

737337

5057
18473

分布式并行处理系统探索

武汉大学分布式系统研究组

武汉大学出版社

分布式并行处理系统探索

武汉大学分布式系统研究组

武汉大学出版社

内 容 简 介

本书是武汉大学 分布式系统研究 组编辑的有关分布式并行处理系统 研究的论文集。

本文集围绕着研制 WuPP-80 分布式并行处理系统, 从其体系结构、分 布式操作系统、并行处理语言、分布并行算法及其应用等方面系统地进行了论述。全书共分三部分十二篇文章: 第一部分为概论, 着重分析了分布式系统与并行处理相结合的现实意义, 并介绍了 WuPP-80系统的基本特征; 第二部分详细论述了 WuPP-80 系统的实现方法; 第三部分介绍了异步并行算法的基本原理及其在 WuPP-80系统上的实际应用。

本文集可供计算技术和计算数学专业科研人员、工程师以及该专业的研究生和高年级学生参考。

分布式并行处理系统探索

武汉大学分布式系统研究组

*

武汉大学出版社出版

武汉大学出版社发行

武汉大学印刷厂印装

*

787×1092毫米 16开本 8印张 195千字

1984年4月第一版 1984年4月第一次印刷

印数1-2000

统一书号: 13279·14 定价: 1.05元

前 言

分布式计算机系统是将多台小型/微型机互连成的一种新型计算机系统，是从分散处理的概念出发来组织计算机系统的，冲破了传统的集中式单机局面，具有高的性能价格比，灵活的系统可扩充性，良好的实时性、可靠性与容错性等潜在优点，是近几年来计算机科学技术领域中极受重视，并成为迅速发展的一個新方向。

任何事物的发展都离不开现实环境，而且各个发展阶段相互渗透。从科学的角度看，我们正处于信息时代。为了提高信息处理速度与增加信息吞吐量，以“空间换时间”（即用计算空间的冗余换取计算时间的缩短和系统性能的提高）的并行处理结构思想，一直不同程度地贯穿于信息处理的全过程，并导致了计算机系统性能的极大改进，产生了多种新型计算机系统。现在，随着大规模集成工艺和微型计算机的迅速发展，又出现了新的动向和可能性，这就是把多台计算机与并行处理相结合而构成另一类新型的分布式并行处理计算机系统。

分布式计算机系统的研究和开发正处于兴起阶段，许多名词尚无统一定义，为区分人们通常所谈及的分布式系统的概念，我们认为分布式并行处理计算机系统是由若干个可分离的、能自治的、而彼此间又有交互作用、并可协同解决同一任务的计算机为单元所构成的，对用户是透明的多机系统。也就是说，分布式并行处理计算机系统是一种多指令多数据流（MIMD）体系结构，既可按原单机软硬功能分布处理，又可联合协同并行处理，具有如下四个基本特性：（1）软硬结构上，具有模块性；（2）工作方式上，具有自治性；（3）系统功能上，具有协同并行性；（4）对用户讲，具有透明性。按照这种定义，多处理机、多功能部件系统、固定分配任务的多机系统、主从系统、计算机网络等不是本文集中所说的分布式并行处理计算机系统。

分布式计算机系统最重要的物质基础——大规模集成电路和微型计算机的发展，将直接影响分布式计算机系统的结构原理和应用的开拓。在分布式并行处理系统研究中，最吸引人的是把数量极多的微处理机（或大规模集成电路单元）有机地组织起来，构成一个具有巨大计算（或处理）能力的高性能通用系统。尽管目前还没有见到这种理想的系统，但已有人进行了许多深入研究，取得了一些进展。

武汉大学计算机科学系（研究所）与数学系联合于1982年10月首次研制成功了用于科学计算的分布式并行处理计算机系统 WuPP-80。该系统既可按原单机分散处理，同时又可按任务组合系统中计算机协同并行处理，提高了处理速度，扩大了存贮容量，为该系统研制的一类新型异步并行算法，以新的思想建立了适合于这类系统的计算方法，并成功地应用在 WuPP-80 系统上，解算了应用广泛的一类数学物理问题，特别对角缘问题、裂缝问题以及解 Navier-Stokes 方程等高难度问题，得到了令人满意的结果，充分发挥并验证了 WuPP-80 系统在并行数值计算上的优越性，开拓了微型计算机的应用范围，推动了异步并行算法的研究，展现出设计 MIMD 系统新的应用前景。当然，在系统规模、容错能力，任务的分划和分配，程序设计的方便性等方面还有许多问题有待进一步探索。特别从科学和人类进步的要求

来看，“人工智能”从本质上讲代表着计算机发展的未来，亦即计算机智能化代替人思考部分问题是发展的必然趋势。不言而喻，分布式结构将是这类高级系统的普遍表现形式，还会有大量课题需要研究。

本文集是我们对分布式系统进行的初步探索。在许多同行的鼓励和支持下，我们将研制 WuPP-80 系统的有关论文，作适当修改、整理，有机地联结成一个整体，将其出版，以供广大读者共同研讨。全书共分三部分十二篇文章：第一部分概论，着重分析了分布式系统与并行处理相结合的现实意义，并概要地介绍了 WuPP-80 系统的基本特征；第二部分详细论述了 WuPP-80 系统的硬件和软件结构，如何在保留原单机软硬件资源的基础上，采用一种改进了的全互连拓扑结构，增添并行处理语句和功能模块构成并行处理语言及对称结构的分布式操作系统；第三部分以“区域分裂”、“异步并行”、“混乱松弛”、“局部拟边界迭代”等新思想、新概念较系统地论述了“异步并行算法”的基本原理，并给出在 WuPP-80 系统上的一些数值试验结果。

本文集可作为计算技术和计算数学专业科研人员、工程师、研究生和高年级学生的参考读物。书中各部分自成系统，相对独立，读者可根据自己的兴趣，有选择地进行阅读。

由于我们对分布式并行处理计算机系统的研究时间不长，经验积累不多，因此缺点在所难免，希望读者予以指正。

本文集在编写与出版过程中得到我校计算机科学系（研究所）有关同志的大力支持，在此表示衷心的感谢。

武汉大学分布式系统研究组

1983年4月

目 录

第一部分 概 论

- 分布式并行处理与 WuPP-80 (1)
- 分布并行算法和分布式并行处理系统研究 (9)

第二部分 分布式并行处理系统 WuPP-80

- 分布式并行处理系统 WuPP-80 的体系结构 (19)
- 分布式并行多微处理机操作系统——DPOS (36)
- WuPP-80 并行 FORTRAN 语言及其编译系统的设计 (46)
- 关于 WuPP-80 系统中存贮共享问题的讨论 (57)

第三部分 异步并行算法及其应用

- 一类新型的异步并行算法 (60)
- Schwarz算法及其在分布式并行处理机上的应用 (I)
- 解线性椭圆型边值问题 (71)
- Schwarz算法及其在分布式并行处理机上的应用 (II)
- 解弱非线性椭圆型边值问题 (80)
- Schwarz算法及其在分布式并行处理机上的应用 (III)
- 解二维定常 Navier-Stokes 方程 (92)
- Schwarz算法及其在分布式并行处理机上的应用 (IV)
- 解奇异性椭圆型边值问题 (102)
- Schwarz算法及其在分布式并行处理机上的应用 (V)
- 解非定常数学物理问题 (115)

第一部分 概 论

分布式并行处理与WuPP-80*

分布式计算机系统及其应用的研究与开发,是当前世界计算机科学界的一个主攻方向,与计算机工业的发展紧密相关。本文试图从“分布式系统”的产生和发展及其深入研究与“并行处理”相结合等方面提出某些看法,并给出实验模型的特征。

一、分布式系统的产生和发展

首先,由于七十年代硬件技术的进步,大规模集成电路的迅速发展,微处理机、微型计算机相继出现,而其成本大幅度地下降,使得能按其性能/价格比将若干台小型/微型机构成分布式多机系统来取代集中式的主机结构开拓了新的途径。

文献[1]给出了作为系统价格函数的价格/性能比(图1.1),指出了分布式多微处理机系统的价格/性能比线是位于一般单处理机系统价格/性能比线之上还是之下的问题。

文献[2]列举了一些很有成效的实例。C. mmp同PDP-10按四个基准程序的平均性能进行比较,以每个美元每秒钟的指令数计算,C. mmp的性能/价格比超过PDP-10的3.59倍;在天气预报方面,分析了1985年期间,两种计算机体系结构的性能/价格比,第一种是按当前的流水线运算单元以及阵列部件等概念设计的,另一种则是微处理机组成的MPS网络,MPS具有40:1的优势[3]。

另一方面,单计算机系统规模日益庞大、复杂,其运算处理速度的提高有限:一是信息传输速度受光速 c 的限制,电信号在电路中的传输速度约为光速的一半;二是物理尺寸受氢原子直径 d_H 的限制[4]。特别在许多实时环境中,目标的分散性[5],集中式的主机达不到快速响应,解决不了复杂的数字问题,不能满足高可靠性的要求,不能提供最佳的或者有效

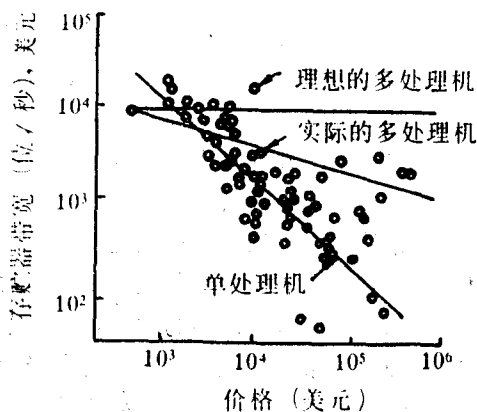


图 1.1 作为系统价格函数的价格/性能比

* 本文于1982年10月在“全国分布式系统学术讨论会”上宣读过。修改后,刊于“人工智能学报”,1983年。此处又略加修改。

的经济处理效果,因此,许多研究工作者提出了“多指令多数据流”计算机结构,特别是利用廉价的微型计算机构成分布式系统[6],分散处理,冗余硬件,模块化结构,使系统灵活、可靠,或者冲破冯·诺依曼的体系结构,象“数据流驱动的数据流计算机”[7],作为提高程序的执行速度。如 CHOPP 系统由32000个 μP 构成,其计算能力可达 16×10^4 MIPS,在运算速度方面超过 ILLIAC IV (速度为120MIPS)的1300倍[8]。Maryland 大学正在研制一种高度并行的多微处理机系统 ZMOB,它由256个 Z80A 组成,每秒可执行1亿条指令(8位字长),具有16MB的高速分布存贮系统,与 Cm* 相似,是一台通用的研究计算机,应用于分布式图形匹配、分布式数据库、计算机视觉和实时判别(感觉)与信件开关等系统。这清楚地表明,分布式处理结构及其应用成为当代计算机科学家们所热衷研究的领域。

再一个方面,计算机应用的社会化,促进了人类社会生活、生产文明的向前发展。计算机网络,单计算机的生产过程控制和社会管理,促使人们向着更高一级的全面自动化处理方向前进,自然地联想着过程控制、社会问题的分散性等特征,必然要冲破传统式的单处理机局面,这是科学技术及管理的必然发展趋势。☆

因此,分布式计算机系统的研究从七十年代初期开始兴起[9]、[10],随着技术条件的更加完善,硬件成本的下降,到了七十年代的后期,就逐渐形成了一个发展迅速且很有前途的新兴领域,吸引着广大的科学工作者去从事研究和开拓新的领域。

二、分布式并行处理

大家都熟知,“分布式系统有高的性能价格比,灵活的系统可扩充性,良好的实时性、可靠性与容错性”等许多优点[11]、[12]、[13]、[14],而并行计算是提高计算机的效率,并加快对应用问题解的一种非常重要的方法。近二十年来,“并行性”成为了计算机设计的一个确定原则,并发展成为“并行处理”这一专门领域[15]、[4]。

因此,能否把分布式系统应用于并行处理领域,特别是数字计算方面,目前还存在着看法上的分歧。有的认为微型机速度低、字长短、有限的编址能力不适于组成高性能的并行处理系统;有的在描述中限制太多或范围太广,强调了某些不正确的特性。因为至今几乎所有的设计都和应用有着密切的联系,所以,不易给出一个确切的定义[13]。但由多台微型计算机组成一种“分布式系统”,其结构本身的“分散性”,每个结点的“自治性”,与许多复杂的应用问题所固有的并行性特征相映,很有研究意义。这是我们探索研究“分布式并行处理系统”的基点。

文献[16]考虑递归矢量函数

$$X(k+1) = f[x(k), x(k-1), \dots, x(k-m+1)] \quad (1)$$

其中 $x(k) \in R^n, k \in \{0, 1, \dots, k\}$ 。方程(1)可以表示通常在控制论和信息处理中用一组差分方程所描述的某些动态系统的行为,矢量 $X(k+1)$ 依赖于前 m 个值,当 $X(k+1)$ 的所有分量同时计算时,计算(1)的第一级并行性就实现了。又假设每一个分量可以写成

$$X_i(k+1) = f_i(\phi_{i1}, \phi_{i2}, \dots, \phi_{ij}, \dots, \phi_{ir_i}) \quad (2)$$

其中 $\phi_{ij} = \phi_{ij}[x(k), x(k-1), \dots, x(k-m+1)]$ 。如果把 ϕ_{ij} 分配给不同的计算机,则可独

立地同时计算所有 $\phi_{i,j}$ ，显示了第二级的并行性。这些分析清楚地表明了这类递归算法所固有的并行性，其结构很容易用微型计算机成分布式处理系统来实现。

随着32位微处机（例如16032微处理器）的出现，其处理速度比大型机更快，又可配置具有虚拟存贮功能和浮点运算功能等，使分布式系统具有适于构成并行处理的新型计算机系统的现实意义越来越明显[11]。

为了使新的计算系统能有效地运行，需要提出新的算法或修改旧的算法。在现实生活中，许多应用问题显示出高度的局部化程度，即一个问题可分划成许多相对独立的小部分。文献[17]、[18]、[19]、[20]基于“证明了解一类偏微分方程及优化问题的 Schwarz 算法的收敛与松弛的顺序无关，而敛速与之有关。”明确地提出了三个任意性：区域分块的任意性，松弛因子的任意性，及松弛顺序的任意性等，这为建立“一类新的异步并行算法”提供了理论基础。大量的数值模拟试验（包括在 WuPP-80 系统上试算）表明，这类算法特别适于在“分布式并行处理机”上解算“线性椭圆型边值问题”、“弱非线性椭圆型边值问题”、“二维定常的 Navier-Stokes 方程”及“奇异性椭圆型边值问题”等（参看第三部分），甚至于可以在不同的区域上，采用不同的离散化方法及相应的不同解法。

文献[21]正在研究流体性质的 Monte Carlo（蒙特卡罗）Simulation，亦提出了“异步算法”，使用 Cm* 中的50台 LSI-11 处理机，计算256个水分子的系统，他们确信 Cm* 这样的多处理机结构对流体结构的 Monte Carlo 计算能提供一种有效的解法。

在非数值计算方面，象古典的“D 氏问题”，由于其元素的分层结构及相关性较少，有可能运用“分布式并行处理系统”而加快其解决。

从已有的多处理机系统方面的实例测量结果表明，“程序的局部化原则”对于分布式处理系统是很重要的。文献[1]谈到在 Cm* 中，存贮访问可分局部访问、群内访问和群间访问三级，使用基准程序测量的时间分别是 $3.5\mu s$ 、 $9.6\mu s$ 和 $26\mu s$ 。由于“程序的局部化原则”，使局部访问的命中率达到90%左右时，其平均存贮访问时间缩短为 $4.1\mu s$ 。

文献[22]对于代码和数据在公用共享存贮器与私用局部存贮器中给出了吞吐量的渐近线如图1.2所示。

为了对系统中任何地点的数据或可执行的代码进行快速存取[23]，提出了存贮器采用分

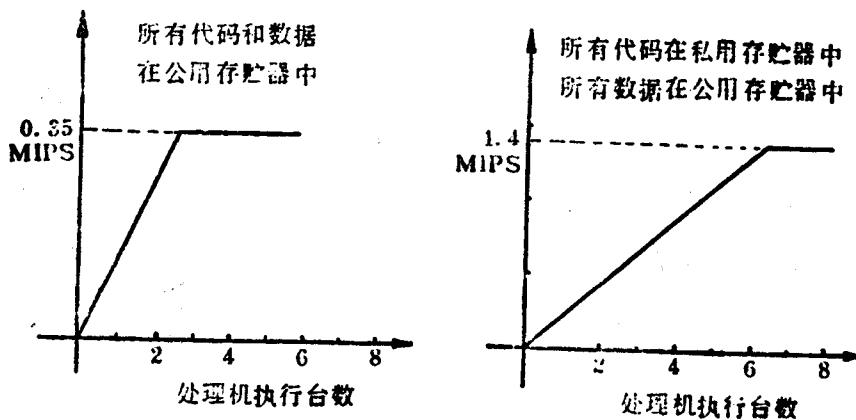


图 1.2 吞吐量渐近线

散方式，把共享内存和每个处理器结合在一块比较好，因为这样对一个共享数据字的传送，只需要进行一次存取，而且因为它的对称性，有可能实现硬件的高度模块化。

因此，我们所探讨的“分布式系统”应由若干台具有同样功能和处理能力的基本处理单元所组成。每个基本处理单元带有自己的控制器、局部存储器 and I/O 通道，形成一个具有自治能力的子系统。这些基本处理单元再通过特定的联结方式互连起来，实现处理机之间、处理机与存储器之间的相互通讯。这种联结方式是分布并行处理计算机体系结构的标志，而且互连接口愈深入处理机的内部，其耦合度愈紧密，而并行处理能力就愈强。在控制方式上，采用分散的方式动态地管理和分配多个分布而又相互联系的资源，特别能协调地完成同一个任务。换句话说，分布式并行处理系统是由若干个可分离的、但彼此间有交互作用并可以协同起来解决同一个实际任务的能自治的计算机所构成的系统。这种系统，可以把一个大型的实际任务（数值的或非数值的）分成若干个彼此可以通讯、但又可以并行地运行的子块进行处理，亦可多个作业分而自治独立地在其子系统上运行或可动态地重新组合，适于多方面的应用。简单说来，可以认为分布式并行处理系统是一种多指令多数据流计算机网络，具有如下四个基本特性：（1）模块化；（2）自治性；（3）协同并行性；（4）透明性。

三、WuPP-80 系统模型

在教育部、国家科委的直接资助下，我们于1980年初开始探讨“分布式系统”的“分布并行算法”、“体系结构”、“分布式操作系统”、“并行语言”以及“可靠性、性能评估”等诸方面的研究工作。

目前已用4台（可扩充）LSI-11 微型计算机互连成了一个完整的系统，即武汉大学分布式并行处理系统（简称为 WuPP-80）。建造 WuPP-80 的目标有两个：一是为算法研究与系统软件研究提供一个理想的环境和较大自由度的实验样机；二是在现有条件下，尽量探求一种高性能/价格比的互连结构。这部分将概要地介绍 WuPP-80 的某些特征，有关详细论述参看本书后两部分。

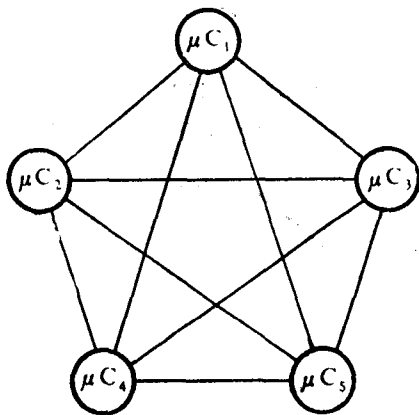


图 1.3 全互连结构

1. 互连结构的特征

因为全互连专用路径结构 [24]（如图 1.3 所示）不存在对算法的影响，所以在 WuPP-80 中，采用了一种类似的互连方案（参看图 1.11）。

在 WuPP-80 中，任一 μC_i 直接同所有其他的 $\mu C_j (i \neq j)$ 相连接，每一对 μC 可以同时进行通讯，这比单总线（甚至于多总线）的连接提供了宽得多的通讯频带，有利于提高整个系统的执行速度和流通量。

为了使几台 μC 能并行地协同执行一个任务的不同部分，要求通讯速度快、方便。为此，我们采用将“源 μC ”的总线周期（DATI、DATIO、DATIOB、DATO、DATOB）[25] 并入“目标 μC ”的总线周

期，使两总线勾通变为由“源 μC ”控制的单一总线一样来完成其总线周期。这样一种深入处理机内部的紧耦合，在总线周期级的通讯方式，可能是最快速的，且亦简单，性能效率高。

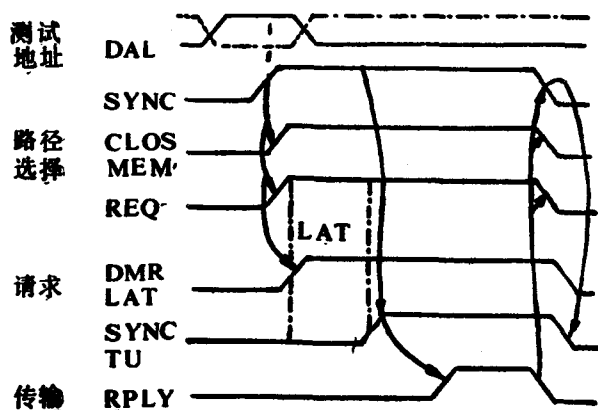
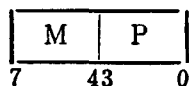


图 1.4 通讯的基本过程

如图1.11所示，没有总的公用存贮器，采用分散方式，每一个 μC 都带有局部存贮器。对于地址的变换，采用增设“通讯控制字”（CCW）的办法，其格式如下：



CCW 是可编程序的。

μC 之间进行一次通讯的基本过程如图1.4所示。机间通讯开销的主要部分是“源 μC ”

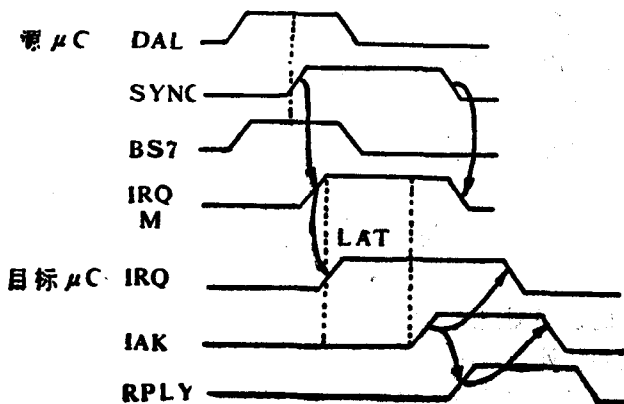


图 1.5 中断过程的波形

请求的等待时间，这与算法及程序的局部化原则有关。

很显然，分散的局部存贮器构成了一种共享内存，使 WuPP-80 具有协同解算较大应用题的可能。

从系统的分散控制、能分布处理出发，考虑了 μC 之间的中断功能，其主要过程如图1.5

所示。

2. “竞争”、“死锁”与“互斥”等问题的解决办法

采用随机扫描方法，使系统的成员都有相等的机会得到“提出了申请”的响应，不会出现“轮不上”的现象。

由于“总线超时自陷”能释放总线主控权，这使相互对访的“死锁”现象可以解开。

在总线周期中，有一种“读—修改—写”的形式，其执行过程不允许被其他任何中断源切断，这可用于实现“互斥”，比 C_m^* 所采用的封锁段描述的方法更方便，效率更高。

3. 软件结构

为了充分利用原有单个 μC 的软件资源，而又能适应 WuPP-80 系统的工作要求，为此配置了“分布式并行操作系统 (DPOS)”，扩充修改了原有高级语言，适于并行计算的汇编语言、FORTRAN 语言。

基于目前的条件，研制分布式并行多微处理机操作系统 DPOS 的目标旨在为用户提供一个单用户多任务级高度并行的系统，以期在 WuPP-80 系统上实现高速解算尽可能多的同步、异步并行算法的课题。

在 WuPP-80 中，任一 μC 里都存放着 DPOS 的同一副本；任何一台 μC 既可在自身 RT-11OS 的监督下工作，又可通过 DPKNL 协同操作。为了实现多机多任务间通讯，在 DPOS 中提供了三种方式：

- 带进出口参数的派生、终结原语；
- 实现同步、异步通讯的高级通讯原语；
- 直接发送原语。

同时还考虑了多机多任务的调度及资源的管理问题。

在 FORTRAN 语言中，增加了 FORK (分叉)，JOIN 或 JOINS (汇合)，BTASK (派生) 等并行语句，以解决并行程序段的派生、通讯及资源共享等问题。

这样就为我们联合“小机”解算“大型”课题提供了软件环境。

四、结 束 语

有关 WuPP-80 的性能测试还要作大量的研究工作，如何提供一种高效率的、对用户方便的通用的“分布式并行处理系统”，还有许多方面值得进一步调查研究。我们计划考虑模块性较好，互连方式灵活，容错能力较强，可靠性高，易于适应多种应用系统的新体系。

作者：彭德纯 朱冠华 张帮华

参 考 文 献

- [1] S. H. Fuller, et al, "Multi-Microprocessors: An Overview and working Example," Proc. of the IEEE, Feb. 1978, pp. 216~228.
- [2] B. H. Liebowitz, "多处理机方式的小型计算机系统, 第一部分: 设计原理", 计算机设计, 1978, 10, Vol. 17, No. 10(中译本)。
- [3] J. E. Wirshing, "Computer of 1980's—Is It a Network of Microcomputers?" Digest of Papers, Compcon Fall, 1975, pp. 23~26.
- [4] 金兰等, "并行处理计算机结构", 国防工业出版社, 1982.
- [5] F. G. Carty and R. H. Ries, "Goodyear Aerospace Corporation's Microcomputer Array Processor System," Proc. 1980 Int'l Conf. Parallel Processing, Aug. 1980, pp. 140~142.
- [6] J. K. Ousterhout, et al., "Medusa: An Experiment in Distributed Operating System Structure," Comm. ACM, Vol. 22, No. 2, Feb. 1980, pp. 92~105.
- [7] A. Plas, et al., "LAU System Architecture: a Parallel Data-driven Processor Based on Single-assignment," 1976 International Conf. on Parallel Processing, Aug. 1976, pp. 293~302.
- [8] 白英彩, "多微处理机系统概论", 中小型计算机, 第一期, 1981.
- [9] D. J. Farber, and C. L. Kenneth, "The Structure of a Distributed Processing System—Software," Proc. Symposium on Computer Communications Networks and Teletraffic, April 1972.
- [10] D. J. Farber, et al, "The Distributed Computer System," Proc. The Annual IEEE Computer Society International Conf., Feb. 1973.
- [11] "中美计算机讨论会介绍", 计算机学报, Vol. 5, No. 3 (1982, 5), p. 239—240.
- [12] G. Popek, "Notes on Distributed Systems of Microprocessors," in Microcomputer System Design: An Advanced Course, Vol. 126 of Lecture Notes in Computer Science, Springer-Verlag (1981), pp. 303~348.
- [13] H. Lorin, "Aspects of Distributed Computer Systems," John Wiley & Sons, New York, July 1980.
- [14] D. Aspinall & E. L. Dagless, "Overview of a Multi-Microprocessor Development Environment," UMIST, Microprocessors and Microsystems, Vol. 3, No. 7, Sept. 1979.
- [15] D. J. Kuck, "Parallel Processing of Ordinary Programs," Advances in Comp., Vol. 15, 1976, pp. 119—179.
- [16] D. I. Moldovan, "A Multi-Microcomputer Architecture for an Iterative Algorithm," Proc. of 1980 Inter. Conf. on Parallel Processing, Aug. 1980, pp. 155~156.
- [17] 康立山, "一类新型的异步并行算法", 见本书。
- [18] Kang Lishan, "The Schwarz Algorithm," WUHAN UNIVERSITY JOURNAL, Natural Science Edition Math., No. 1, 1981.
- [19] 康立山等, "Schwarz 算法及其在分布式并行处理机上的应用(I)——解线性椭圆型边值问题," 见本书。

[20] 康立山等, "Schwarz 算法及其在分布式并行处理机上的应用 (I) ——解弱非线性椭圆型边值问题," 见本书。

[21] P. H. Hibbard and N. S. Ostlund, "Numerical Computation on Cm*," Proc. of the 1980 Inter. Conf. on Parallel Processing, Aug. 1980, pp. 135~136.

[22] R. J. A. Buhr, et al., "Why Multiple Microprocessors?" MIMI 77, Nov. —Montreal, pp. 283~287.

[23] A. D. Hirschman 等, "标准模块为多处理机系统设计提供了灵活性", 计算机设计 (中译本), 第 5 期, 1979.

[24] G. A. Anderson, et al., "Computer Interconnection Structures, Taxonomy, Characteristic, and Examples," ACM-Computing Survey, Vol. 7, No. 4, Dec. 1975, pp. 197~213.

[25] DEC, "Microcomputer Processors," Published by DEC, 1978—79.

[26] S. E. Madnick and J. J. Donovan, "Operating System," McGraw-Hill Co. 1974.

[27] P. H. Enslow, "Multiprocessors and Parallel Processing," John Wiley & Sons, 1974.

分布并行算法和分布式并行处理系统研究*

分布式计算机从七十年代初期开始兴起，到了后期就逐渐形成了一门发展迅速且很有前途的新兴分支，并成为当代计算机科学家和工业界所热衷研究的领域。

大量的研究实例[1]、[2]、[3]、[4]表明了“分布式系统有高的性能价格比，灵活的系统可扩充性，良好的实时性、可靠性与容错性”等优点。但是，如何把多台微型计算机，构成一种“分布式并行处理系统”，扩充存贮空间，加速解算一般复杂大型课题等却是有争议的，且有许多难关。不过就其结构本身的“分散性”，每个子系统的“自治”能力，蕴含着并行处理许多复杂的应用问题所固有的并行性特征的能力，很有研究意义。因为至今几乎所有的计算机系统设计和应用有着密切的联系，所以，我们从探索某些类型的实际问题入手，构造一个实用系统 WuPP-80，以期为分布式并行算法和系统软件的研究提供一种较理想的环境，亦即在现有条件下，尽量探求一种高性能/价格比的互连结构，使之既可分布处理，又能快速并行处理，适于多方面的应用。本文将从算法、结构及其软件进行综合性论述。

一、分布并行算法

大家都知道，并行处理是提高计算机的效率，并加快对应用问题解的一种非常重要的方法。把分布式系统与并行处理相结合，是我们研究的重点。

分布式计算机系统结构的一个基本特征是：软硬件结构的“分散性”及其成员——子系统的“自治”能力。这样一种结构使以往的串行算法，乃至并行算法均不能使之有效地运行。为此必须修改旧的串行、并行算法，或提出新的分布并行算法。

在数值计算中，文献[5]基于区域分裂法——Schwarz 交替法的推广，“证明了解一类偏微分方程及优化问题的 Schwarz 算法的收敛性与松弛的顺序无关，而敛速与之有关”，由此而明确提出了三个任意性：“分块的任意性、松弛因子的任意性及松弛顺序的任意性。”这为“一类新的异步并行算法”的研究奠定了理论基础。

在数值模拟实验中，[6]采用步长为 h 的5点差分格式解二维 Poisson 方程的 Dirichlet 问题：

$$\begin{cases} \Delta u = 2(3x + x^2 + y^2) & \text{于 } \Omega \text{ 内, } \Omega = \{0 < x, y < 2\} \\ u|_{\Gamma} = x^2(x + y^2) & \text{其中 } \Gamma \text{ 为 } \Omega \text{ 的边界} \end{cases}$$

为适应分布式系统处理，把区域 Ω 分成如图1.6(a)或(b)，按照模拟程序方法进行计算（参看图1.7）。实际计算表明，无论怎样分块，无论按那样的顺序由逐线松弛法计算，结果都是收敛的。

* 本文刊于“人工智能学报”，1983年。此处略加修改。

在 $\epsilon = 10^{-7}$ 控制下, 用一台计算机来计算, 对 Seidel 迭代法, 需迭代 265 次; 对 Schwarz 算法, 需迭代 144 次; 把区域 Ω 分成 4 块 (图 1.6(b)) 模拟 4 台计算机同时计算 (亦即每台计算机算一块, 所谓“程序的局部化”), 而相应的拟边界点值由其相邻块进行交换 (亦可随机交换), 结果需迭代 155 次。虽然迭代次数有所增加, 但这时每台计算机需计算的点却减少了四分之三, 而计算机之间只增加了拟边界点值的交换, 且不需要“等待”新的值, 仍能继续迭代 (所谓“异步通讯”), 甚至于可以在不同的区域上, 采用不同的离散化方法及相应的不同解法。这与一般多处理机系统, 用网格法把所有网格点平均分配到若干台处理机上并行计算是不同的 [7]、[8]。由此看来, 用分布式并行计算机解这类题, 可使解题的时间大大缩短 (注意图 1.6(a) 的分划使交换的信息量和变量比至少为 1:1, 其分划不太理想)。

类似地对于“弱非线性椭圆型边值问题”、“二维定常的 Navier-Stokes 方程”、及“奇异性椭圆型边值问题”等亦进行了模拟计算, 同样获得了满意的结果 (参看第三部分)。这样就从理论到实践导致了“一类新型的异步并行算法”的产生, 它既适用于解算一类最优化问题, 也适用于解算数学物理问题。这类算法不仅可用于串行式计算机, 也适用于并行式计算机系统, 更适于分布式并行处理系统。应用在 WuPP-80 系统上获得成功, 充分发挥并验证了这类系统在并行计算上的优越性。因此, 它的发展将直接影响“异步并行处理系统”的发展。

在定理证明中, 对于某些典型问题亦提出了适于分布并行处理的算法。

下面用“代德景问题”中, 求 $[1, 1, 7, n]$ 的表达式作为例子, 说明用怎样的算法可以在 WuPP-80 上把表达式求出来。

首先从定义出发, 规定什么是 n 向图。

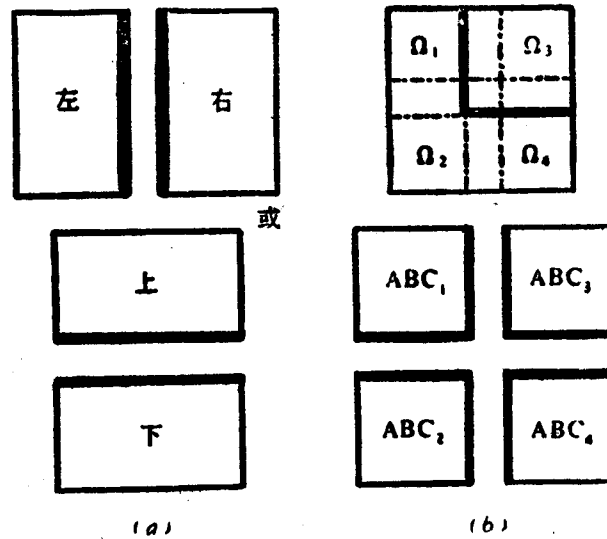


图 1.6 区域分划法

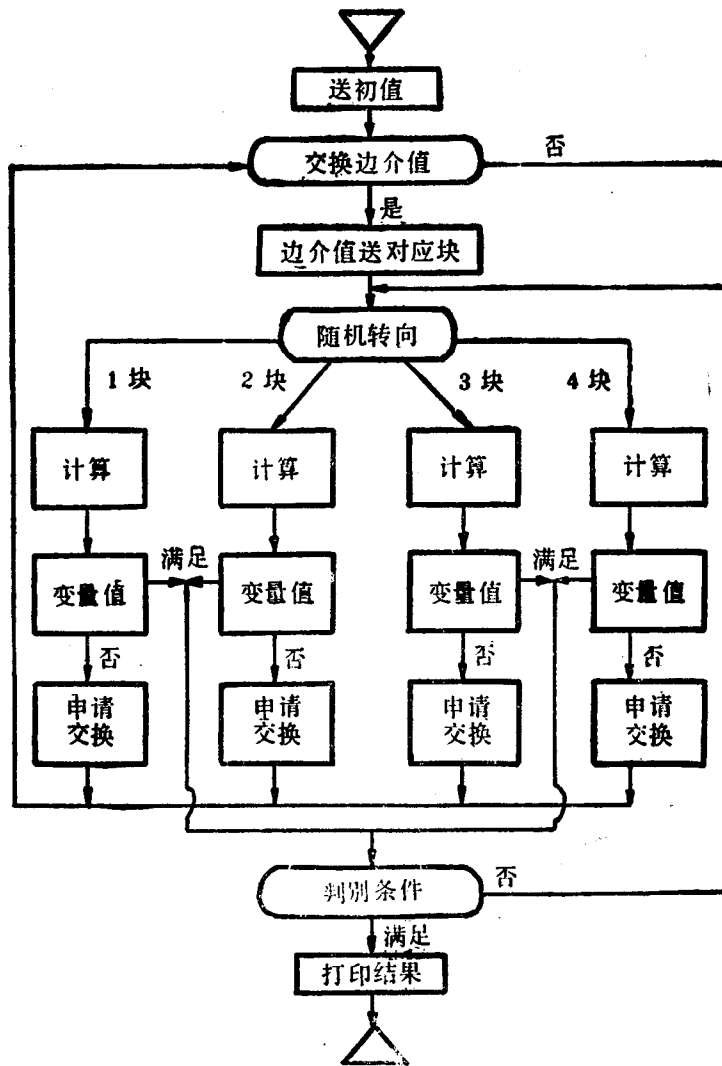


图 1.7 模拟程序框图

一般用下列三个步骤，从 k 向图可以构成 $k+1$ 向图（图1.8是2向图，生成3向图的过程）。

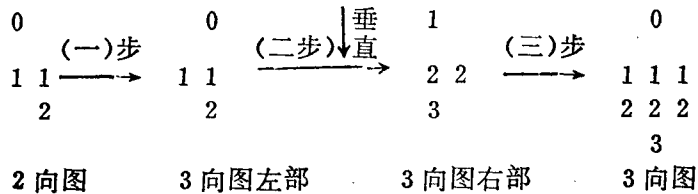


图 1.8 n 向图的递归生成法

(一)假定 k 向图($k=0, 1, 2, \dots$)中数字0不动；其它数字向左平行移动，使各水平线右端数字排在一条垂直线上。这样得到的图形，就叫做 $k+1$ 向图的左部。