamcast 及 Nintendo 的 GameBoy)及高级终端机等。

在微控制器开发方面,以架构而言,可分为两大主流:RISC(如 HOLTEK 的 HT48xxx 系列)与 CISC(如华邦的 W78 系列)。RISC(Reduced Instruction Set Computer,精简指令集计算机)表示微控制器的所有指令都是利用一些简单的指令组成的,使得微控制器的线路可以更易做到最佳,从而提高执行速率,使得执行一个指令所需的时间减到最短。HOLTEK 的 HT46XX(A/D 微控制器系列) HT47xx(R 到 F 微控制器系列) HT48xx(一般 I/O 微控制器系列) HT49xx(LCD 微控制器系列)便是采用 RISC 结构来设计。对于 RISC 微处理器来说,由于指令集的精简,许多工作都必须组合简单的指令,因此对于较复杂组合的工作需要通过编译器来执行,而对于 CISC 微控制器,由于硬件所提供的指令集较多,所以许多工作可通过一个或是数个指令来实现,因而减少了许多编译工作。例如对于一个积分表达式,使用 CISC 指令集的微控制器运算可能只需要十个机器指令,而 RISC 微控制器在执行相同的程序时,由于 CPU 本身不提供浮点数乘法的指令,可

需要执行上百个机器指令(但每一个指令可能只需要CISC指令十分之一的时间),而程序语言转换成机器指令的动作是由编译器来执行的,所以在RISC微控制器上执行,编译器比较复杂。RISC与CISC之间的取舍,似乎也是微控制器硬件架构与软件(编译器)的平衡之争,应该没有绝对优势的一方,只能说因应不同的需求而而有不同的产品。

11.4.2 微控制器的基本结构

微控制器产品架构由早期以累加器为基础的CPU,演进至现今含RISC指令或同时含RISC指令和DSP的CPU,如Motorola的68356,也有如DEC的SAIIC、与Hitachi的SH-DSP系列等32位嵌入式微控制器,每一系列产品又因应不同的应用与接口需求,衍生出不同规格的产品。

微控制器的基本结构如图11-3所示。

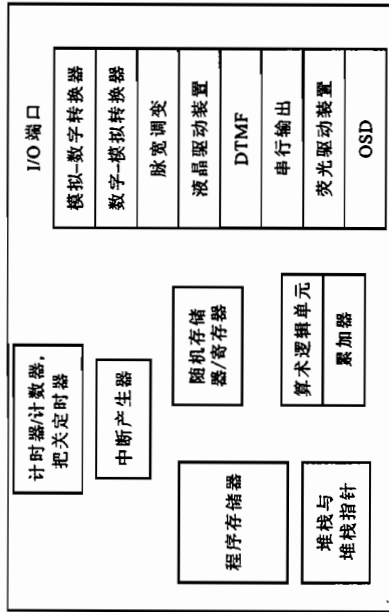


图 11-3 微控制器的基本结构

1. 程序存储器

微控制器中的程序内存是只可读取而不可写入的存储单元,主要用来存放用户所开发的程序,微控制器的动作便是依据存储于其中程序指令来执行的。在8位单芯片中常见的程序内存容量有0.5KB、1KB、2KB、4KB及8KB,而通常应用在微控制器的程序内存有下列几种,见表11-3。

程序存储器	特征
掩模ROM	在ROM的制作中,利用光罩将记忆内容设定好,程序代码在制造过程中就直接放入芯片中,属于量产阶段使用,掩模费用需数千美元,量产时的成本却相当低。

程序存储器	特征
一次可编程(OTP)ROM	OTP ROM是PROM组件,只容许烧录一次,可使用标准的EPROM烧录器或者OTP ROM提供厂商所提供的烧录器。一旦烧录完成后就无法清除或修改烧录内容,较适合试产或小量生产阶段使用。OTP ROM的单颗价格约是掩模ROM的1.5~2倍。
具有开窗的EPROM	提供用户更改程序的空间,可用紫外线透过开窗清除程序,可重复烧写,包装成本非常高,仅适合少量生产或实验室使用。
EEPROM	此类内存可重复写入或清除,故可以删除或修改其中的程序,而无须使用开窗之包装,可节省包装之成本,亦方便重复使用,但生产制程较复杂。
Flash EPROM	当需要能够清除或写入较大量程序的非易失性内存时,Flash PROM比传统的EEPROM可提供较好的解决之道,因为Flash EPROM可擦写次数及速度方面表现更好。利用Flash PROM作为程序内存,由于不需要EPROM特殊的开窗式陶瓷包装,价格上与OTP ROM相差不多,又具有多次重复烧写的功能。

为了防止程序被未经授权者窃取、反汇编、篡改等,须在OTP ROM、EPROM、EEPROM及Flash ROM上选择程序保密措施(可采用加密或熔丝的方法),因为这些存储器可借由烧录器来读取程序代码。掩模ROM则不需要进行程序保密。

2. 随机存储器

随机存储器(Random Access Memory, RAM)用来暂时存放数据或程序。以HOLTEK 8位微控制器为例,其RAM容量有64B、96B、160B及224B。

3. 累加器

累加器(Accumulator)是微控制器的运作中枢,80%的指令都与累加器有关;数据可以被存放在累加器中,直到总线或其他单元或程序读取其中的数据为止。

4. 寄存器

寄存器(Register)是微控制器内部用来暂时存放数据的地方,每个寄存器的功能各不相同,但却有一些共同的特性,就是可以直接读写,因其位于微控制器的内部,故减少了一些不必要的等待及寻址时间。有些微控制器的I/O端口也可以像寄存器一样直接存取。

5. 堆栈及堆栈指针

堆栈(Stack)及堆栈指针(Stack Pointer)又称为后进先出队列(Last-In-First-Out Queues),其功能如下。

- ① 暂时存放程序计数器(Program Counter, PC)的值,适用于子程序调用或中断发生时将程序计数器的值暂时存储起来。
- ② 可作为寄存器使用,以指令Push、Pop来完成。
- ③ 有些微控制器的堆栈级数是固定的(如HOLTEK μ C系列),有些则可自定义。

6. 算术逻辑单元

算术逻辑单元 (Algorithm Logic Unit, ALU) 的功能是执行算术运算及逻辑运算,除了产生结果之外,也产生相关的标志(分别针对零、进位、借位、状态),每一个微控制器也不完全一样,尤其是进位标志,一定要查看指令说明表。

7. 输入输出(I/O)端口

在单片机应用系统中,I/O的扩充不是目的,而是为了给外部设备提供一个输入输出通道,作为外界与微控制器间的沟通管道。例如接入键盘、显示器、驱动开关控制或测量等。在I/O扩充时必须考虑与之相连接的外围设备硬件的电路特性,如电平匹配、干扰抑制、驱动能力等。

微控制器的I/O端口有多种电路形式,其中有许多端口可以由软件以比特位来设定输入与输出方向。I/O端口附加有大电流、耐高压的缓冲器,以直接驱动LED、高功率晶体管以及模拟信号的输入输出等。

8. 计时器与定时器

计时器 (Time Counter) 与定时器 (Timer),通过外加振荡晶体,经分频电路为微控制器提供数个不同的时基 (Time Base)。常应用于时钟之时基 (如 1 s、500 ms、62.5 ms、15.625 ms 等)、脉宽调制 (Pulse Width Modulation, PWM) 之时基、键盘和 LED 扫描、频率输出、自动关机 (Auto Power OFF, APO) 等时基。

9. 计数器

计数器 (Event Counter) 专用于累计外部的事件个数,可能为脉冲或其他数据,也可用以产生正确的时间延迟。常用于产生另一种时基,外加固定频率;可规划成另一种立即中断输入以及速度计 (Speed Meter)、转速表 (Tachometer) 等。

10. 把关定时器 (Watchdog Timer)

把关定时器用来监督微控制器是否处于不正常的工作状态。当微控制器受到噪声干扰或操作不当出现异常时,需有防范措施确保微控制器能够自动重置,让微控制器能够继续运作。许多微控制器都已把它列为标准配置。

把关定时器实际上就像一个自跑式的 RC 振荡器,不管芯片的振荡脉冲有无停止,它还是继续计数而不随中断而停止,即便是芯片已进入省电的暂停状态 (在暂停状态下,芯片的定时脉冲停止振荡,而把关定时器并不停止计时,当计时逾时将使本芯片自动重置,I/O 引脚输出保持不变,保持相当低的耗电)。

11. 中断

中断 (Interrupt) 用来处理立即事件或列为优先处理的事件,负责时间计数器超时中断及外部事件产生中断请求等工作。大部分微控制器的中断处理系统是多层的,内设有关中断优先级电路,以决定先后顺序。常应用于:微控制器处于等待状态时,由外加信号唤醒;需要立即处理由传感器、开关、警报器、电源故障预警器等传来的信息;需要一个固定间隔来处理显示、键盘扫描、实时时钟等相关事件的情况。

12. 微控制器上的外围设备

微控制器还可以加挂更多外围设备,以扩充、延伸其功能。主要包括以下设备。

(1) 串行 I/O 端口

微控制器内含串行 I/O 端口用于为外围设备提供通信通道。常见的有以下几种:

- ◆ 通用异步收发器 (Universal Asynchronous Receiver Transmitter, UART)
- ◆ 通用同步/异步收发器 (Universal Synchronous/Asynchronous Receiver Transmitter, USART)
- ◆ 串行外围接口 (Serial Peripheral Interface, SPI)
- ◆ 串行通信接口 (Serial Communications Interface, SCI), 这是 UART 的加强版
- ◆ 集成电路间总线 (Inter Integrated Circuit bus, I²C bus)
- ◆ 三线同步串行接口 (MicroWire/Plus)

(2) 液晶驱动装置 (LCD Driver)

在显示接口上,LCD 是常用的显示装置,例如,在一些多功能的电话、数字温度计、呼叫器、掌上型游戏机上皆可以发现它的踪迹。因此内含 LCD 驱动线路的微控制器厂商提供多样内置 LCD 驱动装置的微控制器可供选用,HOLTEK HT49XX 系列也提供 LCD 驱动装置的微控制器。

(3) 真空荧光管驱动装置 (VFT Driver)

LCD 显示器在无光源或无背光的环境下,无法阅读显示器的内容,而真空荧光管 (Vacuum Fluorescent Tube, VFT) 显示器可提供高亮度、且色彩丰富的视觉效果,常应用于高级的家电产品上,如影碟机、DSP 均衡器。

(4) OSD

屏幕显示开关 (On Screen Display, OSD) 是电视及监视器的人性化接口,其接收水平同步信号 (H-Sync) 及垂直同步信号 (V-Sync),并通过红绿蓝三色及消隐信号将屏幕信息送出,其显示颜色至多可达 8 种。微控制器指令执行速率会造成 OSD 的显示行数及字段的,显示行数由两行至数十行,字段则有 15 ~ 26 个字符或更多,通常执行速率较快者可显示较多的行数、字段,速率较慢者在显示上会有直接的受限。

(5) 模拟 - 数字转换器 (ADC)

模拟 - 数字转换器 (Analog to Digital Converter, ADC) 将时间连续的模拟信号转换为微处理器能处理的数字信号,在工业及消费电子上都有很广泛的应用。

(6) 数字 - 模拟转换器 (DAC)

数字 - 模拟转换器 (Digital to Analog Converter) 是上述 ADC 过程的逆过程。

(7) 脉宽调制 (PWM)

脉宽调制 (Pulse Width Modulation, PWM) 是一种开关电源的供电控制技术。通过控制开关电源的充放电开关的导通比,调制脉冲宽度,进而调整输出电压,以满足主板和显示器等对电压的不同需求。具有转换效率高、响应速度快等特点,还能实现过压、过流保护等功能。

(8) DTMF 产生器、接收器

双音多频 (Dual Tone Multiple Frequency, DTMF) 产生器将以数字 0 和 1 编码的脉冲电话呼叫转换为两种音频频率编码的混合信号呼叫,不但提高了译码的速度,也增加了可靠性与抗噪声能力。

13. 双时钟(Dual Clock)

微控制器的时钟脉冲频率愈高,耗电量就愈大。当必须考虑低功耗时,应该考虑使用双时钟技术。

综上所述,微控制器的指令集结构较单纯、容易开发及修改、I/O及中断处理能力强,主要工作集中在软件设计上,只要通过开发系统即可进行在线仿真(In-Circuit Emulator, ICE),进行设计及修改的工作。微控制器具备价格低、系统硬件架构简单、应用程序的开发及修改容易、芯片稳定度佳、可靠性好等优点,其应用领域极广。

作为本节小结,将MCU和前一节的DSP做一比较,如下表11-4所示。

表 11-4 MCU 与 DSP 的性能比较

性能指标	MCU			DSP
	低档	高档	低档	高档
指令周期(ns)	600	40	50	5
乘加时间(ns)	1 900	80	50	5
US\$/MIPS	1.5	0.5	0.15	0.1

11.5 工业控制系统

11.5.1 什么是工业控制系统

工业控制系统是计算机,尤其是专用计算机和处理器(如上述的嵌入式系统、数字信号处理器、片上可编程系统及单片微机系统等)最广泛的应用领域,一般为以微处理器为基础的集中分散控制系统,其主要特征是集中管理和分散控制;与采用直接数字控制或监督计算机控制相比,可将因计算机故障造成的风险分散,使风险减小到最小的程度。分散控制系统具有控制功能多样、操作简便、系统便于扩展、维护方便、可靠性高、便于与其他计算机联网等特点。

图11-4所示为1975年最先推出分散控制系统的 Honeywell 公司的 TDC-3000 系统的结构框图。由图可知,整个系统分为三级。最高级为工厂信息网,构成了整个工厂的管理信息系统(Management Information System, MIS),并通过工厂信息网接口与 TDC-3000 系统的主干网——局域网控制网(Local Control Network, LCN)相连;第二级为局域网接口网,通过高速通道(Hiway)接口与数据高速通道(Data Hiway)相连,或通过网络接口与通用控制网(Universal Control Network, UCN)相连,或通过可编程控制器接口连接可编程控制器;最下层可为与数据高速通道或通用控制网相连的现场管理和控制设备及可编程控制器等。

11.5.2 分散控制系统的特点

1. 分散控制系统的优点

(1) 分级递阶控制

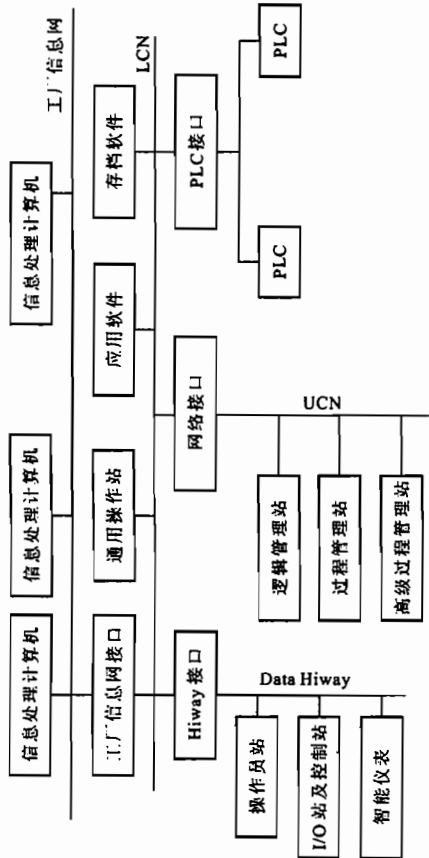


图 11-4 TDC-3000 系统结构

分散控制系统是分级递阶控制系统。它在垂直方向或水平方向都是分级的;最简单的分散控制系统至少在垂直方向分为两级,即操作管理级和过程控制级。在水平方向上各个过程控制级之间是相互协调的分级,它们把数据向上送达操作管理级,同时接收操作管理级的指令;各个水平分级间也进行数据的交换,这样的系统是分级的递阶系统。分散控制系统的规模越大,系统的垂直和水平分级的范围也越广。分级递阶系统的优点是各个分级具有各自的分工范围,相互之间有协调,通常这种协调是通过上一分级来完成的。

(2) 分散控制

分散控制是分散控制系统的另一特点,分散是针对集中而言的。在计算机控制系统的应用初期,控制系统是集中式的,即直接数字控制(DDC)。分散的含义包括:人员地域的分散,设备、控制、操作分散,数据库的分散及风险的分散等。分散的目的是为了提高设备的利用率和分散风险,使风险降到最低程度。

(3) 自治和协调性

分散控制系统的各组成部分是各自为政的自治系统,它们各司其职,相互间又有联系,数据信息相互交换,各种条件相互制约,在系统的协调下工作。

分散控制系统一个部分的故障也会影响其他部分,例如,操作管理站的故障将使操作人员无法知道过程的运行情况;通信系统的故障会使数据传送出错;导致现场控制站的故障使系统无法正常工作。应该指出,不同部件的故障对整个系统影响的大小是不同的,为此,在分散控制系统选型和系统配置时,应考虑重要的部位设置较高可靠性的部件或有必要的冗余措施等。

分散的基础是分散的系统应是自治的系统。递阶分级的基础是分散的系统是相互协调的系统。分散中又有集中,集中的数据管理、集中的控制目标、集中的显示屏幕、集中的通信管理等。

2. 分散控制系统的类型

分散控制系统由三大部分组成。根据现场控制站、操作管理站和通信系统的不同结构,分散控制系统大致可分为下列几类。

(1) 产品结构类型

1) 模块化现场控制站 + 与移动访问协议(MAP)兼容的局域网 + 信息综合管理系统
这是一类最新结构的分散控制系统。作为大系统,通过局域网可在较广的地域内应用。通过现场总线,系统可与现场智能仪表通信和操作。这是一种开放的系统互连的具有互操作性的系统,将成为分散控制系统的主体结构,是第三代分散控制系统的典型结构,即现场总线控制系统(Fieldbus Control System, FCS)。

2) 现场控制站 + 局域网 + 信息管理系统

由于采用局域网技术,使通信性能提高,联网能力增强。这是第二代分散控制系统的典型结构。

3) 现场控制站 + 高速数据通道 + 操作站 + 上位机

这是第一代分散控制系统的典型结构。经过对操作站、现场控制站、通信系统性能的改进和扩展,系统的性能已有较大提高。

4) 可编程逻辑控制器(PLC) + 通信系统 + 操作管理站

这是一种在制造业广泛应用的分散控制系统的结构。尤其适用于有大量顺序控制的工业生产过程。分散控制系统制造商为适应顺序控制实时性强的特点,现已有不少产品可以下挂各种型号的可编程逻辑控制器,组成可编程逻辑控制器 + 分散控制系统的形式,应用于有实时要求的顺序控制和较多回路的连续控制的场合。

5) 单回路控制器 + 通信系统 + 操作管理站

这是一种适用于中、小企业的小型分散控制系统结构。它用单回路控制器(或双回路、四回路控制器)作为盘装仪表。信息的监视操作由操作管理站或仪表面板实施,有较大灵活性和高性价比。

(2) 实际应用中的结构类型

在实际应用中,采用微处理器、工业级微机组成分散控制系统的结构可分为如下几类。

1) 工业级微机 + 通信系统 + 操作管理机

工业级微机,作为多功能多回路的现场控制站。相应的软件也已有软件厂商开发。

2) 单回路控制器 + 通信系统 + 工业级微机

工业级微机作为操作管理站使用。它的通用性较强。软件可自行开发,相应的管理、操作软件也有产品可购买。

3) 可编程逻辑控制器 + 通信系统 + 工业级微机

与上述产品结构类型4)相类似,适用于顺序控制为主的场合。

4) 工业级微机 + 通信系统 + 工业级微机

工业级微机各有不同的功能,前者作为现场控制站,后者作为操作管理站。机型、性能等也可有所不同。

5) 智能前端 + 通信系统 + 工业级微机。

11.5.3 现场总线控制系统

这是一种简易而较通用的小型分散控制系统结构。
目前的分散控制系统也称为传统的分散控制系统,因分散控制系统的检测、变送和执行等现场仪表仍采用模拟信号连接,无法满足上位机系统对现场仪表的信息要求,阻碍了上位机系统功能的发挥。从20世纪80年代起,出现了智能化的现场仪表。这些智能化的现场仪表的功能远远超过模拟现场仪表,可对量程和零点进行远程设定,具有仪表工作状态自诊断功能,能进行多参数测量和对环境影响的自动补偿等。智能化现场仪表的出现,产生了与上位机系统实现数字通信的要求,即要求建立一个标准的现场仪表与上位机系统的数字通信链路,这条通信链路就是现场总线。现场总线是一种连接现场设备和控制室设备的全数字、双向、多点、开放式的通信系统,它使现场控制和管理的集成成为可能。

现场总线是对分散控制系统的拓展,突破了分散控制系统相对封闭的限制。分散控制系统并不是完全意义上的分散控制系统, I/O 和控制站都集中在中央控制室,现场仪表通过一对一的 4 mA ~ 20 mA 电流信号与上位机传递信息,而且各厂家的协议标准不统一,导致不同分散控制系统之间不能互操作。现场总线将测控任务分散到现场设备中,上位计算机只负责监控以及一些复杂的优化和先进控制功能。现场总线是工厂底层信息与数据传递的主体,在整个工厂的信息网络中,现场总线处于重要的基础地位。

现场总线控制系统是分散控制系统向全数字化系统发展的结果。在分级的工业网络的设备层,作用如同过程控制和制造业自动化控制的局域网,并可以内置方式将控制分布在整个局域网。它为系统的互操作性提供了基础,是系统开放的一个重要的方面。计算机与通信技术相结合产生了计算机网络;计算机网络与计算机控制系统相结合产生了现场总线控制系统。

采用现场总线控制系统取代分散控制系统具有很大的经济效益:可节约大量的通信电缆;控制功能下放到现场,使控制信号传输的准确性、实时性、快速性和可靠性大为提高;省去 I/O 端子柜和控制柜后,使控制室占地面积大大减少并使系统简化,带来了系统设计、安装、调试和维护费用的降低及工作量的大大减少;由于机组控制室(指两台机组共用 1 个控制室)只有两根同轴电缆或光缆和几根紧急停机开关的电缆进入,大大有利于控制室的防火和整个系统的安全运行。

目前,尽管存在着多种现场总线模式,但最开放、功能最完备和最具有应用前景的现场总线当属基金会现场总线(Foundation Fieldbus, FF)。

11.5.4 现场总线控制系统的特点

较之传统的分散控制系统,现场总线控制系统有以下一些突出的优点。

(1) 总线式结构

一对传输线(总线)挂接多台现场设备,双向传输多路数字信号。这种结构比一对一的单向模拟信号传送结构布线简单,安装费用低,维护简便。

(2) 开放互操作性

现场总线采用统一的协议标准,是开放式的互联网络,对用户是透明的。在传统的分散控制系统中,不同厂家的设备是不能互相访问的。而现场总线控制系统采用统一标准后,不同厂家的网络产品可以方便地接入同一网络,集成在同一控制系统中进行互操作,因此简化了系统集成。

(3) 彻底的分散控制

现场总线控制系统将控制功能下放到作为网络结点的现场智能仪表和设备中,做到彻底的分散控制,提高了系统的灵活性、自治性和安全可靠,减轻了操作控制站 CPU 的计算负担。

(4) 可靠性高

现场总线控制系统采用数字信号传输数据,提高了数据的精度和抗干扰性。将控制功能放到现场设备中,使危险分散,系统的可靠性提高。

(5) 信息综合、组态灵活

通过数字化传输现场数据,现场总线控制系统能获取现场仪表的各种状态、诊断信息,实现实时的系统监控和管理。此外,现场总线控制系统引入了功能块的概念,通过统一的组态方法,使系统组态简单灵活,不同现场设备中的功能块可以构成完整的控制回路。

11.5.5 现场总线控制网络模型

现场总线本质上是一种控制网络,控制网络直接面向生产过程,因此要求很高的实时性、可靠性、数据完整性和可用性。为满足这些特性,现场总线对标准的网络协议做了简化,一般只包括 ISO/OSI 七层模型中的三层:物理层、数据链路层和应用层。此外,现场总线还要完成与上层工厂管理信息系统的信息交换和传递。综合自动化是现代工厂自动化的发展方向,在完整的工业网架中,现场总线控制网络模型应涉及从底层现场设备网络到上层信息网的数据传输过程。基于上述考虑,统一的现场总线控制网络模型应具有三层结构,即现场智能设备层、现场总线监控层和远程监控层,如图 11-5 所示。

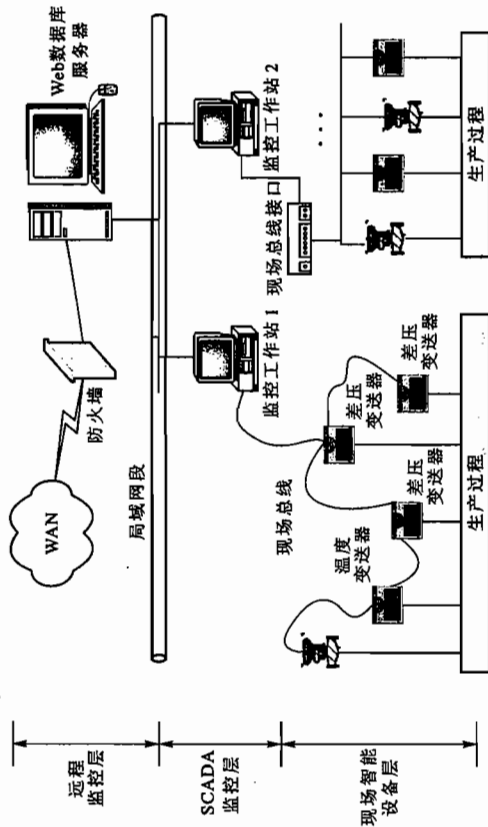


图 11-5 现场总线控制网络模型

1. 现场智能设备层

依照现场总线的协议标准,智能设备采用功能块的结构,通过组态设计,完成数据采集、A/D 转换、数字滤波、温度压力补偿、PID 控制以及阀门补偿等各种功能。智能转换器对传统检测仪表电流电压信号进行数字转换和补偿。此外,总线上应有可编程控制器接口,便于连接原有的系统。现场设备是以网络节点的形式挂接在现场总线网络上,为保证节点之间实时、可靠的数据传输,现场总线控制网络必须采用合理的拓扑结构。常见的现场总线网络拓扑结构有以下几种。

(1) 环状(令牌)网

其特点是时延确定性好,重载时网络效率高,但轻载时因等待令牌而产生不必要的时延,传输效率下降。

(2) 总线网

其特点是节点接入方便,成本低。轻载时时延小,但网络通信负荷较重时时延加大,网络效率下降,此外传输时延不确定。

(3) 树状网

其特点是可扩展性好,频带较宽,但节点间通信不便。

(4) 令牌总线网

结合环形网和总线网的优点,即物理上是总线网,逻辑上是令牌网。这样网络传输时延确定、无冲突,同时节点接入方便,可靠性好。基金会现场总线(FF)采用令牌总线网,通过链路活动调度器(Link Active Scheduler,LAS)在总线上发布令牌进行集中管理和调度,协调网络中周期性报文和非周期性报文的传输。

现场设备必须采用统一的协议标准,实现标准化,并达到不同厂家产品的完全互操作。标准化的现场设备应具备以下特征:

- 全数字双向通信
- 实施标准的功能块功能
- 多变量传输,包括状态信息和诊断信息等
- 集成自诊断、报警和趋势分析等先进功能
- 本质安全,便于现场安装

为实现现场设备的标准化,需要制定统一的协议标准和制造标准。由于当前存在多种现场总线协议标准,需要在采用不同网络协议的现场总线之间加入协议转换器(网关),识别并解释不同格式的数据包,并在不同的网络之间转发。

2. 中间层

中间层是基于数据采集与监控系统(Supervisory Control And Data Acquisition,SCADA)的现场总线监控层。这一层从现场设备中获取数据,完成各种控制、运行参数的监测、报警和趋势分析等功能,另外还包括控制组态的设计和下载。监控层的功能一般由上位计算机完成,它通过扩展槽中网络接口与现场总线相连,协调网络节点之间的数据通信,充当 LAS 的角色;或者通过专门的现场总线接口(转换器)实现现场总线网段与以太网段的连接,这种方式使系统配置更加灵活。这一层处于以太网中,因此其关键技术是以以太网与底层现场设备网络之间的接口,主要负责现场总线协议与以太网协议的转换,保证数据包的正确解释和传输。

监控层除上述功能之外,还为实现先进控制和过程操作优化提供支撑环境。

3. 上层

上层是基于 Intranet 的远程监控层。其主要目的是在分布式网络环境下,构建一个安全的远程监控系统。首先要将中间监控层实时数据库中的信息转入上层的关系数据库中,这样远程用户就能随时通过浏览器查询网络运行状态以及现场设备的工况,对生产过程进行实时的远程监控。赋予一定的权限后,还可以在线修改各种设备参数和运行参数,从而在厂域网范围内实现底层测控信息的实时传递。目前,远程监控的实现途径就是通过 Intranet,主要接入方式是租用专线或者利用公众数据网。由于涉及实际的生产过程,必须保证网络安全,可以采用的技术包括防火墙、用户身份验证以及密钥管理等。在这方面,WorldFIP 现场总线技术具有优势。WorldFIP 的报文比较灵活,兼容 TCP/IP,可以无缝连接 Intranet,同时又不影响实时数据的传送,因此整个控制网络可以采用统一的协议标准。此外,WorldFIP 的现场设备中有内嵌的 Web 服务器,因此用户可以直接通过 Intranet 访问现场设备中的信息,无需中间的协议转换器。

11.5.6 控制网络和信息网络的集成

在整个现场总线控制网络模型中,现场设备层是整个网络模型的核心,只有确保总线设备之间可靠、准确、完整的数据传输,上层网络才能获取信息以实现其监控功能。当前对现场总线的讨论多停留在底层的现场智能设备网段,但从完整的现场总线控制网络模型出发,应更多地考虑现场设备层与中间监控层、Intranet 应用层之间的数据传输与交互问题,即实现控制网络与信息网络的紧密集成。

控制网络的通信技术不同于以传输信息和资源共享为目的的信息网络,其最终目标是实现对被控对象中能量和物质转移的有效控制,使系统安全稳定地运行。因此要求具有协议简单、安全可靠、容错性好、成本低等特点。其网络负载稳定,多为短帧传送,信息交换频繁。实现控制网络与信息网络的紧密集成是建立企业综合实时信息库的基础,为企业的优化控制、调度决策提供依据;通过控制网络与信息网络的结合,可以建立统一的分布式数据库,保证所有数据的完整性和互操作性;现场设备与信息网络的实时通信,使用户能通过信息网络中标准的图形界面随时地了解生产情况;控制网络和信息网络的紧密集成也便于实现远程监控、诊断和维护功能。控制网络与信息网络的集成,可以通过以下几种方式实现。

(1) 在控制网络和信息网络之间加入转换接口

这种方式通过硬件来实现,即在底层网段与中间监控层之间加入中继器、网桥、路由器等专门的硬件设备,使控制网络作为信息网络的扩展与之紧密集成。硬件设备可以是一台专门的计算机,依靠其中运行的软件完成数据包的识别、解释和转换;对于多网段的应用,它还可以在不同网段之间存储转发数据包,起到网桥的作用。此外,硬件设备还可以是一块智能接口卡。例如,Fisher-Rosemount 的 DeltaV 系统就是通过一块机柜中的 HI 接口卡,完成现场总线智能设备与以太网中监控计算机之间的数据通信。

转换接口的集成方式功能较强,但实时性较差。信息网络一般是采用 TCP/IP 以太网,而 TCP/IP 没有考虑数据传输的实时性,当现场设备有大量信息上传或远程监控操作频繁时,转换接口都将成为实时通信的瓶颈。

(2) 在控制网络和信息网络之间采用 DDE 技术

当控制网络和信息网络之间具有中间系统或共享存储器工作站时,可以采用动态数据交换(DDE)方式实现二者的集成,其实质是各应用程序通过共享内存来交换信息。中间系统中的信息处理机是控制网络的工作站,也是信息网络中工作站。其中运行两个程序,其一是接收、校验实时信息的通信程序,为信息网络数据库提供实时数据信息;另一个是数据库访问应用程序接口,它接收 DDE 服务器实时数据并写入数据库服务器中,供信息网络实现信息处理、统计分析等功能。

DDE 方式具有较强的实时性,而且比较容易实现,可以采用标准的 Windows 技术。但是涉及复杂的协议转换时,DDE 方式的软件开销比较大。因此这种集成方式适合配置简单的小型系统。

(3) 控制网络和信息网络采用统一的协议标准

这种方式将成为控制网络和信息网络完全集成的最终解决方案。由于控制网络和信息网络采用了面向不同应用的协议标准,因此二者集成时总需要某种数据格式的转换机制,这将使系统复杂化,也不能确保数据的完整性。如果信息网络的协议标准提高其实时性,而控制网络的协议标准提高其传输速率,二者的兼容性就会提高,实现合二为一,这样从底层现场设备到远程监控系统,都可以使用统一的协议标准,不仅确保了信息准确、快速、完整的传输,还可极大地简化了系统设计。上面提到的 WorldFIP 协议就兼容 TCP/IP,因此可以方便地实现与以太网和 Internet 的集成,使控制网络和信息网络紧密地结合在一起,最终实现统一的网络结构。当前多种现场总线标准并存,信息网络协议也不尽相同,所以要实现控制网络与信息网络采用统一的协议标准,还有很多具体问题需要解决。

11.5.7 现场总线的标准

数字通信技术的发展导致了众多器件级的网络解决方案。目前,存在着多种现场总线模式,可分为传感器/执行器总线网络和基于 CAN I/O 的控制网络。这些网络能满足离散自动控制(例如计数和分类),但缺乏对于 PID 和其他过程自动控制所必需的用户层。

由现场总线基金会组织开发并管理的基金会现场总线(Foundation Fieldbus, FF)是最开放、功能最完备和最具有应用前景的现场总线,除了具备一般总线的特点外,在客户服务、功能模块调用、设备描述、诊断和保护、通信方式、数据精确传递、安全供电、设备的即插即用等方面,都体现了先进水平。

基金会现场总线技术与专用网络协议不同。基金会现场总线既不为个别公司所拥有,也不受某国或者某一管理机构控制。基金会现场总线是开放式、可操作的现场总线,建立在国际标准化组织的开放系统互连(ISO/OSI)七层通信模型的基础上,采用 OSI 参考模型的第一层、第二层和第七层规范,与国际测量和控制学会(International Society for Measurement and Control, ISA)及国际电工委员会(International Electrotechnical Commission, IEC)正式批准的 SP50 标准兼容,非常适用于对数据传送和数据处理十分敏感的应用场合,是目前国际上大多数地区和国家通用的总线标准。

习 题

1. 什么是嵌入式系统? 嵌入式系统有什么特点?
2. 嵌入式系统的有什么样的结构? 由哪些重要部分组成?
3. 嵌入式处理器大体可分为哪几类? 各有哪些特点?
4. 根据应用场合的不同, 嵌入式操作系统可以分为哪两类? 何谓实时系统?
5. 什么是 DSP?
6. 简述 DSP 的发展历程。
7. DSP 有哪些特性? 简述这些特性。
8. 什么是 SOPC 技术?
9. 什么是 IP 核?
10. 分别简述什么是软核、硬核、固核以及其各有什么特点。
11. 简述构成 SOPC 的几种方案。
12. 什么是微控制器? 它可分为哪几类?
13. 按架构而言微控制器分为哪几类? 各有什么特点?
14. 简述微控制器的基本结构和各部分的功能。
15. 比较 MCU 与 DSP 的性能。
16. 什么是工业控制系统?
17. 简述分散控制系统的优点。
18. 简述分散控制系统结构的分类。
19. 什么是现场控制总线系统?
20. 简述现场控制总线系统的特点。
21. 什么是现场总线控制网络的模型?
22. 常见的现场总线网络拓扑结构有哪几种?
23. 控制网络与信息网络的集成可以通过哪几种方式实现? 各有什么特点?
24. 简述现场总线的标准。

ADC	Analog to Digital Converter
AGP	Accelerated Graphics Port
ALU	Arithmetic Logical Unit
ANSI	American National Standards Institute
AP	Access Point
APA	All Points Addressable mode
APO	Auto Power OFF
ASCII	American Standard Code for Information Interchange
ASK	Amplitude Shift Keying
ATA	AT Attachment
ATX	AT eXtend
BCA	Ball Grid Array
BIOS	Basic Input-Output System
CAN	Controller Area Network
CAS	Column Address Strobe
CAV	Constant Angular Velocity
CCD	Charge Couple Device
CD	Compact Disc
CD-I	CD-Interactive
CDMA	Code Division Multiple Access
CD-R	CD-Recordable
CD-ROM	CD-Read Only Memory
CD-RW	CD-Rewritable
CF	Compact Flash
CGA	Color Graphics Adapter
CHS	Cylinder-Head-Sector
CIS	Contact Image Sensor
CISC	Complex Instruction Set Computer
CL	CAS Latency
CLV	Constant Linear Velocity

CMOS	Complementary Metal-Oxide Semiconductor
CPI	Cycle Per Instruction
CPU	Central Processing Unit
CR	Cyclic Redundancy Check
CRC	Cyclical Redundance Checking
CRT	Cathode Ray Tube
CSP	Chip-Scale Package
DAC	Digital to Analog Converter
DCE	Data communication Equipment
DDR SDRAM	Double Data Rate SDRAM
D-ILA	Direct-Drive Image Light Amplifier
DIME	Direct Memory Execution
DIMM	Dual-In line Memory Module
DIP	Dual In-line Package
DLP	Digital Light Processor
DLV	Digital Light Valve
DMA	Direct Memory Access
DMD	Digital Micromirror Device
DMI	Direct Media Interface
DMT	Discrete Multi Tone
DPI	Dots Per Inch
DRAM	Dynamic Random-Access Memory
DSL	Digital Subscriber Line
DSP	Digital Signal Processing
DTE	Data Terminal Equipment
DTMF	Dual Tone Multiple Frequency
DUN	Dial-Up Networking
DVD	Digital Versatile Disc
DVD + R	DVD Recordable
DVD + RW	DVD Rewritable
DVD-R	DVD Recordable
DVD-ROM	DVD Read Only Memory
DVD-RW	DVD Rewritable
DVI	Digital Visual Interface
ECC	Error Correcting Code
ECP	Enhanced Capability Port
EDO DRAM	Extended Data Out DRAM
EEPROM	Electrically Erasable Programmable Read-Only Memory

ECA	Enhanced Graphics Adapter
EIA	Electronic Industries Association
EISA	Extended ISA
EMC	ElectroMagnetic Compatibility
EMI	ElectroMagnetic Interference
EPP	Enhanced Parallel Port
EPROM	Erasable Programmable ROM
FAT	File Allocation Table
FCS	Fieldbus Control System
FDM	Frequency Division Multiplexing
FF	Foundation Fieldbus
FFT	Fast Fourier Transform Algorithm
FIFO	First In First Out
FIR	Fast Infrared
FLOPS	Floating Point Operation Per Second
FM	Frequency Modulation
FPGA	Field Programmable Gate Array
PPM DRAM	Fast Page Mode DRAM
FPU	Floating Point Unit
FRC	Frame Rate Control
FSB	Front Side Bus
FSK	Frequency Shift Keying
GPBS	General Packet Radio Service
GSM	Global System of Mobile communication
GUI	Graphical User Interface
HD Audio	High Definition Audio
HDMI	High Definition Multimedia Interface
HDSL	High bit-rate Digital Subscriber Line
HFC	Hybrid Fiber-Coaxial Network
HCA	Hercules Graphics Adapter
I2C	Inter-Integrated Circuit
ICH	I/O Controller Hub
IDE	Integrated Device Electronics
IDSL	Integrated Digital Subscriber Line
IHA	Intel Hub Architecture
IP	Internet Protocol
IrDA	Infrared Data Association
IRQ	Interrupt Request

ISA	Industry Standard Architecture	PLD	Programmable Logic Device
ISDN	Integrated Services Digital Network	PLT	Power line Telecommunications
ISO	International Organization for Standardization	PhP	Plug-and-Play
ISP	Image Signal Processor	POTS	Plain Old Telephone Service
IT	Information Technology	PPPoE	Point-to-Point over Ethernet
JPEG	Joint Photographic Expert Group	PROM	Programmable ROM
KVM	Keyboard, Video and Mouse	PSK	Phase Shift Keying
LBA	Logical Block Mode	PSTN	Public Switched Telephone Network
LCD	Liquid Crystal Display	PWM	Pulse Width Modulation
LIDE	LED InDirect Exposure	QAM	Quadrature Amplitude Modulation
LPT	Line Print Terminal	QPSK	Quadrature Phase Shift Keying
MAC	Micro Channel Architecture	RADSL	Rate Adapted Digital Subscriber Line
MCH	Memory Controller Hub	RAID	Redundant Array of Independent Disks
MDA	Monochrome Display Adapter	RAM	Random Access Memory
MGCP	Media Gateway Control Protocol	RAS	Row Address Strobe
MIDI	Musical Instrument Digital Interface	RDRAM	Rambus DRAM
MIM	Metal Insulator Metal	RISC	Reduced Instruction Set Computer
MMC	MultiMedia Cards	ROM	Read-Only Memory
MODEM	Modulation and Demodulation	RS-232C	Recommended Standard 232C
MPC	Multimedia Personal Computer	RSVP	Resource Reservation Protocol
NIC	Network Interface Card	RZI	Return to Zero, Inverted
NID	Network Interface Device	S. M. A. R. T	Self-Monitoring, Analysis and Reporting Technology
NRZI	Non Return to Zero Invert	SAS	Serial Attached SCSI
NTFS	NT File System	SATA	Serial ATA
OFDM	Orthogonal Frequency Division Multiplexing	SC	Subscriber Connector
OS	Operation System	SCADA	Supervisory Control And Data Acquisition
OSD	On Screen Display	SCSI	Small Computer Standard Interface
OSI/RM	Reference Model for Open Systems Interconnection	SD	Secure Digital Card
P&D	Plug and Display	SDRAM	Synchronous DRAM
PBX	Private Branch Exchange	SDSL	Single pair Digital Subscriber Line
PC	Personal computer	SDSL	Super high bit-rate Digital Subscriber Line
P-CAV	Partial-Constant Angular Velocity	SDSL	Symmetric Digital Subscriber Line
PCB	Printed Circuit Board	SIMM	Single-In line Memory Module
PCI	Peripheral Component Interconnect	SIP	Session Initiation Protocol
PCI-X	PCI eXtended	SIR	Serial InfraRed
PCM	Pulse Coded Modulation	SM	Smart Media
PIO	Programmed Input Output	SOC	System On Chip
PLC	Power Line Communication	SOPC	System On Programmable Chip

SPD	Serial Presence Detect
SPDIF	Sony/Philips Digital Interface
SPI	Serial Peripheral Interface
SPP	Standard Parallel Port
SRAM	Static RAM
SRC	Sample Rate Converter
STN	Super Twisted Nematic
STP	Shielded Twisted pair
SVGA	Super Video Graphics Array
TDM	Time Division Multiplexing
TFT	Thin Film Transistor
TN	Twisted Nematic
UADSL	Universal ADSL
UART	Universal Asynchronous Receiver/Transmitter
UCN	Universal Control Network
UFD	Universal Disk Format
USB	Universal Serial Bus
UTP	Unshielded Twisted pair
VCD	Video CD
VDSL	Very high bit-rate Digital Subscriber Line
VESA	Video Electronics Standards Association
VGA	Video Graphics Array
VIVO	Video In & Video Out
VoIP	Voice over IP
VPN	Virtual Private Network
WDM	Wave Division Multiplexing
Z-CLV	Zone-Constant Linear Velocity

- [1] 陈庆章. 多媒体概念漫谈[J]. 新浪潮, 1996, 12: 30-34.
- [2] 钟玉琢, 等. 多媒体计算机技术基础及应用[M]. 北京: 高等教育出版社, 1999.
- [3] 王军, 等. 多媒体技术发展现状及未来趋势浅谈[J]. 情报理论与实践, 2001, 24(2): 145-147.
- [4] 史济民, 等. 多媒体应用与开发基础[M]. 北京: 清华大学出版社, 2003.
- [5] 王新成. 多媒体技术概论[J]. 金融电子化, 2003, 10: 41-46.
- [6] Bates Tony. Teaching, Learning and the Impact of Multimedia Technologies, EDUCAUSE Review, 2000, 35(5): 38-44.
- [7] 蒋林涛. 多媒体业务的现状和发展趋势[J]. 世界电信, 2001, 8: 3-7.
- [8] 冯玉琨. 加速开拓多媒体视讯多应用与多业务市场[J]. 中国多媒体视讯, 2004, 1-2: 11-14.
- [9] 张春田, 等. 数字图像压缩编码[M]. 北京: 清华大学出版社, 2006.
- [10] 李昌立. 数字语言—语音编码实用教材[M]. 北京: 人民邮电出版社, 2004.
- [11] Gilbert Helg. 以太网[M]. 3版. 戴志清, 等译. 北京: 人民邮电出版社, 1999.
- [12] 黎连业. 计算机网络基础和网路工程[M]. 北京: 人民邮电出版社, 1998.
- [13] 胡嘉玺, 张治文, 等. 宽带接入 DIY[M]. 北京: 人民邮电出版社, 2000.
- [14] Michael Gallo. 计算机通信和网路技术[M]. 王玉峰, 等译. 北京: 人民邮电出版社, 2003.
- [15] 张登银, 等. VoIP 技术分析与系统设计[M]. 北京: 人民邮电出版社, 2003.
- [16] 朱卫东, 等. 计算机硬件技术基础[M]. 北京: 高等教育出版社, 2001.
- [17] 章振业, 等. 计算机外部设备[M]. 北京: 电子工业出版社, 2000.
- [18] Scott Mueller. PC 硬件工程师手册[M]. 吕俊辉, 等译. 北京: 机械出版社, 2002.
- [19] 渠云田. 电工电子技术——数字与电气控制技术基础(下册)[M]. 北京: 高等教育出版社, 2003.
- [20] 郑纬民 汤志忠. 计算机系统结构[M]. 2版. 北京: 清华大学出版社, 2003.
- [21] 冯博琴, 吴宁. 微型计算机硬件技术基础[M]. 北京: 高等教育出版社, 2003.
- [22] 蒋本珊. 电子计算机组成原理[M]. 3版. 北京: 北京理工大学出版社, 2004.
- [23] 黄洁副. 数字电子技术基础[M]. 武汉: 华中科技大学出版社, 2004.
- [24] 吉本久泰. 图说数字技术[M]. 丁志俊, 等译. 北京: 科学出版社, 2003.
- [25] 赵效民. 高手进阶, 终极内存技术指南——完整进阶版[J/OL]. 存储时代[2004-2-18]. <http://www.stor-age.com>.

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其为人将承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人给予严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话：(010)58581897/58581896/58581879

传 真：(010)82086060

E-mail: dd@hep.com.cn

通信地址：北京市西城区德外大街4号

高等教育出版社打击盗版办公室

邮 编：100011

购书请拨打电话：(010)58581118