DOI:10.16262/j.cnki.1000-8217.1995.03.003

并行算法的发展及其前沿研究课题

李晓梅

(国防科技大学，长沙 410073)

[摘要] 介绍并行算法的产生及其重要意义；阐述并行算法的基本概念、分类和发展情况；综合提出并行算法研究内容、研究层次以及当前的前沿研究课题。

[关键词] 并行计算机，并行算法，并行算法复杂性，并行计算模型

1 并行算法的产生和它的重要意义

并行算法的产生可归因于两个方面：一方面并行计算机的迅猛发展，促进了并行算法的产生与发展。并行计算机从70年代初诞生以来已经历了从分布主存SIMD阵列机(如1972年的ILLIAC TV)发展为向量机（如1978年的Cray-1)，又从向量机发展到紧耦合共享存储的并行处理机系统（如1983年的Cray X-MP)，再发展到松耦合分布存储的大规模并行处理机系统(如1993的Cray T-3D)。并行机的发展，不仅促使人们对已有的传统数值和非数值计算方法进行并行性改造，同时也促使人们从实际问题的物理模型、数学模型、计算方法到算法的程序实现各层次上进行并行性研究，研究新的并行算法，使之适合于蓬勃发展的并行计算机体系结构。另一方面，由于科学技术高速发展，许多领域，尤其是高新技术领域中问题的解决，特别需要更高性能的并行机和更先进的算法给以支持。有人作了这样一个实验比较：对某一规模为n的问题，设计了5种算法A_1，A_2，A_3，A_4和A_5，它们求解该问题所需时间$T(n)$分别为n，$n\log_2 n$，n^2，n^3和2^n。假设它们在单位时间内所能处理的数据量分别为S_1，S_2，S_3，S_4和S_5，现要考查计算机的速度提高10倍、1万倍后，上述5种算法的数据处理能力变化。通过简单的运算后可得出表1。表1说明：当计算机速度提高10倍，1万倍后，算法A_1和A_2处理数据能力也差不多提高10倍，1万倍，算法A_3和A_4提高甚少，最差是A_5，只增加3个或13个数据量。也就是说，如果采用A_5，一台高速计算机在一分钟之内处理的输入量用一台只有它的10^{-4}低速计算机来处理，不到两分钟就处理完毕，可见设计先进算法是多么的重要。

为有效地使用并行计算机，发挥其最高效率，就必然要有一套与之相适应的并行算法支持。并行算法对发挥并行计算机效率起着至关重要的作用。下面我们介绍一个简单的例子：用于石油地震数据处理的褶积计算

$$C_j = \sum_{i=1}^{NB} B_j \cdot A_{j+i+1}, \qquad j = 1,2,\cdots,Nc \tag{1}$$

我们针对公式（1）设计了“纵向”并行计算方法，在YH-1向量机上进行了数值计算，结果

本文于1994年11月28日收到.

表明，当 $NB \geqslant 100$，$NC \geqslant 100$ 时，该问题的向量加速比（公式（1）的串行计算时间与向量计算时间之比值）可提高 100 倍。这说明，如果给一台每秒 1 亿次的计算机，对公式作串行计算，则一台每秒 1 亿次计算机就变成为一台每秒百万次的计算机了。从而并行机的效率得不到真正发挥。

表 1 五种算法数据处理能力变化情况

算法	$T(n)$	计算机速度提高前单位时间内所能处理的数据量	计算机速度提高 10 倍后单位时间内所能处理的数据量	计算机速度提高 1 万倍后，单位时间内所能处理的数据量
A_1	n	S_1	$10S_1$	10^4S_1
A_2	$n\log_2 n$	S_2	S_2 很大时接近 $10S_2$	$S_2 \geqslant \log_2 9000$，超过 $9000S_2$
A_3	n^2	S_3	$3.6S_3$	$100S_3$
A_4	n^3	S_4	$2.15S_4$	$21.54S_4$
A_5	2^n	S_5	$S_5+3.3$	$S_5+13.32$

对大型科学和工程计算，以及大型数据处理问题，研究并行算法，提高求解问题的计算速度显得更为重要。我们在 YH-2 紧耦合共享存储多处理机系统上对 14 个实际应用问题设计了并行算法，并进行了数值计算，结果如表 2 所示。分析其中的用于石油地震数据处理的“测网加密程序”，其单机向量加速比为 3.22，四机加速比（单机向量计算时间与四机向量计算时间之比值）为 3.27。因此四机向量计算速度比单机串行计算速度快 10 倍多。这表明在单机上对“测网加密程序”进行串行计算，需要 10 小时完成，而在四机向量上对其作并行计算，则只要 1 小时即可完成计算。这不仅节省人力，也产生了可观的经济效益。

表 2 在 YH-2 单机和四机处理系统上对 14 个大型计算题的测试结果

序号	题目名称	YH-2（单机）与 YH-2（四机）加速比
1	多极扩散室的消波率数值模拟	3.98
2	相对论电子束数值模拟	3.99
3	回归模型的自变量选择	3.87
4	电磁模相对论性自由电子激光的粒子模拟	3.42
5	二维平面爆轰程序数值模拟	2.31
6	对称区域法解高维 poisson 问题	3.43
7	倾角动校正计算	3.94
8	分子链表程序数值计算	3.23
9	不变嵌入法求解积分方程本征值问题	3.99
10	测网加密程序计算	3.27
11	中高能重离子碰撞的半经典蒙特卡洛模拟	3.32
12	能带论 LMTO 方法数值计算	3.17
13	载人飞船与机动弹头喷流干扰流场数值模拟	3.65
14	中期数值天气预报 $T_{63}L_{16}$ 数值模拟	3.54

2 半行算法的基本概念与分类[1,2]

我们知道，并行计算机是按某种方式互相连系的多机系统，这些处理机能够独立地执行各自操作，处理各自的数据，因此，同一时刻可按一条或多条指令处理多个数据。这一特征是设计并行算法的基本依据。

并行算法是适合于在并行计算机上求解的一类算法。它是 K 个并发进程集合，这些进程

可以同时执行，并且相互作用和协调工作，以达到对给定问题求解。当 $K=1$ 时就是传统的串行算法。例如，计算 8 个数求和，即 $S=x_1+x_2+x_3+x_4+x_5+x_6+x_7+x_8$。用串行算法需 7 步（加法），如果有一台由 4 个处理机组成的并行计算机，则计算 S 的一个并行算法仅需 3 步，其中：

第一步计算：$u_1=x_1+x_2$，$u_2=x_3+x_4$，$u_3=x_5+x_6$，$u_4=x_7+x_8$；

第二步计算：$r_1=u_1+u_2$，$r_2=u_3+u_4$；

第三步计算：$S=r_1+r_2$

它是由 4 个并发进程集合：进程 1 计算 u_1，r_1 和 S；进程 2 计算 u_2；进程 3 计算 u_3 和 r_2；进程 4 计算 u_4。其中进程 1 在步 2 需从进程 2 获得信息 u_2，在步 3 需从进程 3 获得信息 r_2；进程 3 在步 2 需从进程 4 获得信息 u_4。这就是进程之间的相互作用和协调工作以达到对 8 个数的求和。

一般情况下，并行算法可分为数值并行算法和非数值并行算法两大类。所谓数值计算（Numerical Computation）是基于代数关系运算的一类计算问题，诸如矩阵计算、线性方程组求解等，它基本上是属于数值分析范畴。而非数值计算（Non-numerical Computation）则属于关系运算一类问题，诸如排序、选择等，基本上是属于符号处理范畴。因此，研究数值和非数值问题的并行算法分别称为数值并行算法和非数值并行算法。

在数值和非数值二类并行算法上，每一类又有同步并行算法和异步并行算法之分。所谓同步并行算法（Synchronized parallel Algorithm）是指算法的诸进程集合中，某些进程中的若干操作要等待另一些进程中的某些操作执行之后才能执行的算法。由于进程的执行要依赖于信息交换和数据输入，因此，导致所有进程需要同步在一个给定时钟上，以等待较慢进程。所谓异步并行算法（Asynchronized parallel Algorithm）是指诸进程集合中的所有进程间无需等待，进程间的信息交换或通过动态地读公用存储器中更新的数据来实现，或随机地相互使用对方送来的信息。这里的“动态地”与“随机地”是强调进程既不等待其它进程更新数据，也不等待传送信息，因此，进程不会出现暂时的封锁，但对存储器的并行存取会发生冲突，使某些进程在存取公用变量时有小的延迟。

并行算法好坏，主要通过加速比 S_p 和效率 E_p 来度量。即 $S_p=T_1/T_p$ 和 $E_p=S_p/p$，其中 T_1 表示求解某一问题的最优串行算法在单机上的执行时间，T_p 表示求解同一问题的并行算法在 p 台处理机上的执行时间。由于 $T_p\leqslant T_1$，$S_p\leqslant p$，所以 $S_p\geqslant 1$，而 $E_p\leqslant 1$。

3　并行算法的发展及其研究内容[1,2]

进入 20 世纪 60 年代，人们已开始研制新型结构计算机——阵列式结构和流水线结构。它们是为适应当时大型科学计算的需要，特别是解偏微分方程的需要。随着新型计算机结构研制，人们开始了并行算法的研究，不过，这时期并行算法研究是建立在公理化假设的理想化并行计算模型上，即假设处理机台数无限且相同；有任意大的存储器且能同时为所有处理机存取；四则运算均花同一单位时间，且存、取、同步和通信不计时间。因此，我们称这一时期研制的并行算法为理想化同步并行算法，且以当时研制的 SIMD 机为背景。

进入 20 世纪 70 年代后，由于并行计算机问世，并行算法研究迅速发展。由于这时期并行计算机的结构特点以及向量运算具有内在并行性，并行算法研究首先在数值线性代数方面

取得了较为丰富成果。然而与串行算法不同的是，并行算法与并行计算机的体系结构密切相关，对于同一问题，依据不同体系结构可以设计出不同的算法。因为这一时期投入使用的大多是阵列机和流水线机，故这一时期主要研究同步并行算法，特别是向量机上并行算法研究，且基于 PRAM 并行计算模型及其扩充模型进行算法研究，同时使算法逐步软件化。

进入 20 世纪 80 年代，一方面紧耦合共享存储多处理机系统研制成功并投入使用；另一方面并行计算机的研制向着 MPP（Massively parallel processing）机，即大规模并行处理机方向发展。与此同时，出现机群系统并投入使用。因此，这一时期同步并行算法成熟，并软件化；异步并行算法研究非常活跃，特别是大规模异步并行算法应运而生。在数值计算领域，一是对原有串行算法进行并行化改造，二是根据分而治之和重新排序两个基本原则设计新的并行算法，而且对算法复杂性分析和性能评估作了重新考虑。

进入 90 年代，并行算法研究领域由大型科学和工程计算以及大规模数据处理，扩展到了智能计算和神经计算等广泛的领域，研究方法有了新的突破，涌现出一些全新算法。这其中最主要的流派是回归大自然，对问题重新建模或借鉴与学习大自然的演化规律来设计大规模并行算法。例如，基于细胞自动机（Cellulor Automaton）理论模拟流体力学的格子气方法（Lattice gas method）就是对流体流动问题的重新建模。它根据分子运动论的观点，从流体的微观结构出发建立完全并行的流体微观运动模型——格子气自动机，由格子气自动机的并行运行和统计平均来模拟流体的宏观运动。

并行算法的研究内容可归结为两个方面：

（1）数值计算。主要研究相关问题计算、矩阵与多项式运算、线性与非线性方程组求解、矩阵特征值问题计算、非线性方程求根、微分方程数值解、离散变换、最优化问题计算、素性测试与大整数分解等基本数学问题的并行算法研究，并将其应用于各领域科学与工程计算和大型数据处理中。

（2）非数值计算。主要研究分类、排序、选择、搜索、图论计算、图象分析、优化判定、组合遍历、语音识别、电路设计、并行数据库、计算机辅助设计、智能计算和神经计算等各种非数值问题的并行算法研究，并应用于各种实际问题中。

由于并行计算机有紧耦合共享存储多机体系结构、松耦合分布存储多机体系结构和群机系统之分，因此，就存在这些体系结构上的各种数值和非数值并行算法研究。对于前一种体系结构，影响并行算法效率的关键问题是：（1）算法进程划分与生成；（2）任务分配与负载平衡；（3）进程调度、通信与同步关系，算法加速比与问题规模关系，算法加速比与算法中串行部分所占比例关系。对于后两种体系结构，影响算法效率的关键问题是：（1）数据输入组织与数据分割；（2）平衡的计算调度和平衡的通信调度；（3）计算与通信的重叠执行；（4）如何减少远程调度，如何提高 Cache 命中率和减少程序中的伪共享现象。对群机系统要尽量减少处理机间通信量。

并行算法的研究层次是：

（1）并行算法理论研究，包括计算问题可并行性研究；不同计算模型的能力、限制及它们等价关系研究；各种因素对并行算法加速比的影响，特别是负载平衡问题研究；并行算法复杂性下界研究；并行算法稳定性及其收敛性研究等。

（2）并行算法的设计，包括并行计算模型研究（如 PRAM 模型及其改进模型、$\log P$ 模型、

分布共享程序设计模型等等)；并行算法设计技术（如分而治之技术、划分技术、流水线技术以及加速级连接技术等)；并行算法性能分析。

(3) 并行算法的实现，包括并行算法到体系结构的映射，并行程序语言的优化以及并行算法的运行环境与工具的研究等。

4 并行算法前沿研究课题[3-5]

1993 年，美国 David 和 Culler 等人分析了当前各种并行计算机系统的特点，根据 VLSI 技术和网络技术的发展，看到了今后新的并行计算机的发展趋势，即 90 年代和未来的并行计算机将由上千个结点机通过高性能的互连网络连结。每个结点机由高性能通用微处理器、高速缓存和容量有限的动态随机存储器（DRAM）组成，结点机间通过互联网络以消息传递方式相互通信。

与此同时，近一二年来，随着工作站的性能迅速提高和价格下降，一种新的并行处理系统应运而生，这就是工作站群机（Workstation cluster）系统。它是通过以太网或 FDDI 将一些工作站相互联结，并配以相应的支撑软件，主要是消息传递程序库（Message Passing Library）和并行程序开发环境等（如 PVM，Express)，它具有很好的性能价格比，现已成为并行处理的热门话题。

另外，近年来国内外掀起了一股人工神经网络的研究热潮。利用机器模仿人类的智能是长期以来人们认识自然、改造自然和认识自身的理想。人工神经网络以其独特的结构和处理信息的方式，使其在许多实际应用领域中取得了显著的成效，能够解决一些传统计算机难以求解的问题。

综上所述，当前并行算法前沿研究课题是：

(1) 大规模并行处理机上的并行算法研究：MPP 并行计算模型研究；MPP 并行计算模型上并行算法研究；并行算法到实际并行计算机的映射；MPP 机上并行算法效率评价准则研究，特别要研究影响并行算法效率的关键问题；MPP 机上并行语言和编程环境研制。

(2) 工作站群机系统上并行算法研究：对于工作站群机系统上并行算法研究，我们仍然需要研究工作站群机系统的并行计算模型及其上面的并行算法设计；并行算法到实际工作站群机系统的映射以及并行算法效率评价准则的研究；消息传递程序库和并行程序开发环境研究。

通常工作站群机系统计算速度很高，但与 MPP 系统比较，首先是群机系统处理机间的通信速度很慢，这是由其网络延迟（Latency）高、传输速率低所决定的。如果采用的是 Ethernet 网络型群机，则网络延迟几乎是典型 MPP 机的 30 倍，而传输速率就更低了。其次，工作站群机系统是异构型的，面向许多不同种类的机型；第三，工作群机系统容错能力差，因此，在这种系统上进行算法设计时，要特别注意数据组织上数据划分，并尽量采用动态数据调度，以使处理机间达到负载平衡，把处理机间的通信量减少到最低限度，同时，算法设计要考虑自身的容错能力。

(3) 基于神经网模型并行算法研究：神经计算复杂度理论研究，主要评价神经计算能力，研究神经网能否解决常规计算机无法解决的问题；神经网并行计算模型研究及其上并行算法设计、算法到实际神经网的映射；组合优化问题的神经网络并行算法研究，特别是模拟退火并行算法和均场退火并行算法研究；基于神经网络的并行学习和训练算法研究，特别要研究

神经元的划分和并行度开拓，并行学习算法设计与实现；遗传算法研究，研究遗传算法的性能分析，遗传算法的并行实现，遗传算法在神经网络中应用以及遗传算法用于机器学习中程序设计等。

参 考 文 献

[1] 李晓梅，蒋增荣. 并行算法. 长沙：湖南科技出版社，1992 年 9 月.
[2] 陈国良. 并行算法设计与分析. 北京：高等教育出版社，1994 年 5 月.
[3] 刘勇. 遗传算法的理论与应用 [学位论文]：武汉大学，1994 年.
[4] 戴葵. 通用高速可变结构并行神经计算机系统的研究与实现 [学位论文]：长沙，国防科技大学，1994 年.
[5] 黄晓安，郗春明. 并行处理与工作站群机系统，计算机世界报，1994 年 3 月.

THE DEVELOPMENT OF PARALLEL ALGORITHM AND SOME QUESTIONS TO BE RESEARCHED

Li Xiaomei
(*National University of Defence Technology, Changsha* 410073)

Abstract In this paper, the development of parallel algorithms and its importance are introduced. The basic concept, the sorting and the development of parallel algorithms are stated. The problems and the levels to be researched and some questions of present forntier research for parallel algorithms are given.

Key words parallel computer, parallel algorithm, complicated parallel algorithm, a model of parallel computing