

并行计算机

体系结构 程序设计及算法

[英] R.W. 霍克尼 C.R. 杰斯荷普 著

夏绍琛 栾毓敏 译 李润斋 林定基 校

清华大学出版社



并行计算机

体系结构 程序设计及算法

[英] R.W.霍克尼 著
[英] C.R.杰斯荷普

夏绍璠 栾毓敏 译

李润斋 林定基 校

清华大学出版社

内 容 简 介

本书主要介绍新型的计算机体系结构、并行语言和并行算法，最后展望了超级计算机的发展前景。
书中第一章阐明了计算机的体系结构的概念、定义、组成、分类和评价。
第二章、第三章分析了流水线计算机和处理机阵列这两类并行计算机结构。
第四章引出了并行性的表达式和并行语言并说明并行计算机的有效性依赖于选择合适的并行算法。
第五章提出了评价并行算法性能的某些原则并用到一些具体问题的算法中。
第六章论述了并行计算机的一些限制并展望了超级计算机的发展趋势。
本书可供从事计算机设计和研制的科技人员以及高等学校有关专业的师生阅读和参考。

PARALLEL COMPUTERS ARCHITECTURE, PROGRAMMING AND ALGORITHMS

R. W. Hockney
C. R. Jesshope 著

Adam Hilger Ltd, Bristol 1981年第1版

并行计算机

体系结构 程序设计及算法

[英] R. W. 霍克尼 著
[英] C. R. 杰斯荷普

夏绍璠 栾毓敏 译
李润斋 林定基 校

☆

清华大学出版社出版
北京 清华园

河北省蔚县印刷厂印刷
新华书店北京发行所发行

☆

开本：787×1092 1/16 印张：17³/₄ 字数：477千字

1987年7月第一版 1987年7月第1次印刷

印数：00001~6000

统一书号：15235·300 定价：2.95元

译 者 的 话

信息时代已经到来。作为信息处理的主要工具——计算机已经广泛地应用到各个领域。随着信息量的飞速高涨，计算机的计算能力也相应地提高，目前国外计算机的计算能力几乎每五年就要增加十倍。我国在本世纪末要实现四个现代化，这不仅要求在计算机的数量上有较大的发展，而且在质量上，在计算能力上也必须有显著的进步。不久前银河机和757大型向量机的研制成功，标志着我国计算机系统已进入到并行计算机的先进行列，但离开世界先进水平差距仍然很大。因此，我们要努力研究和学习外国的经验。本书全面、系统地介绍了国外各种并行计算机系统，是一本很好的学习资料。

本书共分六章。前三章着重介绍并行机的种类、性能。并以实例，如CRAY-1、CYBER-205、FPS AP-120B、ICL-DAP及BSP等，这些都是典型的并行流水线计算机和大型的阵列处理机。其中尤为重要的是引出一个半性能长度 $n_{1/2}$ ，它是测量并行度的一个重要参量。书中并给出了设计机器时应考虑的问题。

后三章主要阐明了并行语言和并行算法。以并行结构的计算机为例，根据各种题目所提出的要求，分析了并行算法的效率。讨论了用并行语言如何提高机器解题能力和大大缩短机时等一系列的问题。此外，通过典型的例子，说明如何同时考虑并行语言和并行算法二个因素，来设计出一个好的计算机体系。本书用FORTRAN语言对一阶和二阶线性递归方程、矩阵乘法、三对角线系统、傅里叶变换及偏微分方程等作了一些实际研究。这些实例都是在上述所提到的各类并行计算机上计算过的。并给出明确的结果，以说明并行语言和并行算法的意义及用途。本书最后一章展望未来超级计算机的发展前景。

本书非常适合于从事计算机科研工作与教学工作的同志学习和研究，对于高年级本专业的大学生及本专业的研究生都是一部不可缺少的教材。本书不仅对从事大型计算机工作的同志非常有用，而且对从事微型计算机工作的同志也是相当有用的。因为大型阵列处理机ICL DAP就是由很多小的处理机组成的。今天，微型机和网络有密切的关系，特别是计算机网络的发展，给并行处理提供了很强的功能。此外，也可以把流水线放在微型机的基本系统内，如书中提到的FPS-164，FPS AP-120B等。在今后的十多年我国的电子工业技术必然会有很大的发展，无疑半导体工艺也会有新的突破。与此同时，一定会有新的计算机体系结构相应地出现。因此，本书会对我们起特殊的指导作用。

本书前三章由夏绍璠译，李润斋校，后三章由栾毓敏译，林定基校。

前 言

这本书是讨论计算机科学若干特殊问题的新系列专著中的第一本专著。这些特殊问题是指对决定计算机未来的形式已经具有或正在具有重要意义的那些问题。过去20年内发生了迅速的变化，至今没有任何减弱的迹象。的确，所有迹象表明，在今后十年内将会有更大的变化。尤其是计算机的体系结构，由于半导体工艺出现了先进技术，以及综合地使用了计算机和通讯技术，使得有可能引入新的形式。因此，对于从事实际工作的科学工作者和工程师来讲，了解超大型计算机的体系结构、并行算法、局部网络及安全保密等方面的发展情况，现在变得特别重要。专著的目的在于对这些课题以及其他一些课题提出最新解释和叙述。

第一本专著选择讨论并行处理，这是特别合适的。已有几台向量机和阵列处理机成为商品出售，并且在过去几年内，在设计、研制、并行算法等几方面又继续进行了努力。在很广的应用领域内已采用了并行处理机，例如，在气象学、蛋白质结晶学、动态程序和优化、信号处理和分子电子勘测等这样一些完全不同的课题中得到广泛的应用。计算机网络的发展，提供了并行处理能力。这就为越来越多的科研人员 and 工程师提供了并行处理手段。对于这些科研人员、工程师以及大学三年级的学生来说，本书的目的在于给出关于这一课题的一般评述，并且提出实现这些有效的技术的可行的方法。

序

本世纪八十年代多半是属于并行计算机的十年。本书的目的就是要介绍这个题目。虽然从五十年代起，很多计算机已显示了并行或者并发操作的功能，但是，只是到了1974—1975年才出现有效地进行向量和数组运算的专用并行计算机。设计这类计算机是建立在这样的基础之上的，它或者是并行执行运算操作的不同子功能，这与工厂装配线上的工作方式相同（如CDC STAR和TIASC流水线计算机），或者是复制多台完整的运算器（处理机阵列如ILLIAC IV）。在早期的设计中遇到很多问题，但是到了1980年，与早期的研究计划相反，几家主要厂家提供了能够大量生产的并行计算机。主要的例子有CRAY-1（实际上最初安装是在1976年）和CDC CYBER205流水线计算机，以及ICL DAP和Burroughs BSP处理机阵列。很遗憾，从我们写这本书开始，Burroughs公司就对BSP生产中的问题做了实验，之后安装了一台样机，但后来还是取消了这个课题。然而，这仍然是一项十分有意义的设计，工作虽然放弃了，但从中增加了很多这方面的见识。现在流水线设计已经象处理机一样愈来愈普及，如把流水线放在用作信号处理和地震数据分析的小型机上，此外也可以把流水线放在微型机基本系统上。例如FPS AP-120B, FPS-64, Data General AP/130和IBM 3838等机器。

上述设计引入了并行原则，这是因为只靠改进线路的速度不可能得到所希望的性能。在为美国国家航天局（NASA）计划采用的国家航空学模拟设备所开展的设计研究工作中，这一点是很明确的，就是它必须基于计算机能够执行每秒 10^9 次浮点运算。CDC公司建议该机包含四条高性能流水线，而Burroughs公司的设计方案是用512个相同的运算器组成一台阵列机。可靠的超大规模集成电路生产工艺出现以后，就提供了一种设计技术，即可用很多个简单的处理单元（PES）制造出很大的阵列。ICL DAP（4096处理单元PES）和Goodyear Aerospace的MPP是早期处理机阵列的一些实例。利用超大规模集成电路技术的优点，处理机阵列在十年内可能会发展。有人可能认为，对上述设计只是大型科学实验室才有兴趣，而不可能对广大计算机用户有影响。然而，许多事实表明，不管怎样，在计算机体系结构方面的那些改进首先是为科学研究需要，随后也一定会成为一般计算机领域的一个组成部分。

因此，似乎需要设计出并行度更高一些（也就是要求其性能比只靠线路速度所能提供的要高）的计算机，同时还要考虑高度并行用的超大规模集成电路的工艺。由于这种需要，很可能使计算机体系结构并行性在80年代成为一个新的发展领域。本书的目的就是叙述流水线机和阵列机设计的原理与分类，并指出这些原理是如何在现有的几种设计（CRAY-1、CYBER205、FPS AP-120B、ICL DAP和Burroughs BSP）中成功地得到应用。还对几种主要用途（矩阵计算，快速傅里叶变换，泊松解）的各种设计特性作比较。高度并行体系结构的出现，也引出了能在并行机上高效执行的数值算法的设计问题。此外，还有能表示这种算法的计算机语言问题。我们认为并行算法和语言是和体系结构同等重要的，因此，有专门章节进行讨论。

在我们的讨论中有一个特点可能是新的，这就是对一台计算机的性能描述过程中，用了

一种双参量描述的概念 (The concept of a two-parameter description)。即除了通常用的每秒浮点运算最高性能 r_{∞} 外,我们还引入半性能长度 (half-performance length) $n_{1/2}$, 它是为了得到二分之一最高性能而使用的向量的长度。这个向量也可以看作是测量设计并行程度的一个参量, 它的值从 $n_{1/2}=0$ (对一台串行机) 到 $n_{1/2}=\infty$ (对一台无限多个阵列的处理机)。因此, 这第二个参量恰好把并行设计按型谱分类, 它的特点就是可用数量来衡量并行性, 从而回答了下面这个问题: 我所设计的计算机的并行程度究竟如何? 一个同样重要的问题是 $n_{1/2}$ 可以定量地选择最有效的算法。于是, 它也回答了下面的问题: 在并行计算机上使用的最好算法是什么? 我们也引入一种代数符号表示法, 用它能把整个计算机体系结构表示成一行, 这样, 绕过了罗嗦的说明, 有助于用一般表示公式的描述方法对设计进行分类。

该书中多数资料是为Reading大学计算机科学系收集的教材。课程的题目是“新型计算机体系结构”。在过去的五年内, 这门课程占用40个授课单位。由于缺少一种合适的教材, 也就促使我们写了这本书。这本书可供大学三年级学生选用, 也可作为具有科学硕士水平人们的专用教材, 或者也可作为准备作关于并行计算方面的论文的博士生的教材。

在手稿的准备过程中, 不少人提出了建议, 通过讨论, 得到了很多人的帮助。其中, 我们愿意提到Reading大学的同事们, 特别是Jim Craigie, John Graham, Roger Loader, John Ogden, John Roberts和Shirley Williams。还有Southampton大学的Henry Kemhadjian和Reading大学副校长Ewan Page博士。Bristol大学Mike Rogers总编辑教授也对手稿提出过改进意见。我们也得到计算机工厂非常慷慨的帮助, 从样本中获得资料和照片。其中, 我们愿意对下面各位表示谢意。ICL Stevenage 研究发展中心的Pete Flanders, David Hunt, Stewart Reddaway; ICL在Euston工作的John Smallbone; ICL和伦敦大学、Queen Mary学院的Dennis Parkinson教授; Bracknell研究CRAY计算机的Stuart Drayton, Mick Dungworth, Jeff Taylor; CDC在英国的David Barkai, Nigel Payne; CDC在Minneapolis的Patricia Conway, Neil Lincoln和Chuck Purcell; Burroughs公司在Paoli的J H Austin; Burroughs公司在伦敦的G Tillot; 在Akton的Goodyear航空航天中心工作的C T Mickelson; Bracknell浮点系统(Floating Point System)的John Harte, David Head和Steve Markham。Jill Dickinson小姐以她高超的技术打印了手稿, 并校对出很多错误, 在此也表谢意。

我们把这本书献给计算机设计者。要是没有他们使人鼓舞的献身精神的话, 我们就不会有这样一种兴趣去研究、学习、分类、使用各种各样的设计。设计任何计算机必须靠一批人的努力, 我们愿意向Seymour Cray, Neil Lincoln, George O' Leary和Stewart Reddaway表示特别的谢意。因为他们是我们详细讨论的计算机的主要设计者。

R. W. Hockney

C. R. Jesshope

目 录

前言 序

1. 引言	(1)
1.1 并行计算机历史.....	(1)
1.2 设计分类.....	(15)
1.3 特性描述.....	(31)
2. 流水线计算机	(44)
2.1 选择和比较.....	(44)
2.2 CRAY-1.....	(45)
2.3 CDC CYBER 205.....	(63)
2.4 FPS AP-120B.....	(82)
3. 处理机阵列	(95)
3.1 什么是处理机阵列.....	(95)
3.2 开关网络.....	(104)
3.3 ICL DAP	(114)
3.4 BURROUGHS BSP.....	(124)
4. 并行语言	(136)
4.1 一般原理.....	(136)
4.2 并行结构.....	(138)
4.3 向量化编译程序.....	(149)
4.4 机器并行编译程序的表示.....	(153)
4.5 问题并行编译程序的表达式.....	(160)
4.6 可移植性.....	(164)
5. 并行算法	(166)
5.1 一般原理.....	(166)
5.2 递归.....	(170)
5.3 矩阵乘法.....	(179)
5.4 三对角线系统.....	(182)
5.5 变换.....	(194)
5.6 偏微分方程.....	(218)

6. 未来的发展	(233)
6.1 工艺	(233)
6.2 设计的限制	(242)
6.3 未来的超级计算机	(252)
附录	(259)
参考文献	(264)

1 引 言

大规模并行技术是本世纪80年代在设计成批供应的计算机时出现的主要革新。在引言中我们追溯了发展史 (§1.1)。然后,展开介绍这些并行计算机分类的原则 (§1.2) 和描述其有关的特性 (§1.3)。为了有效地使用这些并行计算机,程序设计员必须对计算机系统总的组织有所了解。我们合理地将这些组织称为体系结构:就是说体系结构指的是处理机数目和类型,存储器模块,输入/输出通道,以及这些部件是如何被控制和连接的。在第二和第三章中分别叙述了二种主要类型的并行计算机结构,即流水线计算机和处理机阵列结构。要在这些范畴内对全部设计给出一个综合的说明是不实际的,我们主要选择一些成批供应的计算机来加以讨论。这些计算机现在已经有了或者今后会有很大的销售额。并且对这些计算机而言,描述同一问题所选用的任何一种方法,彼此都是很不同的。在第二章中叙述的流水线计算机有CRAY-1、CYBER205和FPS AP-120B。第三章中叙述的处理机阵列有ICL DAP和Burroughs BSP在第六章讨论将来的发展时,我们研究了工艺基础和并行计算机的其他设计限制;然后介绍美国国家航天局(NASA)委托的二个80年代系统结构的设计研究问题。

如果用户要利用这个发展成果,那么新的设计特点需要新的计算机语言和新的数值算法。就并行计算机来说,由于将向量或矩阵的数学概念引入到计算机语言,并且从适合于执行向量数据的观点分析其算法,因此对一维或有时二维数据表,自然可以获得很高工作效率。一种好的计算机语言能大大便于编写程序。第四章是从标准FORTRAN (§4.3),机器并行性表达式 (§4.4)和题目并行性表达式 (§4.5)的观点来从事并行语言的研究。并行计算机的有效性依赖于发明或选择合适的并行算法,在第五章中我们建立了在任何一台并行计算机上测量某种算法性能的原则。然后,将这些原则应用到选择递归、矩阵乘法、三对角线性方程、变换式和一些偏微分方程的算法中。

虽然我们的说明是以1980年前后的情况为基础的,但我们试图建立当计算机体系结构进一步发展时读者仍能应用的一些原则。我们的目的是使得读者通过对一些新体系结构进行分类和对性能进行描述等手段达到对新设计做出评价,从而能为解决特殊问题选择合适的算法和语言。

1.1 并行计算机历史

我们从考虑减少简单算术操作时间,例如减少浮点乘法所需的时间来开始讨论计算机体系结构并行性的历史,时间是从1951年第一台批量生产的计算机,即UNIVAC1出现以后。发展过程如图1.1所示,该图粗略地说明每隔五年算术运算速度增加10倍。由于工艺的改进,硬件性能有很大的提高,再加上计算机各级体系结构的并行性愈来愈高,这样,就能使得计算速度增加得非常快。

50年代第一代计算机使用电子管作为开关元件,门延迟时间约为1微秒^{*}。大约到1960

^{*} 门延迟时间是指从一个逻辑门的输入到下一逻辑门的输入,信号传输所需要的时间(参阅Turn 1974年的著作,147页)。这些数字仅仅表示延迟时间的数量级。

年，象IBM7090这样一类第二代计算机，其开关元件由分立锗晶体管所代替（门延迟时间接近0.3微秒）。大约在1965年生产出硅双极平面型小规模集成电路（ICS），每片具有几个门，

门延迟时间大约10毫微秒。一直到1975年此电路逐渐得到改进，门延迟时间确实可以做得稍小于1毫微秒。用另一种金属氧化硅（MOS）半导体工艺，虽然速度约慢5—10倍，但可能做到更大的组装密度。到80年代初速度和容量略等于第一代真空管计算机的微处理机出现了，它是放在几平方毫米的单片硅上。这一工艺的发展，显然有可能实现各种高度并行的体系结构，但是，在这以前仅是停留在理论研究方面。

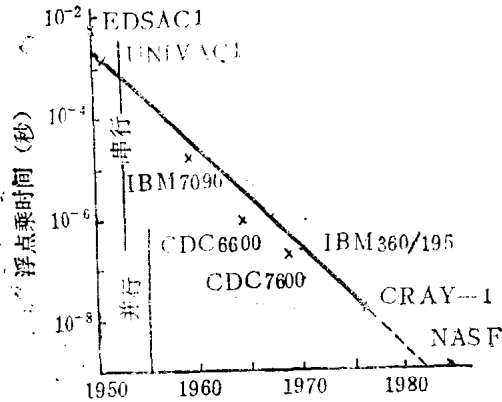


图 1.1 从1950年以来计算机运算速度发展史
(每五年增长10倍)

考查一下从1950年到1975年这段期间，若用门延迟时间的倒数来测量元件的基本

速度，人们能看到它大约增加了 10^8 倍，而当用乘法运算需要的的时间的倒数来测量计算机性能时，则大约增加了约 10^6 倍。这是由于体系结构的改进，主要是并行性的引入，才有可能使机器速度增加了多余的倍数，这就是本书的主题。在第六章中详细地讨论了工艺发展以及它们对计算机设计的影响。Turn (1974) 在他所著的《80年代计算机》(Computer in the 1980s) 一书中也讨论了80年代工艺发展，并提出一些重要的预言。工艺进步对计算机设计的影响，特别是对高性能科学计算机设计的影响，在Parkinson (1982) 所著的《巨型计算机》(Supercomputers) 丛书的下一卷中会有更完整的描述。接着，以后我们的讨论将集中在从结构方面引入并行性这一点上，而不是讨论硬件执行过程的细节。

上面的比较没有考虑到不同的算术和逻辑操作有可能重叠，我们可以把作为测量工艺速度的时钟周期和对某一实际问题每秒执行的有用算术操作数目交替地进行比较，这些操作数目是作为测量其性能用的。Wilkes等1951年列举了在EDSAC1(Wilkes和Renwick 1949) 机器上计算 n 项级数 ($2n$ 次算术操作) 所用的执行时间为 $18n$ 毫秒，该机器的时钟周期为2微秒。这台机器的平均速度为每秒100次算术操作，对比之下，我们可以看到在CRAY-1 (Russell 1978) 机器上做矩阵乘法，则每秒能执行一亿三千万次算术操作。因为CRAY-1的时钟周期是12.5毫微秒。在大约30年期间，整个性能改进了 10^6 倍，其中仅有160倍能归功于工艺方面的改进。人们也应注意在算术操作质量方面也有很大的改进，从EOSAC1机器的36位定点运算发展到CRAY-1机器上的64位浮点运算。

第一代计算机的全部结构是按串行方式工作的，它遵循Burks等1946年设计的存储程序计算机的基本概念，通常简称为Von Neumann结构。这样的计算机包含有一个输入输出(I/O) 部件，一个既存储数据又存储指令的简单存储器，一个为解释指令用的简单控制部件，及一个为处理数据用的简单运算和逻辑部件。后面二个部件叫做中央处理部件或CPU。本书上下文中指的串行工作的重要特点是计算机每一操作必须按顺序执行，或一次只执行一个操作(例如存储器的取数或存数，运算或逻辑操作，输入或输出操作)。而并行性指的是重叠能力或同时执行许多工作。

并行性引入计算机体系结构的主要方法在§1.2中将要全面叙述。他们可以概括为：

(a) 流水线结构——应用相当于工业生产中的装配线技术，改进运算和控制部件性能。

(b) 功能结构——为执行不同功能提供几个独立部件，例如逻辑部件，加法或乘法部件，并允许这些部件同时处理各种数据。

(c) 阵列结构——在统一控制下，提供一个由很多相同的处理部件 (PES) 组成的阵列，这些部件同时执行相同的操作，但是与其相对应的是专用存储器中的不同数据。即所谓连锁步伐操作。

(d) 多道处理结构——提供几台处理机，但每台处理机只执行其各自的指令，通常经过一个共同的存储器相互通讯。

当然，某些特定的设计可以兼有一些或全部上述的并行特点。例如，一个处理机阵列可以把流水线运算部件作为他的处理部件 (PES)，同时多部件计算机中，某一功能部件可以是一个处理机阵列。

到1980年，多处理机设计主要牵涉到几台独立计算机的连接方法，以便最大限度地增加计算机装置的解题能力。这的确是重要的，但是超出了本书的范围。因此，当多处理机作为计算机一部分并行工作时，我们将指出多处理机实例，请读者参看由Enslow (1974, 1977) 和Infotech (1976) 写的评论，这些评论在很广泛的范围内讨论了多处理机系统的性能。在1975年前后，由于出现了便宜的微处理机，使得用许多相互联接的微处理机共同解决一个问题的设想成为现实，这样的系统在80年代很可能成为十分重要的系统。现已提出许多试验系统，并已实现。Lorin (1972) 在他的《硬件和软件的并行性》(Parallelism in Hardware and Software) 和《真正和明显的并行性》(Real and Apparent Concurrency) 这两本书中，对并行性问题作了很好的讨论，这个讨论对计算机设计的所有各方面都有影响 (如运算部件、存储器组织、指令调度、输入/输出、多道处理和多道程序)。此书被推荐为并行性原理的附加读物。

在我们叙述并行性的历史时，我们不想对此做一个全面的评述，而仅是将注意力集中在并行性引入过程中那些重要的发展阶段上。我们也不想叙述所有机器或论文，而我们仅选那些销售数很大或者在计算机科学界特别有影响的机器，同样也只是提到几篇最有影响的科学论文。在图1.2中图解说明了并行性引入计算机体系结构的主要发展趋势和关系。现在我们分别描述主要的发展道路。对计算机早期历史的有关资料 (大多是1950年以前的) 感兴趣的读者们，请参阅Randell 1975年所编写的论文集，这本文集叫做《数字计算机的起源》

(The Origins of Digital Computers)。也还可以参阅Hartree's (1950) 所写的《计算仪器和机器》(Calculating Instruments and Machines) 一书。Rosen (1969) 写了一篇特别有价值的历史评述，他叙述了从1946年ENIAC到1968年CDC7600这段期间计算机的发展。Bell和Newell (1971) 在《计算机结构——读物和举例》(Computer Structures: Readings and Examples) 一书中给出了我们提到的许多计算机体系结构的详细资料，包括1950年至1970年这个时期主要计算机系列的许多论文。并给出在别处不易得到的设计和详细历史资料。在《计算机协会通讯》(Communications of the Association of Computing Machinery) 专载计算机体系结构的专刊 (ACM 1978) 中叙述了其他机器 (MU5, ATLAS, IBM370, UNIVAC-1100, DEC10, CRAY-1)。Auerbach (1976 a) 出版了一系列关于计算机技术的报告，在这些报告中归纳出了多数成批生产计算机的体系结构，这些归纳小结是有用的。

1.1.1 1960年以前并行计算机的发展情况

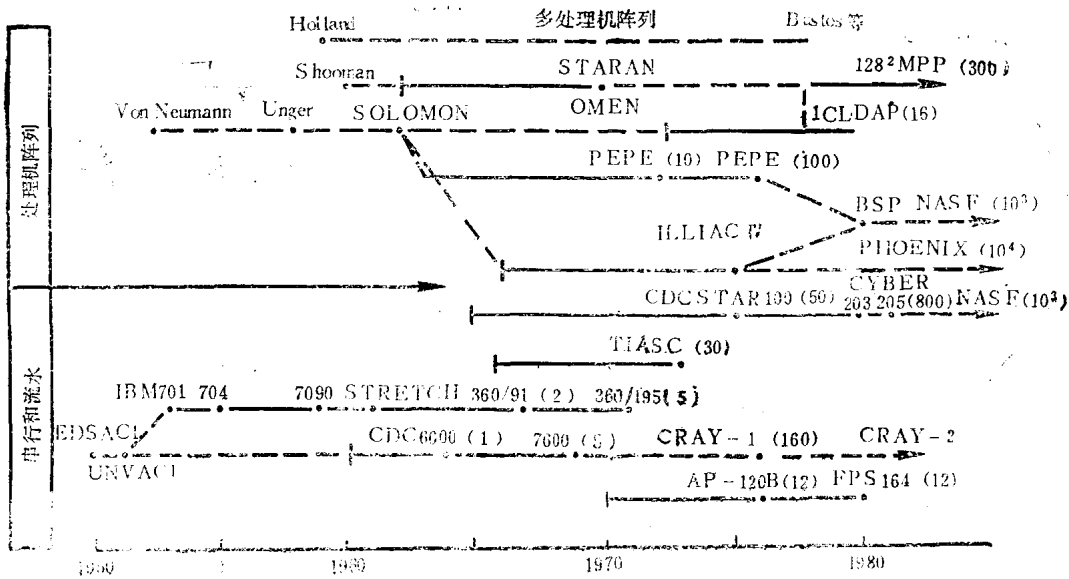


图 1.2 并行计算机从50年代早期到80年代中期得到发展。树状发展图 (Evolutionary tree) 表示出体系结构的联系和影响。括弧内的数字只是表示计算机速度性能, 单位是每秒百万次浮点运算 (Mflop/s)。实线指出了设计和制造过程, 星号表示交付第一个操作系统的时间, 虚线表明各机器体系结构上有密切的关系。

一般认为, 在计算机设计方面, 最早是在日内瓦综合图书馆1842年10月发表的L F Meabrea的文章中提到并行性, 文章题目叫《Charles Babbage发明的解析机概略》(参阅Morrison和Morrison 1961年的著作第244页, 以及Kuck 1977年的著作)。在列举解析机的效用时, 他写道:

“其次, 关于节约时间, 为了使我们确信这一点, 我们只需回想一下, 两个20位长的数相乘, 至多才需要三分钟。同样, 当执行一长串相同的计算, 例如造数字表需要的那些计算时, 要使机器发挥作用, 以便同一时间能给出几个结果。如能这样做, 将可大大缩短总的计算过程”。

Babbage最终设计的计算机没有包含执行并行操作这一性能, 但是比真正由于技术进步到能够实现计算机并行性提前100年, Babbage就已想到使用并行性概念去改进机器性能, 这一点是很清楚的。Charles Babbage在关于机械差分和解析机的著作中, 确实最早提出了许多计算机的基本概念。当时由于需要制作可靠的天文学表格, 在此推动下, 他才写的这部作品 (Babbage 1822, 1864; Babbage 1910; Randell 1975)。

第一批存储程序计算机, 即剑桥大学研制的EDSAC (Wilkes和Renwick 1949) 和宾夕法尼亚大学研制的EDVAC (1952年完成), 以及他们的商业改型 (例如UNIVAC1; 见Eckert等1951) 使用的是几台容量为100个字的水银延迟线存储器。既然在这种存储器中, 一个数是从最低有效位开始按时间顺序连续送到读出线路的头端, 于是可以自然而经济地按位执行二个数的运算。例如, 由非常简单的逐位相加电路来实现二个32位数的加法, 要用32个机器周期才能完成。以这种方式完成的运算叫做位串行方式运算。早期在计算机中使用的并行性是指一个数的所有位并行操作以提高其运算速度。

十分有趣的是, 串行运算结构曾被认为是由于电子元件引入计算机而获得的优点之一, 在这之前, 计算机传统地是并行执行运算的。例如, Babbage曾拒绝在他的差分机上用串行运算

方式，因为用串行方式执行时间太长 (Babbage 1864, Morrison 和 Morrison 1961 34页)，曾预定他的解析机有50个十进制数并行执行运算。十九世纪四十年代的标准台式计算器能做到12个十进制数并行工作。电子元件的速度远远超过以往机器的机械和机电元件速度，性能能够改进 10^3 到 10^4 倍，与此同时，由于一次只处理一位，设备也大为节省。

试验性的自动计算装置 (ACE) 和其商业改型——英国电气公司生产的通用电子数字计算机，DEUCE——这些是具有几种并行特征的第一代真空管计算机。1947年，当 H D Huskey博士到英国国家物理实验室参观时，承担了自动计算装置 (ACE) 试验模型的设计任务，设计是按照 A M Turing博士的思想 (参阅其传记，Turing 1959) 进行的，该装置于1951年首次运行。这台机器有11根水银延迟线，每一根具有32个字，每字32位，一次循环时间为1毫秒(ms)。它的运算器是串行的，每一时钟周期1微秒(μ s)内出现一位。该机具有卡片输入机，卡片穿孔机和乘法器。十分有趣的特征是它们能和其余部分并行工作，并且它的指令 (我们现在将它划为向量指令) 能控制延迟线上所有32个数进行若干次操作。例如，当某一张卡片通过读卡机和穿孔机时，中央处理机 (CPU) 在进行十翻二或二翻十换算。而当乘法器执行一条需要2ms的无符号乘法时，为调整符号，加法器同时在完成一次符号乘法。这些都是功能并行的实例。一条指令能在32个时钟周期内保持有效 (实际上是重复的)，例如，允许延迟线内所有数位由一条指令计算求和。这些装置常用于计算检查和，并且是向量指令最早的一个例子。英国通用电气公司1954年生产了电子数字计算机DEUCE (Haley 1956)，其中快速存储增加到12根延迟线，并补充了8000字的磁鼓。当磁鼓传送数据时，并行性也是明显的，因为机器的其余部分同时能够进行计算。DEUCE提供的独立定点除法器 (除法时间为2ms)，像乘法器一样，能和机器其余部分并行工作。但是它和乘法器使用同一寄存器。三个单字、三个双字和四个四倍字的各种延迟线在DEUCE上被用作快速存取寄存器。一个单字和一个双字延迟线是与定点加法器联用的，也能起累加器作用。

从静态随机存取存储器能方便地并行读出一个字的所有位数，所以，在计算机设计中位并行运算就变得现实可行。1952年在新技术研究所 (IAS) 研制出了使用并行运算的第一台实验计算机，紧接着在1953年完成了具有并行运算的第一台商业计算机，IBM 701机。在这两台机器中使用了 Williams和 Kilburn (1949) 发明的阴极静电射线管。几年以后，出现了第一批使用磁心存储器的机器。这些机器当中最成功的当然是IBM 704机，这种机器大约卖了150台。这些机器不仅有并行计算，而且是首次用了硬件浮点运算部件，于是有效地加快运算速度，使性能超过了以前用软件实现浮点运算的一些机器。第一台IBM 704机是从1955年开始研制的，1975年关闭了最后一台机器。第一代真空管计算机工作了20年，Mclaughlin (1975) 对它的重要历史作用做了叙述。

在IBM 704机以及这个时期的其他机器上，数据的输入或输出都必须通过运算部件的寄存器，因此影响了运算速度。最初用的输入输出设备有联机卡片输入机 (每分钟150—200张卡片)，卡片穿孔机 (每分钟100张卡片) 和行式打印机 (每分钟150行)。不久，第一台磁带机代替了这些初始的联机输入和输出设备 (每秒15000个字符，比卡片输入机或行式打印机至少快100倍)。在单独的输入输出计算机IBM 1401上装有从卡片到磁带和从磁带到打印机的脱机装置。然而，这些磁带的速度比处理机处理数据的速度仍大约慢1000倍。所以，对IBM 704和整个计算机装置来讲，输入输出是一个影响性能的主要瓶颈。

由于允许计算机的运算部件、逻辑部件和数据的输入和打印并行工作，这样至少是部分地解决了输入输出问题。因此，可以增加一台称为输入输出通道的单独计算机，它的唯一工

作是传递数据，把数据从慢速外部设备如卡片输入机、磁带或行式打印机以及计算机的主存储器等设备中取出，或者输入到这些设备中去。输入输出通道一旦由主控制部件启动，它就能进行大批数据的传送，而运算部件继续做其本身的工作。输入输出通道有适合于它本身特点的输入输出指令，也有它自己的指令处理部件和寄存器。在1958年将六条这样的通道加到IBM 704机，并将此机器重新命名为IBM 709。这就是多道处理的早期实例。这些机器的开关逻辑仍旧使用电子管，它们的寿命是不长的，因为到这时候固态晶体管已经成为一种可靠元件。IBM 709按晶体管技术重新工程化，改名为IBM 7090。并从1959年开始销售。IBM 7090再加上它的改进型IBM 7094和7094 II是非常成功的，大约制造和卖了400台。

一些新器件当它处于早期发展的时候，一开始常常想找到很多革新思想，之后转入对某一特殊的设计方案进行重点投资的时期。在这时期以后，由于投资规模的限制，要想进一步的改革将是非常困难的。从汽车的发展过程可看到这一点。在1900年前后使用的发动机原理是很宽的，包括汽油、蒸汽和旋转发动机，以后的巨大投资是用在汽油驱动内燃机方面，这几乎没有人能设想加以改变。同样，计算机设计方面也提出过许多新颖的结构原理，并且在五十年代已被讨论过，但直到1980年，只有建立在单指令和单数据流基础上系统，才在商业上获得成功。这多半是由于出现了超大规模集成电路工艺及便宜的微处理机，才使得一些结构在八十年代能够得到实现。

在1952年Leondes和Rubinoff叙述了一台多操作计算机，该机用的是旋转磁鼓存储器。机器DINA是一台解拉普拉斯方程、扩散方程和波动方程的数字分析机。之后，于1958年Zuse在其‘场计算机’的设计中提出了相似的概念。von Neumann (1952) 提出空间连接阵列处理机原理，他指出，具有29种状态计算单元的一台二维阵列机能够执行全部操作，因为它能模拟图灵 (Turing) 机的性能。继这一理论上的发展之后，Unger (1958) 提出进行实际设计的建议，这些设计建议被认为是七十年代出现的SOLOMON、ILLIAC IV 和ICL DAP等计算机的原始设计。

与此类似，Holland (1959) 的论文叙述了多台处理机的组装，每台处理机只服从于本身的指令流。也把他的论文看作是第一台大规模多处理机设计和后来微处理机连接设计的原始文本。这与20年后由Pease (1977) 和Bustos等(1979)为解决扩散问题提出的那些设计相像。由Pease提出的间接二元 n 立方思想写成的文章是一篇连接 2^n 台微处理机的设计论文，微处理机可按1维、2维、到 n 维立方进行布局，适合于宽范围的普通数字算法，如快速傅里叶变换。预想能够用到16384台微处理机 ($n=14$)。信息管理系统 (IMS) 协会 (Millard 1975) 宣布的超立方体是以Pease思想为基础的，它包含一个四维立方体，此立方体的每个结点带有二台微处理机，其中一台用于内结点通讯，另一台为数据变换所用。

1.1.2 快速标量计算机

标量计算机是这样一种计算机，它提供的指令仅仅对各单个数据进行操作。与此相反，向量机提供的指令则是对有序的成组数组（这叫做向量）进行操作。在六十年代和七十年代这段时期，快速标量计算机的发展史基本上是越来越多地把并行性引入单指令流单数据流计算机所有的串行设计的历史，如IBM 7090，到五十年代末，该机的结构已变得十分标准。

在结构和软件两方面有深远影响的计算机是ATLAS (Kilburn等1962, Sumner等

1962, Howarth等1961,1962, Lavington 1978)。大约在1956年, Manchester大学提出有关这台机器的设想,并在Kilburn教授指导下,由工业部门和大学共同组成的小组联合进行设计。到1961年样机就能在大学里运行。于1963年, Ferranti有限公司完成了第一台生产样机(该公司不久后成为国际计算机和制表机公司的一部分,再后来就是国际计算机有限公司)。ATLAS主要因首创使用了复杂的多道程序操作系统而出了名,该操作系统是建立在庞大的一级虚拟存储和中断系统的基础上。操作系统为当前处于不同执行阶段的程序分配资源。对使用者提供的一级存储虚地址空间接近 10^9 个字。然而,访问多级存储是自动进行转换的,多级存储包括磁心存储(16K字, $K=1024$),磁鼓(96K字)以及作为后备存储的磁带。可以由页式系统实现转换,在存储体系不同级之间,按512字为一页进行信息传送。中断系统使慢速输入输出部件能独立地工作,只有当它需要完全占用CPU时,才中断CPU的计算。

除上述的结构改进外, ATLAS机器早就利用并行性来改进它的计算性能。它把2微秒(μs)的主磁心存储器划分为四个独立的存储体(称为存储栈),在最顺利的情况下,在一个存储周期可并行地取回二条指令和它们的二个操作数。运算本身是按位并行方式工作的(Kilburn等1960)。功能并行性体现在变址计算用的24位独立加法器也叫做B运算部件中。另外还在主要的定点和浮点运算部件上有所体现。该部件使用了单一的主累加器,它有48位长,其中40位为尾数,8位为阶码。B运算部件与128个24位变址寄存器(称为B线)联接使用,此变址寄存器是由周期时间为 $0.7\mu\text{s}$ 的快速磁心存储器组成。指令是单地址的,将地址所指的内容与主累加器的内容进行运算。指令有48位长,其中10位为操作码,24位为主磁心存储器地址,14位为两个变址寄存器地址(每个7位)。取一个或二个指定的变址寄存器的内容与主地址相加而得到操作数地址。使用流水线的原理就是要重叠执行指令的以下各阶段,如在B部件计算24位操作数地址、取操作数以及在主累加器的48位运算等各个阶段。从以下事实能够判断出流水线的效果。在一系列的浮点加法实例中可以看到,若按顺序执行指令的所有阶段,则每一次操作需要6微秒,而用流水线结构,则可减少到平均1.6微秒。在任何时候都有二条、三条、或四条指令处于不同的执行阶段。对应的浮点乘法时间是5微秒,测量得到的执行一条FORTRAN程序指令所用的平均时间近似等于3微秒。Ferranti ATLAS用标准门延迟为12毫微秒(ns)的分立锗晶体三极管和二极管实现其逻辑电路。这台计算机大约用80000支晶体三极管。

为了有效地利用多运算部件,多寄存器和多存储器这样的并行特点,需要首先看一看指令流,从中决定出那些指令能够同时执行而不改变程序的正确性。在查清楚程序执行中的并行性之后,以最佳执行方式调度发到运算部件的指令。Keller(1976)已对先行(look-ahead)这类问题进行了研究。在多数快速的标量计算机如CDC 6600及IBM 360/91中,都考虑了先行问题。

在Seymour R Cray(1957年公司的创始人,副董事长和主要设计师)影响下,数据控制公司在评价串行计算机概念的过程中,逐步引入了功能并行性和流水线概念。最早在1964年交付的CDC 6600是第一台在设计中采用功能并行性的计算机(Thornton 1964)。Thornton是具体设计负责人,他于1970年在其书《计算机设计-数据控制机6600》一书中令人满意地叙述了机器的技术发展。CDC 6600采用1微秒磁心存储器,存储器划分为32个独立存储体,可并行工作。它有为乘(重复)、除、加、长加、移位、布尔演算、转移和递增(重复)等操作设立的10个独立功能部件,以及与慢速输入输出设备灵活连接的10个外围处

理机。虽然这10个外围处理机分时使用一个运算和控制部件，但是它们执行的是独立程序；所处理的数据分别存于各个单独的存储器内。因此在逻辑上它们相当于独立的处理器，这是多道处理的实例。除了为整数循环计数及地址计算的增量部件（只作18位加减）之外，为了适应科学研究需要，该机所有运算器都做成浮点形式（60位字长）。CDC 6600做的很成功，它占领了许多科学应用市场，在以前这些市场是由IBM 7090和7094独占的。继6600之后，数据控制公司在1969年提供了改进型机器CDC 7600。由于时钟周期从6600的100毫微秒减少到27.5毫微秒，所以速度约提高了四倍。6600的十个串行结构功能部件被八个流水功能部件和一个不能以流水方式工作的除法串行部件所代替。由于流水部件额外地增加了速度，不再需重复乘法和增量部件，只加了一个计算一个字中1的个数的特别功能部件。CDC 6600（改名CYBER 70系列74型）和CDC 7600（改名CYBER 70系列76型）的体系结构必定是科学领域方面二个最成功的体系结构。后一种机器已经安装了50台以上。

在六十年代和七十年代这段期间，从IBM生产计算机的历史来看，设计中也是逐步引入了更多的并行性。从1956年开始积极发展了这项技术。它与Los Alamos科学实验室订立合同，要制造一台比IBM 704快100倍的计算机。这台计算机称为STRETCH (Dunwell 1956, Bloch 1959)，后来起名为IBM 7030，只有短时间在市场出售。该机的主要新颖特点是它具有几种先行设备，即具有为指令接受、译码、计算地址和预先为几条指令取操作数用的先行设备。另外，它将存储器分成二个独立的存储组，这样能并行的将数送到运算部件。存储器取送数的最大传输率从而增加一个系数，此系数等于存储组数。这是在存储器上第一次使用并行性设计，使得低速磁心存储器能较满意地与快速处理器相匹配。大概所有以后研制的大型快速计算机都使用了这类存储器分组结构（有时称交叉存取）。第一台STRETCH在1961年交给了Los Alamos。但是它没有达到其设计目标，公司也很不满意这种产品的经济效果，所以，在制造了七个系统以后 [一台被安装在英国Aldermaston核武器科学研究中心 (AWRE)]，这种计算机就被停产了。后来，向主要用户推荐了速度慢一些但用途很普及的IBM 7090系列。

在有了研制STRETCH的经验之后，IBM似乎对高速计算机失去了兴趣。在1964年宣布了IBM 360系列计算机，但是，这个系列中没有一台机器的性能能比得上CDC 6600，而CDC 6600的第一台是与该系列机同年安装的。CDC 6600取代了IBM 7090，并将多数大规模科学中心从IBM公司转到CDC公司，这一惊人事实，在IBM公司引起了强烈的反应。因此，到了1967年，造出了IBM 360/91机 (Anderson等1967)，其性能为CDC 6600的两倍。这台机器有STRETCH的先行设备，跟CDC 6600一样，为浮点和整数加计算设有单独的执行部件，它们当中每一个部件都是流水的，并且能并行工作。也引入了流水线原理，以便加速指令处理、加速取指令、译码、地址计算等连续操作以及加速连续指令的重叠取操作数。用这种方法，当几条指令通过流水线流动时，几条指令同时处于不同的执行阶段。到了1969年，CDC 7600发表了，其工作性能比360/91大约提高一倍。在1971年IBM又公布了360/195，这台机器的性能与CDC 7600差不多。IBM 360/195 (Murphy和Wade 1970) 是360/91结构与360/85引入的高速缓冲存储器结构相结合的产物。在慢速主存和运算寄存器之间引入高速缓冲存储器 (cache) 的概念，相当于又回到了Ferranti ATLAS计算机时期 (Fotheringham 1961)。在360/85中，其高速缓冲存储器用了周期为162毫微秒容量为32768字的半导体存储器，将最近要使用的数据存放在若干个以64个字节为一字组的存储块内。假如指令中需要的数不在高速缓冲存储器，从慢速主存（容量为4兆字节，周期为756毫微秒的磁心存储器