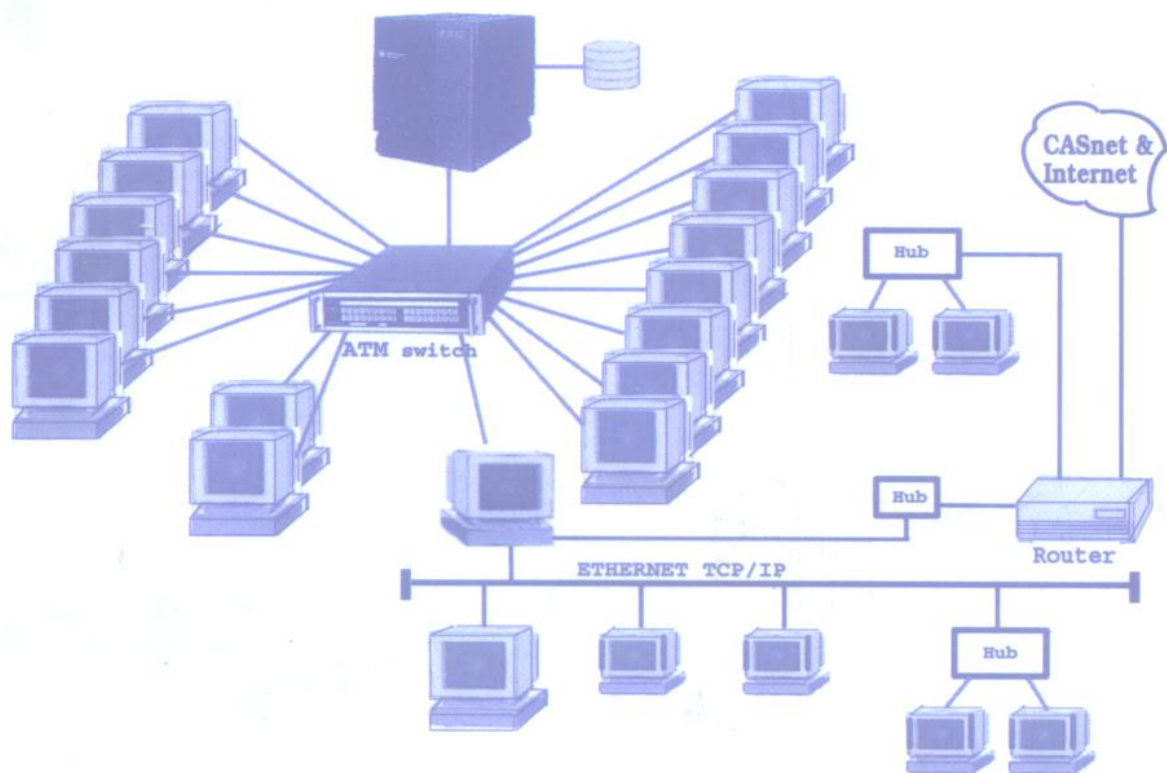




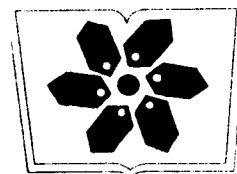
网络并行计算与 分布式编程环境

孙家昶 张林波 迟学斌 汪道柳 编著

科学出版社

1P538.6
593

390897



中国科学院科学出版基金资助出版

网络并行计算 与分布式编程环境

孙家昶 张林波 迟学斌 汪道柳 编著



科学出版社

1996

(京)新登字 092 号

JS/34/22

内 容 简 介

网络并行计算是当前国内外计算机科学研究与应用领域中最引人注目的前沿课题之一。本书是在作者多年来从事并行计算研究与应用的基础上，结合国内外并行机软硬件环境发展的最新动态撰写而成的。书中系统地叙述了网络分布式并行环境与编程的基本概念、主要方法与应用，并引用了 Internet 网上相应的最新资料。

本书分为两大部分。前五章阐述了国内外分布式并行机与国内网络环境的最新发展、分布式并行计算的基本方法以及常用的矩阵并行算法和区域分解方法的应用。后五章详细介绍了可移植网络并行环境 PVM 的获取、安装、测试、编程与实现。为方便读者阅读，书后附有若干 PVM 应用程序及 PVM 系统库函数的调用说明；考虑到我国教育科研部门计算机设备的现状，书中专门介绍了 Linux 操作系统以及微机网络并行计算环境。

本书可供高等学校应用数学、计算数学、计算机以及其它有关专业作为教学参考书，也可供对并行处理、并行算法、网络计算及 Linux 操作系统有兴趣的科研和工程技术人员参考。

图书在版编目 (CIP) 数据

网络并行计算与分布式编程环境 / 孙家昶等编著.

—北京: 科学出版社, 1996.3

ISBN 7-03-005000-2

I. 网… II. 孙… III. ①计算机网络—并行计算方法②分布式网络—程序设计 IV. TP393

中国版本图书馆 CIP 数据核字 (95) 第 2094 号



双青印刷厂 印刷

新华书店北京发行所发行 各地新华书店经售

*

1996 年 3 月第 一 版 开本: 787×1092 1/16
1996 年 3 月第一次印刷 印张: 15
印数: 1-5000 字数: 333000

ISBN 7-03-005000-2/TP·509

定价: 19.80 元

序

自从世界上第一台电子计算机问世以来,计算机高速度的发展已持续了半个世纪,它对科学技术以至整个人类社会都产生了巨大的影响.推动高性能计算机发展的第一动力是科学技术发展对于高性能计算不断增长的需求.然而计算机单机技术的发展受着诸如光速与物理尺寸(不能无限小)等本质的限制,向量计算机 Cray-1 诞生 20 年来单机速度只提高了一个数量级.正是这种计算机单机技术发展的有限性与科学工程计算需求的无限性之间的矛盾决定了计算机发展必然走上多机并行的道路.由于采用了大规模并行,使得计算机的理论峰值在 20 年内提高了四个数量级(从 1 亿次提高到 1 万亿次浮点运算/秒).反过来科学技术需要求解的课题规模愈大、问题愈复杂,并行的潜力也愈大,这也要求高性能计算走并行的道路.

在我国“八五”与“九五”期间,国家各有关主管部门正在加大对并行计算项目的投资强度.石油、气象、航空航天等应用部门都已将并行计算立项,投入强度逐年增加,这将是我国并行计算工作者为国民经济服务的主战场.所有这些都是极好的机遇,同时也是一种挑战.我们认为,发展我国的高性能计算机必须要结合我国国情;更多地考虑我国高性能计算各方面的实际需要.能否抓住当前的有利时机,充分发挥我国并行计算队伍的能力,研究高水平的理论问题,并解决一批有关国民经济或国家安全的“巨大挑战”计算问题,将是决定我国并行计算领域科研与教育工作在 20 世纪最后几年内能否兴旺发达的关键所在.

并行计算机的出现,一方面给并行计算提供了必要的物质基础和机遇,另一方面又给用户带来了许多新的问题和困难.目前并行机种类繁多,互不通用.如何发挥并行机的效率,是用户最关心的问题.一台性能很高的并行机,如果由于各种原因使用中不能充分发挥应有的效率,性能会呈数量级下降.因此,并行计算方法与并行算法的研究是并行计算的一个关键问题.根据实际问题的需要,结合并行机的硬件特点(如处理机的个数、内存方式与向量结构等)进行适当的区域分解及分层分批计算,减少数据传输,充分发挥问题的并行本性,才能设计出高效的并行算法.

为了更好发挥并行机的效率,需要并行机硬件设计、并行系统软件和管理软件研制及并行计算方法与算法研究三个方面的彼此了解与密切配合.并行机软硬件的研制人员要深入了解应用部门的实际需求,在并行机的整个研制过程中,尽量考虑到并行机的使用方便.从事实际并行计算的人员要多懂得一些并行机设计的思路,才能更好发挥并行机的效率.

孙家昶等人编著的《网络并行计算与分布式编程环境》一书是他们多年来从事并行计算工作的部分总结,该书为并行计算研究的这三方面人员在技术方面相互沟通提供了一些必要的知识,因此是一本很好很有价值的参考用书与教学用书.

周毓麟

1995 年 10 月

前 言

经过 20 多年的研究与发展, 并行计算已牢牢确立了它在高科技关键技术中的地位. 作为并行计算物质基础的计算机硬件, 近年来发展尤为迅速. 银河与曙光 1000 的研制成功, 表明我国高性能并行计算机的研制能力跨上了一个新的台阶. 国际上各大并行机公司也正在竞相进入我国市场. 目前不仅大型机市场中并行机的型号层出不穷, 而且工作站中的多处理机种类也愈来愈多, 正从研究领域进入市场, 逐步趋向成熟. 并行机的应用正从传统的科学工程计算逐步渗透到金融管理等部门. 网络技术的开发与应用, 使我国的有关科研与教育部门能通过终端直接与国外交流; 国产并行机的陆续研究成功与投产, 使得国内更多人能够直接接触到并行机; 公开的并行软件环境以及多任务操作系统如 PVM, Linux 等的应用与推广, 使得我国仅拥有少量工作站甚至微机的许多科研教育部门只用少量投资就能建立起自己的并行计算环境.

并行计算研究的主要目标应该是解决一批具有重大挑战意义的科学工程计算问题. 高性能并行计算之所以重要, 就是因为有一大批具有重大挑战意义的科学工程计算问题要求在本世纪内计算性能达到每秒万亿次级运算 (10^{12} 或 Tera ops). 这些“巨大挑战”问题有: 全球天气预报; 分子与原子核结构; 大气污染; 燃料与燃烧机理; 计算生物化学; 新兴材料; 医学解剖与模拟; 新药研制; 催化剂设计; 天文学数据图象处理; 石油勘探开发以及与国家安全有关的重大问题.

网络并行计算是当前国内外计算机科学研究与应用领域中最引人注目的前沿课题之一. 本书是在作者多年来从事并行计算研究与应用的基础上, 结合国内外并行机软硬件环境发展的最新动态撰写而成的, 是编著者们参加国家“八六三”计划及国家攀登计划工作的部分总结. 书中系统地叙述了网络分布式并行环境与编程的基本概念、主要方法与应用, 并引用了 Internet 网上 95 年的最新资料, 可供高等学校应用数学、计算数学、计算机以及其它有关专业作为教学参考书, 也可供对并行处理、并行算法、网络计算及 Linux 操作系统有兴趣的科研和工程技术人员参考.

本书分为两大部分. 前五章阐述了国内外分布式并行机与国内网络环境的最新发展、分布式并行计算的基本方法以及常用的矩阵并行算法和区域分解方法的应用, 后五章详细介绍了可移植网络并行环境 PVM (Parallel Virtual Machine) 的获取、安装、测试、编程与实现. 为方便读者阅读, 书后附有若干 PVM 应用程序及 PVM 系统库函数的调用说明. PVM 是当前国际上公认的一种消息传递标准软件系统, 它比最新标准可移植消息传递界面 MPI (Message Passing Interface) 成熟程度高. PVM 具有通用性强、系统规模小、使用方便等优点, 已能在国际上所有的并行计算机上运行. 一批标准的应用软件已经或者正在移植到 PVM 的平台上. 正是由于这些杰出的成就, PVM 已被美国列为 1994 年整个工业界 100 个最佳研究开发 (R&D) 的成果之一.

考虑到我国教育科研部门计算机设备的现状, 书中第九章专门介绍了微机上的免费 UNIX 操作系统 Linux 以及微机网络并行计算环境.

本书的编著是在孙家昶研究员主持下，与张林波、迟学斌、汪道柳三位博士合作完成的。本书第一章至第四章由孙家昶执笔；第五章由迟学斌执笔；第六、七、九章由张林波执笔；第八章由汪道柳执笔；第十章及附录 B 的内容由迟学斌和张林波提供；附录 A 由曹建文和张林波整理。白德琴同志打印了部分原稿。本书采用的中文 $\text{\TeX}$ 排版系统，是由张林波、汪道柳根据科学出版社的要求提供的，排版主要由张林波和汪道柳负责。

本书在出版过程中得到了很多专家与同行的鼓励与帮助。我国老一辈的科学家，中国科学院周毓麟院士、夏培肃院士与石钟慈院士对本书的出版极为关注，周毓麟院士还特地为本书作序。中国科学院软件研究所冯玉琳研究员、计算技术研究所刘慎权研究员、清华大学陈景良教授审阅了本书的写作提纲。中国科学院计算技术研究所的唐志敏研究员、马影琳研究员与国家智能中心的张浩同志分别对第一章、第二章和第八章中一些技术问题的叙述方式提出了宝贵的意见。本书中提及的几台国产分布式并行机的原始资料分别是由中国科学院计算技术研究所、国家智能中心与江南计算技术研究所提供的，有关国外计算机的最新资料主要取材于 Internet 网。在整个写作过程中，作者们得到了国家“八六三”计划“306-103-3”课题组与中国科学院“八五”重点项目“中大规模并行计算方法研究”项目组诸多同事的支持。科学出版社鞠丽娜同志认真负责的工作态度对于提高本书的出版质量起到了重要作用。我们在此一并表示感谢。

本书从写作稿到出版的全过程是在计算机上完成的。尽管经过多次易稿与修订，书中的疏漏在所难免，恳请广大读者批评指正。由于高性能计算机的发展极为迅速，对书中引用的有关数据感兴趣的读者请随时参照 Internet 网上的最新结果。

作者

1995 年 12 月

目 录

第一章 并行计算机发展的现状与趋势	1
1.1 并行计算机的类型.....	1
1.1.1 SISD 型计算机.....	2
1.1.2 SIMD 型计算机.....	2
1.1.3 共享内存 MIMD 多处理机.....	3
1.1.4 分布式存储 MIMD 多处理机.....	3
1.2 若干分布式国产并行计算机简介.....	5
1.2.1 BJ-02.....	5
1.2.2 BJ-1.....	6
1.2.3 江南 3 号.....	7
1.2.4 曙光 1000.....	9
1.3 国际上若干高性能大规模并行机介绍.....	11
1.3.1 Paragon XP/S.....	11
1.3.2 CM-5E.....	13
1.3.3 T3D.....	15
1.3.4 SP2.....	17
1.4 计算机性能评估与测试程序.....	19
1.4.1 基本测试程序.....	20
1.4.2 Linpack.....	22
1.4.3 其它测试程序.....	22
1.5 全世界 500 台最高性能计算机统计分析.....	23
第二章 网络计算平台	27
2.1 计算机网络简介.....	27
2.1.1 常用的计算机网络.....	27
2.1.2 TCP/IP 协议.....	29
2.2 Internet 及其网络服务.....	30
2.2.1 Internet 的网络服务器.....	30
2.2.2 网络计算数据库 Netlib 与 NHSE.....	31
2.3 中科院网络计算平台 NCFC 与 LSEC.....	33
2.3.1 中国国家计算与网络设施 NCFC.....	33
2.3.2 科学与工程计算国家重点实验室 LSEC.....	35
2.4 网络并行与可移植的消息传递环境.....	36
2.4.1 网络并行的优点.....	36
2.4.2 可移植的异构编程环境 PVM.....	37
2.4.3 消息传递标准平台 MPI.....	37
2.4.4 其它可移植的编程环境.....	39
第三章 并行计算概论	40

3.1 高性能计算与并行计算.....	40
3.1.1 并行处理是大型科学与工程计算之必需.....	40
3.1.2 美国 HPCC 计划.....	41
3.2 并行算法的基本概念.....	42
3.2.1 并行算法的目标.....	42
3.2.2 并行加速比定律与可扩展性.....	42
3.2.3 并行算法的分类.....	45
3.3 并行编程的基本方法.....	46
3.3.1 网络并行编程的基本模式.....	46
3.3.2 负载均衡的基本方法.....	48
3.4 并行计算方法是研制高效并行数值软件的基础.....	50
3.4.1 并行计算的技术路线.....	50
3.4.2 关于“分而治之”的原则与重新排序.....	51
3.5 并行计算模型.....	53
3.5.1 Log P 计算模型.....	53
3.5.2 C^3 计算模型.....	53
3.6 并行测试程序 NAS 与 PARKBENCH.....	53
3.6.1 并行测试程序 NAS 内容简介.....	53
3.6.2 并行测试程序 PARKBENCH.....	54
3.6.3 NAS 的最新测试结果.....	55
第四章 区域分解法.....	62
4.1 Schwarz 方法及其并行化.....	62
4.1.1 基本思想.....	62
4.1.2 模型问题.....	63
4.1.3 子区域分解.....	63
4.1.4 经典 Schwarz 方法.....	64
4.1.5 并行 Schwarz 算法.....	65
4.2 无重叠区域分解法.....	65
4.3 加法 Schwarz 预条件子迭代.....	67
4.3.1 子区域预条件子.....	68
4.3.2 粗网格预条件子.....	69
第五章 矩阵并行计算.....	71
5.1 矩阵相乘的若干并行算法.....	71
5.1.1 行列划分算法.....	72
5.1.2 行行划分算法.....	72
5.1.3 列列划分算法.....	73
5.1.4 列行划分算法.....	73
5.1.5 Cannon 算法.....	74
5.2 线性方程组的解法.....	75

5.2.1 分布式系统的并行 LU 分解算法	75
5.2.2 具有共享存储系统的并行 LU 分解算法	77
5.2.3 三角方程组的并行解法	78
5.3 对称正定线性方程组的并行解法	80
5.3.1 Cholesky 分解列格式的并行计算	80
5.3.2 双曲变换 Cholesky 分解	81
5.3.3 修正的双曲变换 Cholesky 分解	83
5.4 三对角方程组的并行解法	84
5.4.1 递推法	84
5.4.2 分裂法	86
5.5 异步并行迭代法	87
5.5.1 异步并行迭代法基础	87
5.5.2 线性迭代的一般收敛性结果	88
5.6 矩阵乘积的 Occam 语言程序	89
5.6.1 串行矩阵乘法程序	89
5.6.2 并行矩阵乘法程序	90
第六章 PVM 安装调试及应用程序编译链接	93
6.1 PVM 的安装	93
6.2 PVM 控制台	96
6.3 PVM 的配置与调试	98
6.4 PVM 应用程序的编译、链接与运行	100
第七章 PVM 消息传递用户界面	103
7.1 进程控制	103
7.2 虚拟机动态配置	106
7.3 消息传递	108
7.3.1 消息缓冲区	108
7.3.2 数据打包	109
7.3.3 发送与接收	111
7.3.4 数据拆包	112
7.4 任务组	113
7.5 编程实例	114
7.6 XPVM	118
第八章 PVM 系统的实现及网络性能	121
8.1 PVM 的构件	121
8.1.1 任务标识符	121
8.1.2 结点机类型	122
8.1.3 消息模式	122
8.1.4 异步通告	122
8.1.5 PVM Daemon 和程序库	123

8.2 PVM 的消息.....	123
8.2.1 消息片与数据缓冲区.....	123
8.2.2 包缓冲区.....	124
8.2.3 libpvm 中的消息.....	125
8.2.4 pvmd 中的消息.....	125
8.2.5 pvmd 中的消息处理.....	126
8.2.6 用于控制的消息.....	126
8.3 PVM Daemon.....	127
8.3.1 启动与终止.....	127
8.3.2 机器表与机器的配置.....	127
8.3.3 任务表.....	128
8.3.4 等待现场.....	128
8.3.5 pvmd 的影子 pvmd'.....	129
8.3.6 启动从 pvmd.....	129
8.3.7 资源管理器.....	131
8.4 PVM 的通信协议.....	132
8.4.1 消息的头.....	132
8.4.2 pvmd-pvmd 通信.....	132
8.4.3 pvmd-task 和 task-task 的通信.....	133
8.5 消息路由.....	134
8.5.1 pvmd.....	134
8.5.2 pvmd 与远程的任务.....	135
8.5.3 libpvm.....	135
8.5.4 广播.....	136
8.6 任务环境.....	136
8.6.1 环境变量.....	136
8.6.2 标准输入输出.....	137
8.6.3 跟踪.....	138
8.7 资源的限制.....	138
8.7.1 PVM Daemon 的限制.....	138
8.7.2 任务的限制.....	139
8.8 PVM 网络性能测试.....	139
8.9 并行 NAS 性能测试.....	142
第九章 Linux 操作系统与微机网络并行计算环境.....	144
9.1 Linux 操作系统简介.....	144
9.2 Linux 操作系统的安装.....	145
9.2.1 安装时的启动.....	147
9.2.2 硬盘的分区.....	148
9.2.3 基本系统安装与配置.....	149

9.2.4 其它模块的安装.....	155
9.3 X 窗口的安装.....	156
9.3.1 X 窗口系统 XFree86 的安装.....	156
9.3.2 配置文件 XF86Config 的格式.....	156
9.3.3 X 窗口系统的启动.....	160
9.3.4 XF86Config 文件的配置与调试.....	161
9.4 Linux 系统的使用.....	162
9.4.1 退出与关机.....	162
9.4.2 虚拟终端.....	163
9.4.3 X 窗口与虚拟终端.....	163
9.4.4 随机资料.....	163
9.4.5 使用 LoadLin 启动 Linux 系统.....	163
9.5 Linux 的 C 与 Fortran 编译器.....	164
9.6 微机上的 PVM 系统.....	165
第十章 PVM 应用实例.....	167
10.1 随机数发生器.....	167
10.2 矩阵乘积.....	167
10.3 多重网格法.....	168
10.3.1 差分方程与多重网格算法.....	168
10.3.2 区域划分.....	169
10.3.3 PVM 程序简介.....	170
10.3.4 输出结果实例.....	172
附录 A PVM 3.3.7 子程序说明.....	174
附录 B 若干 PVM 应用程序.....	188
B.1 随机数发生器 PVM 程序.....	188
B.2 矩阵乘积 PVM 程序.....	189
B.3 三维多重网格 PVM 程序.....	197
参考文献.....	220
索引.....	223

第一章 并行计算机发展的现状与趋势

1963年2月18日,美国西屋(Westing House)宇航实验室的工程师利用九个CPU部件,搭成一个 3×3 阵列,从而获得了某个偏微分方程的解.现在有人把这个日子当作是并行计算机的誕生日……

1.1 并行计算机的类型

Flynn^[1]根据指令流与数据流方式的不同,曾提出了计算机系统的如下分类方法:

- 1) 单指令流单数据流(SISD, Single Instruction Stream & Single Data Stream);
- 2) 单指令流多数据流(SIMD, Single Instruction Stream & Multiple Data Stream);
- 3) 多指令流单数据流(MISD, Multiple Instruction Stream & Single Data Stream);
- 4) 多指令流多数据流(MIMD, Multiple Instruction Stream & Multiple Data Stream).

直到目前国内外尚没有见到多指令流单数据流(MISD)方面的产品.

并行计算机是指有两个或两个以上的处理器连接起来可以并发操作的机器.这样的机器有时也称为多处理器.并行计算机主要可分为两大类:

- 1) SIMD(单指令流多数据流),各处理器同一时刻执行相同的指令,但数据不同;
- 2) MIMD(多指令流多数据流),各处理器对于不同数据可以独立进行操作. MIMD多处理器一般又可分为两类:共享存储与分布式存储.

Gorden Bell在文献[2]与[3]中提出了并行计算机的进一步分类方法.

世界上真正的第一台并行机系统是1972年研制成功的Illiack IV,当时美国投资3000万美元研制了这台 $8 \times 8 = 64$ 阵列机(每个处理单元的理论峰值为每秒百万次浮点运算,局存16KB,系统实际的总体性能约为每秒200万次浮点运算),以后每年投资200万美元用于维护.这台机器已作为计算机发展史上的划时代产品陈列在波士顿美国国家博物馆.作为并行计算机的产品,Illiack IV不是一个成功的例子.在之后十几年内以向量化结构为主的计算机占了统治地位,其代表是1976年诞生的Cray-1(主存8MB,理论峰值为每秒一亿六千万次浮点运算),到80年代末,主要产品是以Cray序列为代表的向量-并行机.

在计算机结构上,向量计算机的特点是利用流水线的概念,把一个计算部件拆成几段,在流水线的每一部分操作一个子功能,通过时间重叠实现并行加速.这样,如果对于每个加法需要5个时钟周期,对于流水线上的向量加法则几乎平均一个时钟周期出一个结果,这样效率可提高五倍.当然,为了充分利用流水线的效率,必须把所需的数据及时充满流水线.为此向量机都提供若干基本向量操作指令,如向量加法、标量乘向量、按分量乘法、除法等,还有一些宏向量指令,如向量的内积、开方,向量单元的聚集、压缩、合并与累加等.

根据几十个常规应用软件的统计,60-88%左右的标量计算可以被向量化,这表明

向量计算的加速比一般可达到 4-8。但是，单纯依靠单机提高计算机性能的潜力毕竟有限。对于超大型计算问题，人们想要追求的加速比是以数量级为单位的，单纯的向量机不能满足要求。并行化则是转向“多指令流多数据流”方向发展，大规模并行机与网络并行机群可以克服向量机的局限。这就是我们在 90 年代中期讨论并行机的主要出发点。

1.1.1 SISD 型计算机

绝大多数串行计算机属于单指令流单数据流 SISD 类型，其 CPU 主要有：

1) CISC (Complex Instruction Set Computer, 复杂指令型), 主要指通常的大型机 (如 IBM 390)、小型机 (如 VAX) 及 PC 机 (如 Intel 80x86, P6)。

2) RISC (Reduced Instruction Set Computer, 精简指令型), 主要是工作站，目前最主要的芯片有 Sparc, RS6000, HP PA-RISC, MIPS R-8000, Alpha, Power (Power PC, Power-2) 等。

采用单机内的一系列并行处理技术，有可能在精简指令技术的基础上把处理器性能再提高一步。常采用的单机并行化技术有：多功能部件 (multiple functional units)、流水线 (pipelining)、超高速缓存 (cache)、多级存储 (hierarchical memory)、超长指令 (Very Long Instruction Word, VLIW)、多线程 (multi-threaded)、超标量 (superscalar) 技术，以及向量指令与链接 (chaining) 技术等。

由少量向量处理器组成的向量并行机仍是当前国际上所谓超级计算机 (Supercomputer) 的主要模式之一。

1.1.2 SIMD 型计算机

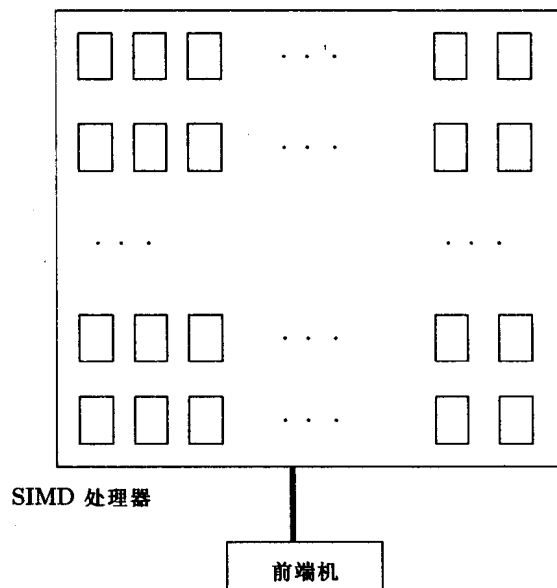


图 1.1 SIMD 并行机基本结构

使用 SIMD 机器通常要通过一个前端机或是一台主机执行指令。这类机器的基本结构如图 1.1 所示。属于 SIMD 的机器有：

- 1) 阵列处理机, 如 DAP;
- 2) 向量机, 如 Cray-1, Fujitsu VPX, Convex C1, 国产机 YH-1;
- 3) SIMD 并行机, 如 Thinking Machine 公司的 CM-2, MasPar MP-1, MP-2.

1.1.3 共享内存 MIMD 多处理机

共享内存 MIMD 多处理机有一个每个处理器都能访问的统一的内存。早期的共享内存机器提供一条宽频带的总线访问内存, 这条总线把所有的处理器以线性方式连接到内存 (见图 1.2), 后来使用交叉开关 (crossbar switch)。图 1.2 中的处理单元可以是标量处理器 (如 SGI Power Challenge), 也可以是向量处理器 (如 Cray X-MP, Cray Y-MP, Fujitsu VPP500, NEC SX-3, Hitachi S-3800, Convex C3, 国产机 YH-2 等)。

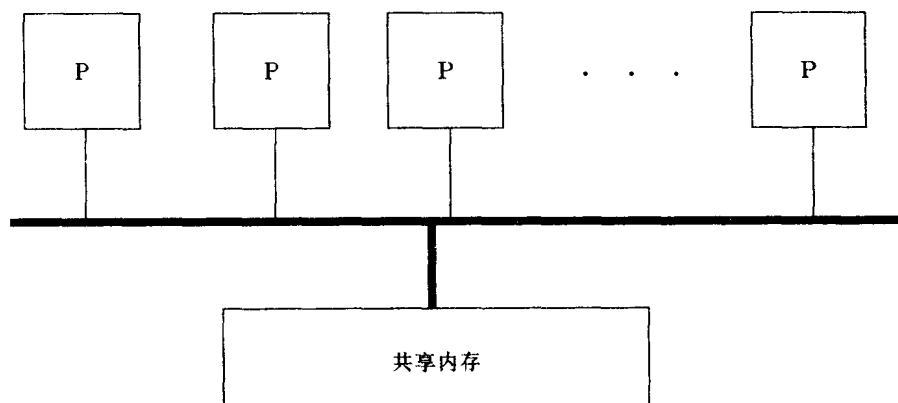


图 1.2 共享内存型

共享内存多处理机的一个主要问题是可扩展性较差, 当多处理机需要同时访问共享内存 (如全局变量) 时, 由于产生内存争用现象而严重影响效率。为此有的共享内存系统采用多总线结构。如 Cray 6400 有四条总线, 每条总线最多挂 16 个处理器, 整个系统最多挂 64 个处理器。

缓解内存争用矛盾的另一个途径是采用非均匀存储访问 (如 KSR, Cray T3D, Cedar 等), 以代替常规的均匀存储访问方式。

1.1.4 分布式存储 MIMD 多处理机

分布式存储 MIMD 中每个处理机都有自己可直接访问的内存, 称为局部存储器, 每个处理器称为系统中的一个结点处理机。结点之间的联结方式 (或称拓扑结构) 有: 总线结构 (见图 1.3)、环状结构、Mesh 结构 (见图 1.4) 或超立方体结构 (见图 1.5)。

消息传递型可扩展多计算机:

1) 大规模并行计算机 (MPP 型) 适合中等粒度并行: 如 IBM SP2, Intel Paragon, Meiko CS-2, nCUBE, CM-5 等。

2) 网络并行机群 (cluster) 系统, 适合粗粒度并行, 如: 共享内存 SPMD 处理器机群、全对称 SMP 机群、局域网连接的各种各样工作站或 PC 机群以及由异步传输模式 (ATM) 连接的广域网络机群。

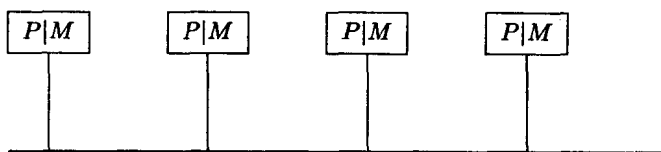


图 1.3 总线结构

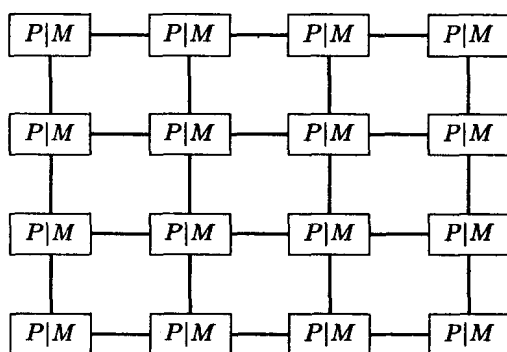


图 1.4 Mesh 结构

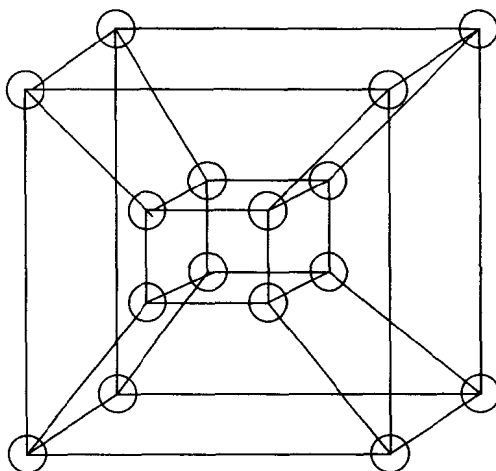


图 1.5 四维超立方体结构

1.2 若干分布式国产并行计算机简介

本节从面向应用的角度出发, 简要介绍目前我们接触到的若干分布式国产并行计算机. 主要分为以下六个方面: 1) 总体结构; 2) 内联网络; 3) 结点处理机; 4) 内外存储与 I/O; 5) 操作系统; 6) 编程语言与环境.

1.2.1 BJ-02

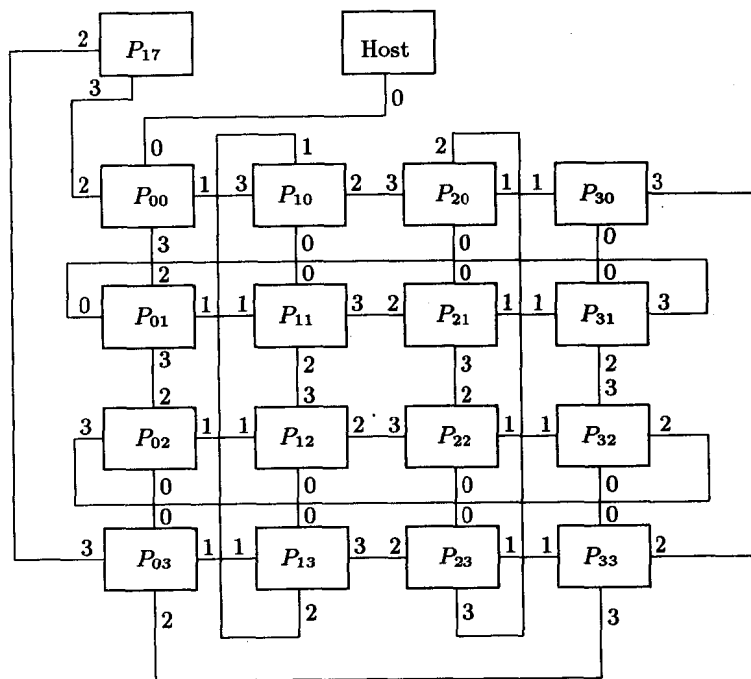


图 1.6 BJ-02 并行机结构框图

(1) 总体结构

- 分布式 MIMD 系统
- 消息传递结构
- 处理器结点个数: 17
- 体系结构: 一个 master, 16 个 slave 组成四维超立方体或 mesh 结构
- 最大存储规模: $16 + 4 \times 16 = 80\text{MB}$
- 一个插件上四个处理器, 外加机箱共有四个插件
- 插件板插到微机总线插槽上可作为加速板
- 完成日期: 1991 年
- 研究部门: 中国科学院计算技术研究所

(2) 内联网络

- 系统与微机通信通过 IBM 总线接口
- 提供直接存储访问 (DMA) 接口

(3) 结点处理机

- Transputer T800/25, 浮点处理速度 2.5Mflops, 带宽 40MB/s
- 每个结点有 4MB 内存
- 每个结点有四条通信链路 (见图 1.6), 传输速率 20MB/s
- Cache 4KB, 数据速率 120MB/s

(4) 内外存储与 I/O

- 前端为 PC 机
- 内存总量 80MB
- 外存为微机硬盘

(5) 操作系统

- 使用微机操作系统 DOS

(6) 编程环境与语言

- 可使用 TDS 环境, 包括编辑、文件处理、编译、运行和调试等功能
- 可使用 OCCAM 并行处理语言
- 可使用并行 C 与并行 Fortran 语言, 由配置文件标明处理器之间的连结方式

1.2.2 BJ-1

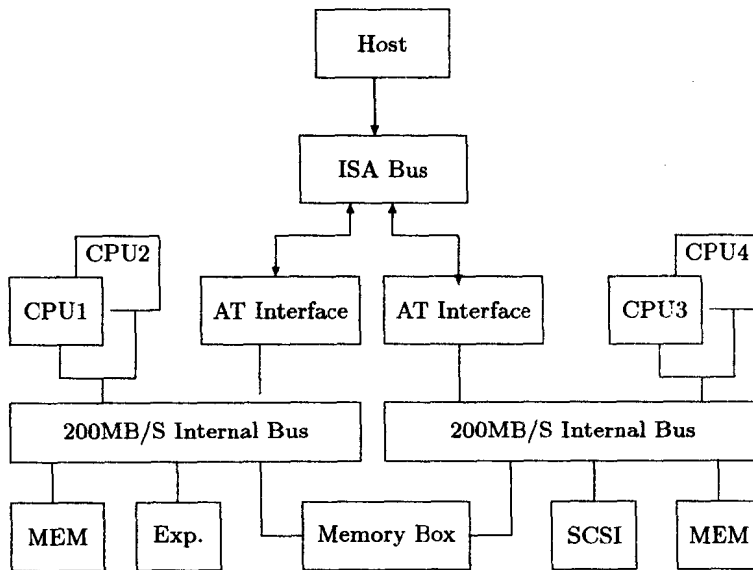


图 1.7 BJ-1 并行机结构框图

(1) 总体结构

- 由共享存储处理器簇组成的分布式 MIMD 系统 (结构框图见图 1.7)
- 处理器簇之间通过双端口存储器进行消息传递
- 两个处理器簇, 每个处理器簇两个 i860/XP 芯片
- 存储方式: 每两个处理器共享 16-256MB 存储器
- 系统访问方式: 通过 80486 PC 机作为前端机