

并行计算方法研究与并行计算机

孙 家 昶
(中国科学院计算中心)

【摘要】 并行处理是当今世界的关键技术之一。本文论述我国发展并行处理与研究并行算法的意义,以及二者之间的关系,介绍国家自然科学基金重大项目“并行计算机与并行算法”中有关并行算法研究一年多来取得的进展,并对今后工作提出建议。

引 言

我国从70年代初开始研制向量机及相应的向量算法。80年代,一些部门开始从事并行计算方法的研究。1989年6月,国家自然科学基金委员会把“并行计算机与并行算法”列为重大项目,以进一步推动我国并行处理关键技术的发展。该重大项目由冯康教授(学部委员、中国科学院计算中心)及夏培肃研究员(中国科学院计算技术研究所)共同主持,复旦大学计算中心、中国科学技术大学及东南大学计算机系分别承担自动并行识别,并行FORTRAN语言及并行操作系统的研制。中国科学院计算所与计算中心分别承担并行机硬件模型机与并行算法研究。

本文主要介绍并行算法研究课题的进展,探讨我国开展并行计算方法研究的若干问题,并对今后工作提出一些建议。

一、并行处理与并行算法研究的意义

并行处理是当今计算技术的前沿领域,很多国家都把发展并行处理列入关键技术的前列。这是因为,尽管计算机的单机速度近年来有了飞速发展,但仍满足不了某些大型科学与工程计算的需要。例如,要提高全球性气象预报的准确性,需要在经度、纬度及大气层方向取网格点($200 \times 100 \times 20$)40万个,要求在2—3小时内完成48小时的预报计算;又如,用计算机模拟电子结构,要计算上千个变量的积分-微分方程,成百万个六重积分,并求解阶高达 10^7 矩阵的本征值;再如,模拟人体心脏跳动,需要求解三维粘性不可压流体方程,目前国际上完成的粗糙模拟一次心脏收缩,就需花超级计算机Cray-II 8个CPU小时。以上这些大型科学与工程计算所要求的计算能力比当前计算机的实际能力高出几个数量级。

同时,还因为单机速度的发展在理论上受着不能超过光速以及物理尺寸不能无限缩小的双重本质限制。事实上,主机速度的增长速度已趋于下降。据统计,1976年以前,主机速度平均每7年提高一个数量级,而从1976年Cray-I(主频周期12.5毫微秒)发展到以后的Cray-II(4.1毫微秒),浮点运算速度从每秒1.6亿次上升到5亿次,10年内只提高三倍。为此,选择并行处理成为极大提高计算机系统性能的必由之路。目前,并行处理机的台数尚没有理论

上的限制,带有 8.5 万个并行处理部件的并行机已经进入市场,带有 100 万个处理部件的并行机正在研制中。

所谓并行,是指一个以上事件在同一时刻或同一时间间隔内同时发生。并行处理的思想最早可追溯到上个世纪,只是由于以前的器件条件不成熟,才出现串行机长期垄断的局面。60 年代开始了计算机体系结构的并行性革新,例如把 I/O(输入/输出)从中央处理器(CPU)分开,建立缓存区;采用多用户分享资源的分时操作系统;采用处理时间重迭的流水线系统等。并行计算机的最初雏形是美国 Westinghouse 航空实验室于 1963 年用 9 个处理部件排成 3×3 阵列机,当时他们用它计算了一个偏微分方程。1972 年研制成功的 Illiac-IV 是一台 8×8 阵列机,这是并行机发展史上的里程碑。随后 CDC 公司研制向量机获得初步成功。向量机形成产业的标志是 1978 年 Cray-I 的诞生。以后,日本的富士通、日立与 NEC 等相继推出“亿次机”乃至“10 亿次机”产品,我国也设计或研制了 757 机、YH(银河)-1 以及 KJ8920 等向量机。另一方面,具有性能价格比优势的小型巨型机(向量并行机)开始陆续进入市场。据不完全统计,到 80 年代末,世界上已安装了 300 多台“亿次机”,1500 台各种并行机,其中绝大部分的结构属于共享存贮系统,这类并行机处理器的个数大多在 4—8 之间,少数达 16—32。处理器个数超过 100 的“高度并行”机目前都采用分布式系统结构,每个处理器自带局部存贮器,不同处理器之间的通讯通过信息传递。

历来,计算方法与所使用的计算工具是密切相关的,计算工具是计算方法实现的物质基础。如 50 年代以前的计算方法是基于台式计算机,近几十年来人们研制的大量应用程序和软件资源,是以串行机为主要对象。并行处理是计算机结构上的重要突破,并行机与传统的串行机在硬件结构上有着本质的区别,并行机为计算机用户实现算法提供了新的手段——并行化,从而促使计算方法的研究进入并行化计算方法的新阶段,这是一个很值得重视的动向。

同时,并行机的研制与发展本身也离不开并行算法研究,这种联系比以往串行机研制时要密切得多。经过长期发展,很多公司生产的串行机都已形成序列,软件兼容,算法研究主要用来开发新软件。而并行机的研制则不同,由于其使用和效率的发挥,强烈依赖于机器结构,使系统软件与算法研究在并行机研制中所占的比重大大超过串行机的研制。所以,很多并行机研制部门都设有自己的算法研究组,负责测试并行性能及研制适合该并行机的特定算法,或反过来检验结构设计的合理性,甚至反复修改结构设计。有的并行机就是根据某个特定算法而设计的。由此可见,研究并行算法是研制发展并行机的一个重要组成部分。

此外,我们之所以强调加强并行算法的研究和应用,是吸取了国际上的经验教训。80 年代中期,国际上出现并行机热,各种型号的亿次机及小型巨型机竞相推出。时隔几年,就出现了反复,不少小型巨型机包括某些巨型机产品积压,某些公司濒于倒闭。造成这一情况的原因很复杂,总的还是属于高技术迅速发展与市场激烈竞争的正常过程。近年来,基于精简指令芯片(RISC)技术的高性能工作站迅猛发展,友好的人机交互界面,功能齐全的真实感图形-图象系统,软件标准化与语言的可移植性,以及有竞争力的性能价格比,为用户提供了大量的应用软件与较理想的硬、软件环境,对于一般用户,基本上不用做算法与程序的更动,这些自然是很有吸引力的。相比之下,小型巨型机用的芯片较落后,又鉴于目前并行软件中自动并行识别(包括向量自动识别)功能不齐全,在多数情况下必须投入较多的人工干预才能很好发挥并行机或向量机效率。换句话说,目前并行机使用还停留在“专家级”水平,对于一般用户,不如使

用工作站方便,价格也比较昂贵。这些是使人们从“并行机热”转向“工作站热”的主要原因。虽然如此,并行机有其自身的特长和用户。对于大规模的科学工程计算,目前的工作站还不能替代并行机。更何况工作站发展的趋势之一是增加并行处理部件,最终仍是要走并行化的道路。而且,由于目前高速芯片以及并行软件的不成熟发展,并行机不景气的状况正在改变之中。

考虑到我国国情,我国发展并行机技术应从中吸取的经验教训是:1. 发展并行机技术不能一哄而上,尤其是在计算机技术飞速发展的今天,器件的更新淘汰周期很快,而我们的资金有限,对造型定型,特别是批量投产必须慎重。2. 最主要的是要加强并行软件和并行算法的预先研究。计算机的应用是计算机发展的主要推动力,不可重研制轻应用,重硬件轻软件。提高并行算法和并行软件研究的比重,是一条投资少、风险小、效率高的多快好省的技术路线。目前国际上很多并行机公司都表示有兴趣与我国合作研制并行软件,我国科技人员数学素质比较好,只要组织得当,就可以抓住这个有利时机,快速发展我国的并行机技术。

二、1990年的工作进展

“并行算法研究”课题组按照本项目学术领导小组的安排,参加项目组有关并行机整体方案的讨论,结合中国科学院的有关项目,1990年着重建立算法研究必要的硬软件环境,测试周围其它的并行机系统,同时开展算法的基础研究和应用研究,一年内完成研究论文16篇,技术报告13篇,在国内外学术刊物或学术会议上宣读论文7篇,这些成果都已收集在“1990年研究年报”中。具体如下:

1. 安装了由五片 T-800/25 组成的 Transputer 分布式并行试验系统,每片峰值为 20MIPS 和 2.0MPLOPS。每个处理器通过 4 条通信链与周围处理器连接,通道传输速度约为 1MB/秒。五片处理器由一片母板与一片子板组成,子板与母板相接,母板把 386 微机作为前端机。主要的软件环境有并行 FORTRAN,并行 C 和 OCCAM 语言,以及 TDS 环境。一年来,我们已经在这个并行试验系统上研制了一批并行程序,其中包括:

(1) 线性代数应用软件包 LINPACK 的 OCCAM 移植,总工作量约两万条 FORTRAN 行程序;

(2) 二维有限元的试验程序;

(3) 一维带激波流场的分区自适应等程序。

(4) 计算 π 精确到 100 万位,在五个处理器上并行计算了 102 个小时,这表明我们并行试验系统的计算能力已经达到国际上 80 年代初的先进水平。

2. 测试共享存贮并行机 Encore Multiman 520,这是国家智能中心引进的四个处理机并行系统,我们的工作主要是测试该系统安装的自动并行识别系统 EPF 的功能。测试结果表明,1988 年底的版本 EPF 对于程序结构较为简单的例子,如矩阵乘法、解线性代数方程组等典型问题,能达到近于线性加速比的并行效率,但对于结构较复杂的应用程序,该自动并行识别的效率不高,甚至出现有处理器愈多,加速比下降的反例。这从另一方面说明,为了充分发挥现有并行机的效率,需要在应用程序的并行算法研究上下功夫。

3. 探讨适用于向量机 KJ8920 的向量算法。KJ 是中国科学院研制的向量机,主频

50MHz, 主存容量 2M 字, 字长 64 位, 向量长度 64—128。测试的目的是发掘计算问题中的向量成分, 提高国产机的使用效率。通过测试表明, 对于简单的程序结构, 在循环语句中直接用向量语句代替标量语句, KJ 效率能提高 10—20 倍。而且, 随着问题规模的扩大, 效率的提高也愈明显。地震石油勘探软件是 KJ 机目前运行的重要应用软件之一, 我们对其中的某些典型模块开始向量化, 取得了初步成效。例如, 对其中一个迭加模块向量化后的速度比原来标量计算时提高 25 倍, 这表明, 现在 KJ 运行的一些应用软件向量化的潜力很大。

4. 并行计算方法的基础研究。并行算法是一门关键技术, 同时有学科上的新思想、新观念和新方法。我们把重点放在并行算法的基础研究和应用基础研究, 已对以下问题开展工作:

- 科学计算中的可并行性与并行潜力
- 分裂与排序在并行计算中的作用
- 并行算法中的舍入误差分析
- 分布式系统不同连接方式对线性代数方程组求解的影响
- 并行计算实现中的软件开销
- 共享内存与分布式系统的并行实现特点
- 向量计算与并行计算的共性及异性
- 有限元、计算流体、优化及几何造型中的并行计算, 等等。

世界上很多客观事物的发展过程是并行的, 彼此相对独立, 相互又有一定的联系和制约。并行算法研究就是要发掘客观过程的内在并行特性, 把一个复杂的计算任务分配到各个不同处理部件, 使每个子任务相对独立, 不同子任务之间减少不必要的通讯, 以协调实现整体的高效率。从这个意义上说, 并行计算有可能使人们对客观事物的认识较为深刻、较为全面。而以往串行计算机上计算方法的一个本质限制是把客观多维实体映象到一维串行机内逐条串行执行。逐步加深认识并行计算与串行计算的关系, 是并行算法基础研究的重要内容。

一个完整的科学工程计算过程, 要经历从形成科学工程模型、数学模型、计算方法, 到程序实现, 乃至软件开发诸环节。我们认为, “并行”的思想应贯穿在每个环节中, 并应尽量从“模型”级开展并行研究。只有这样, 才能更深刻揭示所计算问题客观存在的并行特性, 对于以后的并行算法设计和实现具有指导意义, 并往往能取得意想不到的效果。反之, 若置问题的物理意义与数学算法于不顾, 只是拘泥于现有串行程序的并行改写, 往往事倍功半。

并行机与并行算法的研究, 也推动了有关的计算方法发展。研究表明, 近十年来计算方法的很多重要进展直接或间接与并行算法有密切的联系。现正在逐步形成“并行计算方法”分支学科, 以示区别于原来基于串行机上的计算方法。

什么是“并行计算方法”的主要特点呢? 经过这几年的摸索, 我们初步归纳为“多、分、高、混、模”五个字。“多”, 即多层、多步、多级、多重等方法; “分”, 即分裂、分解、分区、分叉等方法; “高”, 指高精度、高维化等方法; “混”指混乱异步及随机方法; “模”, 指物理模拟方法。我们在这方面的研究成果应邀于 1991 年 5 月初在天津召开的全国计算数学学会会上作大会报告。

基于数学模型级并行计算方法的突出例子是区域分裂法, 这是一个重要的发展方向, 在 1990 年召开的第四届国际会议上被视为继有限元以来计算方法的重大进展之一。我们应邀在大会作了题为“四阶方程区域分裂预条件子”的学术报告, 应用张量积 B 样条方法构造出分裂预条件子, 并证明其条件数与步长细分无关。这是这方面国际上的最好结果, 受到与会国际

著名学者的重视和兴趣。

从数学模型级开始研究计算方法,通过大量的并行计算实验来考验和发展新的并行算法,有步骤地进行并行软件的研制和开发,是我们在研究中采用的主要技术路线。

三、若干体会与建议

1. 我国应抓住当前发展并行处理的好时机,统筹安排,减少不必要的同水平重复与人力物力上的浪费。由于我国的计算机硬件与国际水平的差距较大,而我国科技人员的数学素质相对较好,适当引进并行机硬件器件,加强对并行算法和并行系统软件的基础研究和应用研究,是很有必要的。

2. 并行机主要用途是多机实时控制,图形图象动态模拟,人工智能以及大规模科学工程计算等方面。并行机主要是为“超级计算”设计的(所谓超级计算是指每秒计算量达几亿次浮点运算,总计算量达 10^{12} 以上的计算问题),它及时、准确、安全、可靠”,并行处理的这些特点对于超级计算的意义是不言而喻的。然而,由于大规模并行的分布式并行计算机的使用目前还处于专家级水平,又缺少成熟配套的并行软件和调试工具,为了发挥并行机的效率,需要专门研究和设计新的算法。

3. 并行算法研究属于中长期科研课题,培养青年科技人才极为重要,建议在高校有关专业中增加有关教材。我们课题组中有一半是博士、硕士等青年人,他们接受新事物快,思想敏捷,在工作中发挥了重要作用。

4. 内外开放,并行协作。“并行计算方向”涉及到计算方法的每个方面,我们通过内外交流和协作,加快了研究进展。1990年我们分别派人参加了在美国和苏联召开的国际会议,邀请国外学者来华做学术报告。现正在探讨不久在我国召开一次“并行计算国际会议”的可行性。

ON THE RESEARCH OF PARALLEL COMPUTING METHODS AND PARALLEL COMPUTERS

Sun Jiachang

(Computing Center, Academia Sinica)

Abstract

Parallel procession is one of the key technique in the world today. In this paper, the author emphasize the significance in development of parallel procession and in research of parallel computing methods. The relationship between them is discussed. As one of the main parts of the major project by the NSFC, some progress and advances in the research of parallel algorithms are reported.