

1.1 半导体集成电路技术的进展

自 1946 年世界第一台通用电子数字计算机 ENIAC 诞生以来,计算机已走过电子管时代、晶体管时代,并在 1965 年开始进入集成电路(integrated circuit, IC)时代。现在,不仅仅计算机产业以 IC 为核心,而是整个 IT 业都是以 IC 为核心。下面简要回顾一下 40 多年以来 IC 技术前进的步伐,并了解当前 IC 技术的水平。

1.1.1 IC 技术发展的回顾

1957 年,杰克·基比尔(Jack S. Kilby)在一个硅片上做出四个晶体管,由此发明了集成电路。1958 年美国 TI 公司生产出世界上第一块 IC 产品,由此开始,IC 技术的发展已经历了 4 代,目前仍处于第 5 代——巨大规模集成电路(GLSI)时代。

1. SSI/MSI 时代(1958 年—1965 年)

SSI(small scale integration)为小规模集成,MSI(medium scale integration)为中规模集成。这个时期,集成度小于 10^3 ,集成的晶体管是 TTL 型,芯片的面积小于 10mm^2 。

2. LSI 时代(1966 年—1977 年)

LSI(large scale integration)为大规模集成。这个时期的集成度为 $10^3 \sim 10^5$,集成的晶体管是 MOS 型,推出一批 8 位微处理器产品,如 6800、8080、Z80 等,生产工艺为 $6\mu\text{m}$ 制程。1977 年已推出集成度超过 10 万晶体管的 64K 位 DRAM 和 16K 位 SRAM,以及一些 16 位微处理器产品。

3. VLSI 时代(1978 年—1985 年)

VLSI(very large scale integration)为超大规模集成。这个时期的集成度为 $10^5 \sim 10^7$,集成的晶体管是 CMOS 型。CMOS(complimentary metal oxide semiconductor)为互补式

金属氧化物半导体,是指由 NMOS 和 PMOS 两个逻辑互补器件组成一个 CMOS 反相器,其主要优点是:不管输入端是高电平还是低电平,只有一个 MOS 管处于导通状态;仅当开关瞬变过程才耗费少量电能,从而降低了功耗;同时还可降低栅极延迟时间,从而提高芯片运行速度。包括后来两代都是采用 CMOS 型或在其基础上的改良型工艺。当今 CPU、GPU、南北桥芯片和内存芯片也都采用这种工艺。

这个时期,Intel 公司相继推出 8086、80186、80286 和 80386 微处理器。下面我们用芯片容量每制程来表述 DRAM 芯片技术的进步:1978 年,64K 位 DRAM/ $3\mu\text{m}$;1980 年,256K 位 DRAM/ $2\mu\text{m}$;1985 年 1M 位 DRAM/ $1.2\mu\text{m}$ 。

4. ULSI 时代(1986 年—1994 年)

ULSI(ultra large scale integration)为特大规模集成。这个时期的集成度为 $10^7 \sim 10^9$ 。这个时期,Intel 公司相继推出 80486、Pentium、Pentium Pro 等 32 位微处理器。仍以 DRAM 技术的进步来表征这个时期的 IC 技术水平:1986 年,4M 位 DRAM/ $0.8\mu\text{m}$;1988 年,16M 位 DRAM/ $0.5\mu\text{m}$;1990 年,64M 位 DRAM/ $0.35\mu\text{m}$;1992 年,256M 位 DRAM/ $0.25\mu\text{m}$ 。这个时期的硅圆片直径已达到 200mm,并开始使用 300mm 的硅圆片。IC 产品率大幅上升,1994 年世界 IC 产品销售额已突破 1 000 亿美元大关。

5. GLSI 时代(1995 年至今)

GLSI(gigantic large scale integration)为巨大规模集成。这个时期的集成度大于 10^9 ,硅圆片的直径由 300mm 进展到 400mm。这个时期,Intel 公司相继推出 Pentium II/III 微处理器、三种核心架构的 Pentium 4 微处理器和 Itanium 64 位微处理器。1995 年,1G 位 DRAM 问世,但直到 2000 年才批量生产出 1G 位 DRAM/ $0.18\mu\text{m}$ 内存芯片,采用的是 300mm 硅圆片。2004 年,采用 400mm 硅圆片的 4G 位 DRAM/ $0.13\mu\text{m}$ 芯片成批生产。

集成度的更准确表述应是特征尺寸(feature size),即 IC 器件的电路线宽,因为特征尺寸越小,单位面积内的晶体管集成度就越高。上面所说的 $0.18\mu\text{m}$ 制程、 $0.13\mu\text{m}$ 制程等指的就是特征尺寸。制程和特征尺寸都指的是线宽,表述 IC 工艺技术时常用制程(process)一词,表述 IC 器件性能规格时爱用特征尺寸一词,当然也不是那么决然区分。

说到 IC 技术的进展历程,就必然说到摩尔定律。虽然摩尔定律被称作定律,但它并不是自然定律或法则。它源于 1965 年 4 月戈登·摩尔博士(Intel 公司创始人之一)为庆祝《Electronic》杂志创刊 35 周年所撰写的一篇文章。在这篇文章里,戈登·摩尔推测在未来 10 年中,芯片上晶体管数量每 18 个月都会增长一倍。

从 1978 年进入 VLSI 时代以来,IC 技术的发展一直沿着两个方向进行:一是不断扩大晶圆尺寸,从 $100\text{mm} \rightarrow 125\text{mm} \rightarrow 150\text{mm} \rightarrow 200\text{mm} \rightarrow 300\text{mm}$,直到目前的 400mm ,以提高芯片产量和降低芯片成本;二是不断缩小芯片特征尺寸,从 $3\mu\text{m} \rightarrow 2\mu\text{m} \rightarrow 1.2\mu\text{m} \rightarrow 0.8\mu\text{m} \rightarrow 0.5\mu\text{m} \rightarrow 0.35\mu\text{m} \rightarrow 0.25\mu\text{m} \rightarrow 0.18\mu\text{m} \rightarrow 0.13\mu\text{m} \rightarrow 90\text{nm}$,直到目前的 65nm ,以满足芯片的高集成度化、高速化和高可靠性化需求。近 30 年 IC 技术的发展相当精确地验证摩尔定律的正确。

1.1.2 65 纳米制程 IC 技术

当今,IC 工艺技术已进展到 65 纳米(nm)制程,这是 IC 工艺技术的一道关坎,因为芯片制造工艺、热能及粒子活动已达到物理极限,除非采用更先进的半导体材料、更先进的芯片制作工艺、更先进的晶体管结构和更先进的封装技术,否则摩尔定律会在若干年后失效。

可喜的是,Intel 在 2003 年推出了以 90nm 制程 Prescott 为核心的 Pentium 4 处理器,并在 2003 年 11 月 24 日在北京宣布基于 65nm 技术的全功能 SRAM 芯片已经成功生产的消息。那么,基于 65nm 技术的微处理器芯片问世也就是 2005 年底、2006 年初的事了。

Intel 的 65nm 制程的代号为 P1264,除光刻工艺技术为 65nm 外,还规定栅极长度为 35nm 和使用直径为 300mm 的晶圆片。下面,我们来认识 Intel 65nm 制程 IC 技术的一些主要特点。

1. 采用第二代拉伸硅半导体材料

单晶硅从来都是最重要的半导体材料。但是随着制程的减小和速度的提高,硅自身的一些特征成为进一步提高芯片运行速度的障碍。例如非单一同位素硅(同时含有 Si28、Si29、Si30)在某些情况下,一些硅原子的间距会无规则而降低晶体的完整性,这就会降低导电性,并造成载流子向其他方向散射,是信号不稳定和电能泄漏的重要原因。

拉伸硅(strained silicon),又称应变硅技术有效地解决了这一问题。其基本原理是,从外部施加影响使硅晶体拉伸,迫使硅原子的间距拉大,间距越大,排列就越规整,就可减小载流子通行所受到的阻力,从而可提升沿拉伸方向移动的载流子速率。Intel 将 MOS 晶体管栅极下的沟道做成拉伸硅,这样当打开晶体管时,电流会顺畅地沿拉伸方向由源极流向漏极,速度也能更快,开关反应时间也缩短了。同时,向衬底分散的漏电流也会相应减小,功耗和发热量都会降低。

当然,这种拉伸硅不是靠机械力量实现的,而是靠掺杂(如掺杂 SiGe、Si₃N₄ 等)来实现的。另外,需要说明的是,高性能微处理器都使用的是 CMOS 晶体管,同时包括 NMOS 和 PMOS 管,而 NMOS 只是在张应变、PMOS 只在压应变情况下性能才会得到提高。一种称为双应力衬底(dual stress liner,DSL)的技术,用于在衬底上实施两个应变,分别作用在两种不同类型的晶体管,最终得到 CMOS 晶体管的性能提高。

2. 采用 193nm 波长的光刻技术

目前的光刻技术使用的是深紫外光(deep ultra violet,DUV),波长一般是 248nm、193nm 和 157nm。波长为 248nm 的光刻设备为 Intel 服务了近 10 年,0.25 μ m、0.18 μ m 和 0.13 μ m 制程的芯片都是这类设备生产的。2001 年 Intel 配备了 193nm 光刻设备,生产出 90nm 制程的芯片。157nm 光刻设备还没有完全开发完毕,而且售价高达每套 4 000~5 000 万美元之间。出于成本的考虑,Intel 决定仍使用 193nm 光刻设备来生产 65nm 制程的芯片。

成像图形特征尺寸是曝光波长的 2 倍时很容易实现,特征尺寸与曝光波长相近时就有困难了,现在 90nm、65nm 的特征尺寸是曝光波长(193nm)的 1/2 甚至 1/3,更是难上加难了,它的衍射现象会十分严重,图案边缘形状会严重失真。为克服这种困难,Intel 应用了相移掩模(phase shift mask,PSM)技术和光学邻近校正(optical proximity correction,OPC)技术。

PSM 技术主要是在铬涂层上加入了转换材料,使入射光产生 180 度的相位差,从而降低衍射所引起的干涉效应。OPC 技术是一种针对掩模进行校正的技术,它可以增强比曝光波长更细的特征线的光刻,校正图形的变形与失真。

Intel 正在积极开发极紫外光(extreme ultra violet,EUV)光刻技术,波长为 13nm。Intel 表示有望在未来的 32nm 制程 IC 芯片制造中,全面使用 EUV 光刻技术。

3. 高 K 介质的栅极,长度仅为 35nm

Intel 65nm 制程采用栅极长度仅为 35nm 的晶体管,栅极的厚度仅为 1.2nm。比之 90nm 制程的栅极长度为 50nm,仅此栅极变短这一项,就可使栅极电容降低 15%~20%。

IC 芯片的功耗与 cfv^2 成正比,其中 c 是芯片电容(包括各种寄生电容), f 是芯片工作频率, v 是芯片供电电压。因此,更短的栅极使栅极电容减小,会使芯片功耗减少。

因芯片功耗与供电电压的平方成正比,故降低供电电压是减少功耗的最有效的手段。近年来芯片供电电压逐年降低,已从 5V 降至 1V 左右。但是,电压过低芯片就无法正常工作。目前的技术措施是,通过使用高 K 值介电材料来制作栅极,从而降低阈值电压,以允许较低的供电电压。AMD 使用的是 SiON 和 Si_3N_4 两种材料,而 Intel 使用的是 ZrO_2 。65nm 制程通过使用高 K 介质材料做栅极,并使栅极变短至 35nm,这两项技术不仅降低了功耗,也使开关转换速度提高。

4. 低 K 介质的绝缘层,铜互连,8 层布线

IC 芯片中的层间及线间的绝缘材料都是 SiO_2 ,随着芯片中导线密度不断增加,导线宽度和间距的不断减小,互联中的电阻电容所产生的寄生效应越来越明显。克服阻容效应所引起的信号传播、线间串扰以及功耗等已成为重要问题。用低 K 介质材料做衬底和填充于芯片的层间和线间,不仅可以减少漏电流降低芯片功耗,还能降低线间寄生电容,有效抑制线间串扰提高工作的稳定性。

铜的导电性能(电阻率为 10.7)要优于铝(电阻率为 17)。但由于工艺技术上的原因及成本的考虑,以前各种制程的 IC 芯片内部都是采用铝互连。到了 65nm 制程后,线宽变得这样窄,铝线的电阻就不能忽视了,而必须采用铜互连了。

在保证面积少量增长的前提下,大量增加芯片内晶体管数量的有效办法是增加布线层数。90nm 制程的 Prescott 核心 Pentium 4 处理器采用了 7 层布线。Intel 的 65nm IC 技术采用了 8 层布线。

总之,Intel 采用了上述的材料工艺、光刻技术、栅极结构、高绝缘材料填充、铜互连和 8 层布线后,不仅可保证特征尺寸为 65nm,而且为降低功耗和高频下的工作稳定性做了不少重大改进。65nm 制程的微处理器产品即将问世,摩尔定律继续有效。

1.2 计算机体系结构的进展

上一节对作为计算机基石的 IC 技术予以简要回顾,并介绍了当前的 IC 技术水平。本节对计算机体系结构的进展与分类进行简要介绍,以使读者了解奔腾 PC 体系结构的类型及其发展方向。

1.2.1 冯·诺依曼机体系结构及其发展方向

世界第一台通用电子数字计算机 (electronic numerical integrator and computer, ENIAC) 于 1946 年在美国宾夕法尼亚大学建成。ENIAC 在输入和更换程序方面特别繁冗。对此,ENIAC 课题组的顾问著名数学家冯·诺依曼提出将程序的指令与指令所操作的数据一起存于存储器的概念。这个著名的存储式程序 (stored-program) 概念,成为计算机工作的基本机理。这一概念也被图灵大约在同时期提出。

1. 冯·诺依曼机器

50 多年来,计算机体系结构已取得重大进展,但大多数机器的计算机体系结构仍没能摆脱冯·诺依曼机器结构的范畴。冯·诺依曼型机器由运算器、控制器、存储器和输入输出设备组成,如图 1.1 所示。它的计算机体系结构的基本特点可归纳为以下几点。

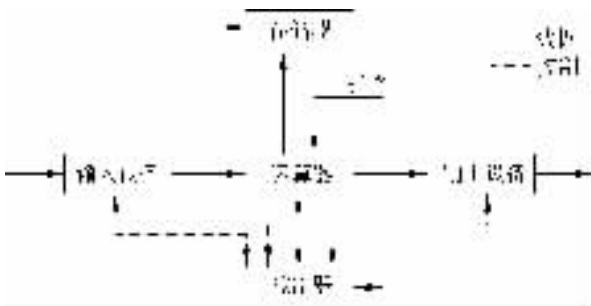


图 1.1 冯·诺依曼机器的结构

第一,采用存储程序方式,程序的指令和数据一起存放到存储器。存储器由线性编址的单元组成,每个单元的位数是固定的。

第二,指令由操作码和地址码组成,操作码指定操作类型,地址码指明操作数地址。操作数的数据类型由操作码确定,操作数本身不具有数据类型标志(如是定点数,还是浮点数等)。

第三,存储器中的指令和数据,从它们本身是区别不了的,它们都是以二进制编码表示的,只不过指令的地址应由指令计数器给出。换句话说,以指令计数器的值为地址,由存储器读出的内容将被机器看成是指令,被送往控制器去解释和执行。

第四,每执行完一条指令,指令计数器一般自动加“1”,以指示下一顺序指令的地址。虽然执行顺序可以根据运算结果来改变,但解题算法依然是(也只能是)顺序型的。

第五,机器以运算器为中心,采用二进制运算,输入输出设备与存储器之间的数据传送都经过运算器。控制器实施对各部件的集中控制。

总之,冯·诺依曼机器实现了存储式程序这一基本原理;但指令的执行次序受指令计数器(也称为程序计数器)的控制,是一种控制驱动方式的机器。尤其是它的串行执行方式,使得解题算法和编程模型只能是顺序型的;它的更深远影响是使目前绝大多数程序设计语言、编译器、操作系统都是面向顺序式编程模型。

2. 计算机体系结构的发展方向

50 年来,电子计算机已经历了四次更新换代,现在正处于第五代。各代的划分主要依据半导体技术水平,并有明显的硬件和软件技术为标志,如表 1.1 所示。

表 1.1 计算机的发展

	技术和系统结构	软件和应用
第一代 (1945 年—1954 年)	电子管和继电器。单 CPU,以程序计数器(PC)和累加器顺序完成定点运算	机器语言或汇编语言,单用户,用 CPU 程序控制 I/O
第二代 (1955 年—1964 年)	晶体管和磁芯存储器。用印刷电路互连。变址寄存器,浮点运算;多路存储器,I/O 处理机	有编译程序支持的高级语言,子程序库,批处理监控程序
第三代 (1965 年—1974 年)	中小规模集成电路(MSI/SSI)。多层印刷电路。微程序设计,流水线,高速缓存,先行处理机	多道程序设计,分时操作系统,多用户应用
第四代 (1975 年—1990 年)	大规模集成电路(LSI/VLSI)。半导体存储器。多处理机,多计算机,向量超级计算机	用于并行处理的多处理机操作系统、专用语言和编译器;并行处理或分布计算的软件工具和环境
第五代 (1991 至今)	特大规模集成电路(ULSI/GSI)。高密度高速度处理器和存储器芯片,可扩展系统结构,因特网	大规模并行处理,Java 语言,分布式操作系统,万维网

当前,计算机体系结构正在沿着两个方向发展。

一是改变冯·诺依曼机器的串行执行模式。如在处理器中加入更多的译码部件、更多的执行部件和更多的寄存器而实现的超标量处理器,以做到同时执行多条指令。又如,以多个处理器共享集中式存储器或分布式存储器的多处理机系统,以及大规模并行处理机(MPP)系统,它们在系统中注入强大的并行处理能力并提出了并行编程模型。

在个人计算机与工作站迅速发展和计算机网络技术成熟和普及的情况下,近年来成为研究热点的机群系统不仅具有良好的并行性,而且具有良好的可扩展性和可编程性以及良好的市场前景。

二是改变冯·诺依曼机器的控制驱动方式。自 20 世纪 70 年代以来,提出了数据驱动、需求驱动和模式匹配驱动三种新型驱动方式。数据流计算机是一种数据驱动式系统结构的计算机,只要指令所需的操作数已经齐备,就可立即启动执行,一条指令的运算结果又流向下一条指令,作为下一条指令的操作数来驱动该条指令的启动执行。程序中各条指令的执

行顺序仅由指令间的数据依赖关系决定。需求驱动方式是一个操作,仅在要用到其输出结果时才开始启动,如果此操作的操作数未到齐,则它去启动能得到各输入数的操作,需求链一直延伸下去,直到遇到常数或外部已输入的数据为止,然后再反方向地去执行运算。归约机就属于需求驱动系统结构的、使用函数式程序设计语言的计算机。在模式匹配驱动方式中,计算的进行是由谓词模式匹配来驱动的,而谓词是代表客体之间关系的一种字符串模式,主要用来求解非数值的符号演算。面向智能的计算机,如 Lisp 机、Prolog 机、神经网络等,属于这种模式匹配驱动式系统结构的计算机。

近 30 年来,前一种发展方向,即控制驱动方式下并行处理体系结构的计算机,已取得重大进展。不论是硬件技术还是软件技术都已相当成熟,并有不少产品走向市场,它们代表当前计算机体系结构发展的主流。而后一种发展方向,除数据流计算机已有一些成型机之外,大多还属于探索、研究阶段,还有大量的工作有待完成。

1.2.2 计算机体系结构的 Flynn 分类法

计算机系统曾普遍被分为巨、大、中、小、微五类,这是按规模、性能、速度以至价格的一种大致划分。

大、中型计算机,国外常用 mainframe 一词表示,它们大都是通用机,在计算机工业中占很重要地位,很多新的系统结构技术都是首先在大中型计算机上被采用。著名的产品,早期有 IBM 360 系列、370 系列,后来有 IBM 4300 系列等。

巨型计算机是为气象预报、国防工程、核物理研究等专门设计的具有极高运算速度和很大容量的计算机。其中,著名的有 Cray-1 计算机,它的向量运算速度达 8 000 万次/秒,并兼顾了一般的标量运算。1983 年研制成功的 CrayX-MP 机向量运算速度达 4 亿次/秒。近年来,以微处理器为阵列结构的巨型机(常称为小巨型机)得到了发展,例如古德伊尔公司为美国宇航局研制的巨型计算机系统 MPP,由 16 384 个微处理器组成 128×128 方阵。我国自行研制的银河-I、银河-II 型机也属于巨型机范畴。

因小型机规模小、结构简单,设计试制周期短、便于及时采用先进技术和工艺,硬、软件成本低等优点,曾在 20 世纪 70 年代被普遍采用,著名的有 DEC 公司的 PDP-11 系列和 VAX-11 系列,以及近年来曾在我国高校、银行、企事业单位中广泛流行的、用于 Client/Server 计算的 AS/400 亦属于小型机。

微型机的出现与发展,引发了世界范围的计算机大普及浪潮。1971 年以 Intel 4004 的 4 位微处理器组成的 MCS-4 是世界第一台微型机。30 多年来,微型机获得惊人的飞跃式发展,从 4 位、8 位、16 位到现在的 32 位机,目前正在向 64 位计算机发展。32 位微型机已采用过去大中型计算机中所采用的技术,故现在的微型机性能已达到 20 世纪 70 年代大中型计算机的水平。

可见,这种按巨、大、中、小、微五类来划分计算机系统只能对同时期的计算机大致分类,规模、价格、尤其是性能和速度的指标随时间变化而变化。而且,这种分类法也不能反映计算机的体系结构特征。

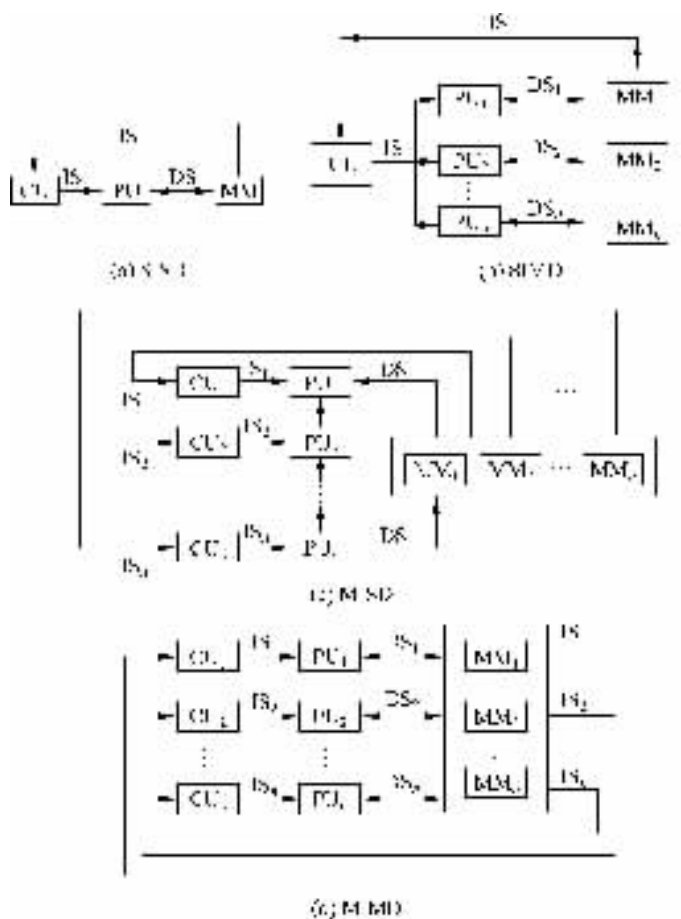
1972年, Michael J. Flynn 提出按指令流和数据流的多倍性对计算机体系结构进行分类。由于当前的计算机体系结构的主流发展方向是控制驱动方式下的并行处理, 故这一分类法获得普遍赞同。Flynn 分类法提出如下定义。

- 指令流(instruction stream): 机器执行的指令序列。
- 数据流(data stream): 指令流使用的数据序列, 包括输入数据和中间结果。
- 多倍性(multiplicity): 在系统最受限制的部件上, 同时处于同一执行阶段的指令或数据的最大数目。

按指令流和数据流的多倍性, Flynn 将计算机体系结构分成如下四类:

- 单指令流单数据流(SISD)。
- 单指令流多数据流(SIMD)。
- 多指令流单数据流(MISD)。
- 多指令流多数据流(MIMD)。

图 1.2 分别表示了它们的基本结构(不包括 I/O 设备)。下面再对每类予以简要解释。



CU=控制部件; PU=处理部件; MM=存储器模块; IS=指令流; DS=数据流

图 1.2 Flynn 分类法各类机器结构

1. SISD 体系结构

这种结构代表了传统的冯·诺依曼机器,即大多数的单机(单处理器)系统。处理器串行执行指令;或者处理器内采用指令流水线,以时间重叠技术实现了一定程度的指令并行执行;或者处理器是超标量处理器,内有几条指令流水线实现了更大程度上的指令并行执行。但它们都是以单一的指令流由存储器取指令,以单一的数据流由存储器取操作数和将结果写回存储器。

2. SIMD 体系结构

这种体系结构有单一的控制部件(CU),但有多个处理部件(PU)。以 CU 由存储器取单一的指令流,一条指令同时作用到各个 PU,指挥各 PU,对来自不同数据流的数据组进行操作。这种系统结构的典型代表是阵列处理机,有的学者认为应将向量处理机划入此类,但也有人持否定态度。值得一提的是,当前的处理器(如 Pentium II/III 和 Pentium 4)已具有多媒体指令功能,这类指令能对打包数据中的多个数据元素同时进行操作。这是一种 SIMD 的变异,也常简称为 SIMD 类的指令。

3. MISD 体系结构

这种体系结构中,有几个处理部件 PU,各配有相应的控制部件 CU。各个 PU 接收不同的指令,对来自同一数据流及其派生数据流(例如中间结果)进行操作。这已被证明是不可能的或不实际的,至少是目前没有这类机器。

4. MIMD 体系结构

这种体系结构中,也有几个 PU 并各配有相应的 CU,但各 PU 接收不同的指令,对不同的数据流进行操作,这是可行的。把一个 PU 及其相应的 CU 看成是一个处理器(CPU),则这是一个多处理器系统。这类系统的典型代表是多处理机和多计算机。如果提供 n 个指令流和 n 个数据流的各存储器模块来自(集中或分布的)共享存储器,则是多处理机系统。如果各存储器模块各有自己独立的地址空间,那么各处理器为协同操作所需进行的相互通信就只能依据消息传递来进行,而不能依赖共享变量来进行,则是多计算机系统。

总之,Flynn 分类法能反映大多数计算机系统的并行性、工作方式和结构特点,但这仅是对控制驱动方式下的串行处理、并行处理计算机进行分类。对非控制驱动方式的计算机,如数据流计算机,就不能包括进去。此外,还有其他一些分类法,如美籍华人冯泽云(Tse-yun Feng)于1972年提出的按最大并行度分类法;又如 Wolfgang Händler 于1977年提出的根据并行度和流水线进行分类。这里就不予具体介绍了。

1.2.3 并行计算机系统分类

1989年,Almasi 和 Gottlieb 曾给并行计算机下了一个定义:“并行计算机是相互通信和协作的多个处理部件的集合,用来快速求解大型问题。”依此定义,SIMD 的阵列处理机、MIMD 的多处理机和多计算机均被看成是并行计算机。下面对这三类系统的系统结构特

征予以说明。

1. 阵列处理机

阵列处理机(array processor)面向高速数值计算,特别适合求解有限差分、矩阵、线性规划等计算问题。这些问题的共同特点是,以各种途径最终可归结为数组和向量处理。流水线向量处理机也适合于这类计算问题,但在实现并行性所采用的主要技术上有所不同:向量处理机采用的是时间重叠技术;阵列处理机采用的是资源重复技术。阵列处理机重复设置多个同样的处理单元(process element, PE),按照一定的方式相互连接,在统一的控制部件(control unit, CU)作用下,各自对所分配的数据并行地完成同一指令所规定的操作。

依据存储器的不同组成形式,阵列处理机又可分为分布存储器式和共享存储器式两类。下面仅介绍共享存储器式阵列处理机。

这种阵列机是若干个存储模块构成一个统一的存储器,通过互连网络(inter-connection network, ICN)为所有处理单元共享,如图 1.3 所示。共享存储模块的数目等于或略大于处理单元数目。通过在共享存储模块之间合理分配数据,使各 PE 所需数据能来自不同的模块,将有利于克服存储器冲突。ICN 能提供多种连接方式,使每个 PE 能与不同的存储模块相连。

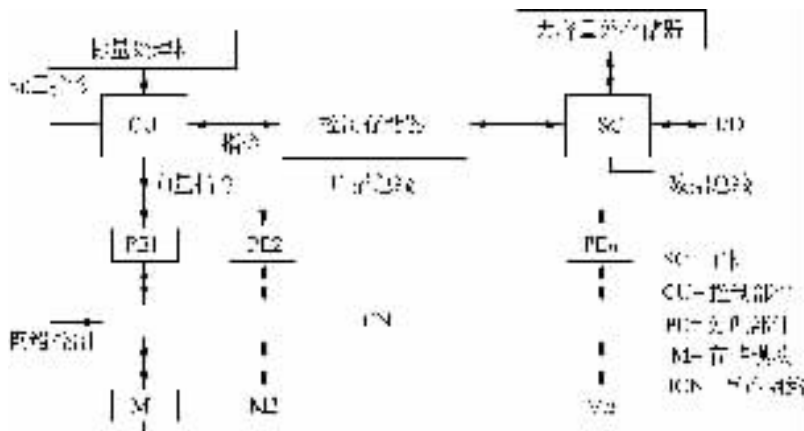


图 1.3 共享存储器阵列处理机

美国宝来公司和伊利诺伊大学于 1979 年制造的 BSP 计算机是共享存储器阵列处理机的代表。它有 16 个算术单元 AE,能对不同的数据组执行同一指令的操作,17 个并行存储模块,每个模块有 512K 个字,组成了一个无冲突访问存储器。

阵列处理机的系统结构比较固定,直接与一定的算法相联系,其效率取决于计算程序的向量化程度。这种结构比较固定的专用机,其可扩展性不好。

2. 多处理机

在 Flynn 分类法中, MIMD 类可分成共享存储器的多处理机(shared-memory multiprocessors)和消息传递的多计算机(message-passing multicomputers)。下面先介绍多处理机。