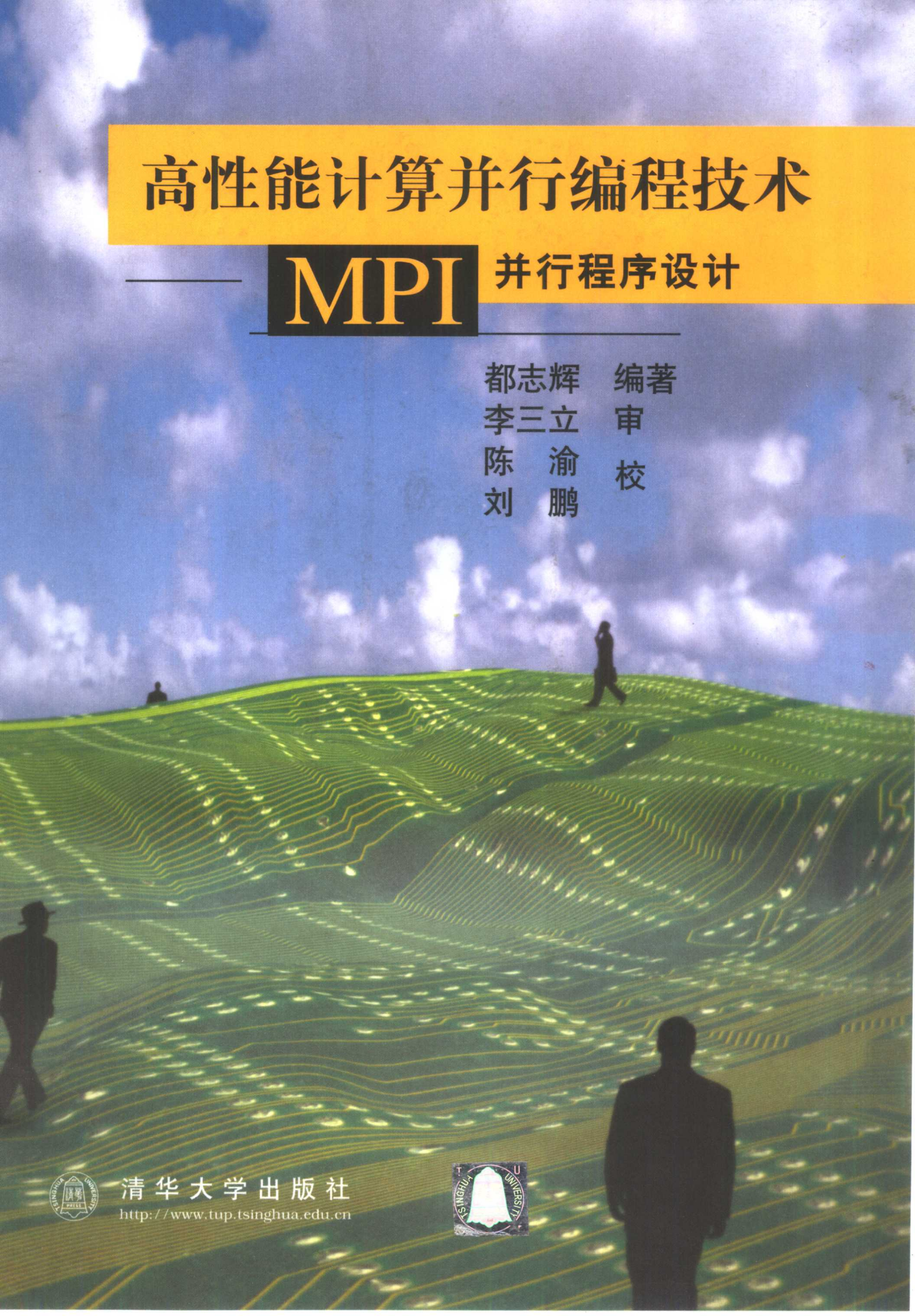
高性能计算并行编程技术

——MPI并行程序设计

都志辉 编著
李三立 审
陈 渝 校
刘 鹏



清华大学出版社
http://www.tup.tsinghua.edu.cn

高性能计算并行编程技术
——MPI 并行程序设计

都志辉　编著
李三立　审
陈渝　刘鹏　校

清华大学出版社

(京)新登字 158 号

内 容 简 介

本书介绍目前最常见的并行程序——MPI 并行程序设计方法,它适合高等院校计算机专业高年级本科生、非计算机专业研究生作为教材和教学参考书,也适合广大的并行计算(高性能计算)用户作为自学参考书。具有 FORTRAN 语言和 C 语言编程经验的人员都可以阅读并掌握本书的内容。

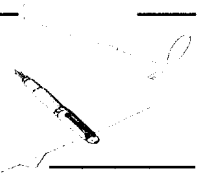
书中首先介绍了并行程序设计基础,提供给读者进行并行程序设计所需要的基本知识;然后介绍 MPI 的基本功能,从简单的例子入手,告诉读者 MPI 程序设计的基本过程和框架,这一部分是具有 C 或 FORTRAN 串行程序设计经验的人员很容易理解和接受的;接下来介绍 MPI 程序设计的高级特征,这是已经掌握了 MPI 基本程序设计的人员进一步编写简洁高效的 MPI 程序、使用各种高级和复杂的 MPI 功能所需要的;最后一部分介绍了 MPI 的最新发展和扩充 MPI-2,其中包括三个部分,动态进程管理、远程存储访问和并行文件读写。

本书包括了 MPI-1 的全部调用和 MPI-2 的关键扩充部分的调用,并附以大量的图表和示例性程序,对程序的关键部分给出了讲解或注释。读者若能将例子和对 MPI 调用的讲解结合起来学习,会取得更好的效果。

本书的目的,不仅是教给读者如何去编写从简单到复杂的 MPI 并行程序,更重要的是,希望读者通过本书的学习,在以后解决问题的过程中能够树立并行求解的概念,使并行方法真正成为广大应用人员和程序开发员手中的重要工具。

书　名:高性能计算并行编程技术——MPI 并行程序设计
作　者:都志辉　编著
出版者:清华大学出版社(北京清华大学学研大厦,邮编 100084)
http://www.tup.tsinghua.edu.cn
印刷者:北京广益印刷有限公司
发行者:新华书店总店北京发行所
开　本:787×1092　1/16　印张:22　字数:518 千字
版　次:2001 年 8 月第 1 版　2001 年 8 月第 1 次印刷
书　号:ISBN 7-302-04566-6/TP·2703
印　数:0001~6000
定　价:36.00 元

序



高性能计算技术在国内外受到高度的重视，它在科学研究、工程技术以及军事等方面的应用，已取得了巨大的成就。国际上科学家普遍认为，没有万亿次以上的高性能计算机，21世纪人类所面临的基因工程、全球气候准确预报、海洋环流循环等“巨大挑战”问题（Grand Challenge）是无法解决的。军事上的核爆炸模拟，也必须使用万亿次以上的高性能计算机，这也是高性能计算受到各国高层重视的一个重要原因。美国20世纪90年代有关高性能计算技术的研究规划，如HPCC（High Performance Computing and Communications）和ASCI（Accelerated Strategic Computing Initiative），都是在总统直接参与下制定的。在我国，高性能计算技术也受到各级领导部门的关注和重视。

从更广泛的意义上来看，“计算”，已经和“理论”与“实验”并列，被普遍认为是人类认识自然的三大支柱之一。这种计算，主要是指应用于科学与工程的、以高性能计算机为平台的大规模并行计算。它已发展成为一门新的学科——大规模科学与工程计算。大规模并行计算已成为研究科学与工程技术问题的一种崭新的手段和方式，采用这种手段和方式进行的科学与工程技术问题研究，被称为“计算科学与工程”（Computational Science and Engineering）。著名的波音777飞机的设计，基本上就是依靠高性能计算机的“无纸设计”，它可以大量节省传统设计技术中昂贵的风洞实验，同时大大缩短设计时间。

以高性能计算为平台的大规模并行计算，在我国也取得了很大的成果，并且有力地推动了交叉学科的发展。在我们研制的清华大学THNPSC－1和THNPSC－2，以及与上海大学合作研制的“自强2000”等高性能计算平台上，已经在多种领域进行了计算科学与工程的研究，这些研究处在各自学科的前沿。例如我们和化学专家进行了“高分子链”的计算研究，和化工专家进行了“石油化工超临界化学反应”的计算研究，和机械铸工专家进行了“模具充模流体场”的计算研究，和材料科学专家进行了“定量电子晶体学”的计算研究等等。更重要的是，这些计算科学与工程的研究可以减少或免除昂贵的、条件苛刻的或者是长时耗的物理实验，大大节省时间和成本，解决那些原来只有通过高代价的实验方法才能解决的问题，或者仅靠实验方法根本无法解决的问题。最近，我们（作者都志辉是重要的参与者）又完成了教育部重点项目“先进计算基础设施ACI北京上海试点工程”，并通过了鉴定。此项ACI工程，进一步推动了高性能计算的应用，并为我国开展以高性能计算机为平台的跨学科

和跨地区的计算科学与工程创造了良好的环境。

采用大规模计算方式来进行科学与工程领域问题的研究，是一种跨学科创新的源泉。计算科学与工程的研究，在我国越来越重要，也越来越迫切需要在各个领域推广普及。然而由于它是一种跨学科的新生事物，通过计算进行科学与工程问题的研究在实践中会遇到很多困难。主要问题是：非计算机学科的科学和工程技术人员，能够熟练编写串行应用程序已属不易，而能够在高性能计算机上编写并行的应用程序，就更少了。这种情况严重影响了我国高性能计算技术的发展，对于我国推广“计算科学与工程”，以至对于我国发展高科技都很不利。

都志辉博士的这本书，《高性能计算并行编程技术——MPI 并行程序设计》，就是根据我国当前国情和需要而编写的，其目的是促进和推广“计算科学与工程”的发展。此书可以满足非计算机专业的科学与工程人员学习并行编程的需要，不熟悉并行编程的计算机专业科技人员，也可学习这本书。

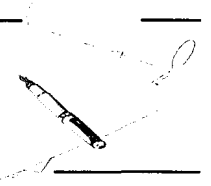
目前，国内外在高性能计算机系统中，最广泛使用的并行编程环境是 MPI，它已成为国际上的一种并行程序的标准。作者以 MPI 并行编程设计作为专门阐述高性能计算并行编程技术的第一本书出版，是很有意义的。作者本人在北京大学博士生研究期间，在清华大学从事博士后研究和出站后的工作期间，专门从事并行程序及并行编译的研究，并在 THNPSC－1、THNPSC－2 和“自强 2000”高性能计算机的设计、应用和并行编程等方面开展了大量的实际工作，因此，这本书是在他多年的实际研究和工作经验的基础上写成的，这就使这本书更有参考价值。

作者由浅入深，对并行编程技术进行了介绍。先介绍了并行程序设计的基本知识和 MPI 的基本功能、过程和框架，然后阐述 MPI 程序设计的高级特征，使读者可以编写一些高效的 MPI 并行程序。作者在这本书的后面部分，还介绍了高级和较复杂的 MPI 并行程序编写技术。最后，作者还向读者介绍了 MPI 的新发展，即 MPI－2。本书的一个特点是：作者在向读者讲解 MPI 编程技术时，采用了大量的编程实例，对于关键部分还给出了注释，使读者能更好地掌握 MPI 编程的方法，而不只是理论知识。比较深入了解高性能计算技术和应用的读者都知道，高性能计算的应用，实质性问题是计算的效率问题，这和串行程序有很大的不同，作者在本书的很多章节中，都贯穿了如何提高并行编程效率的思想，这也是本书的另一个重要特点。这本书理论和实际结合得较好，对于计算机领域的高年级学生和非计算机专业的研究生，可以用作教科书；对于有一定经验的科技人员，可以帮助他们较快地掌握并行编程技术；对于短期培训班，可采用此书的部分章节来使用。

我们已经进入了一个新的世纪，我们的祖国也将进入一个“科技强国”的新时期。高性能计算技术的推广应用，肯定将对我国高科技的发展，起到重要的作用。我衷心希望这本书的出版，将对高性能计算技术在我国的推广，对于科技人员尽快掌握并行编程技术，作出应有的贡献。

中国工程院院士
清华大学教授
李三立
2001 年 6 月 24 日

前言



此书是在李三立院士的直接支持和指导下完成的。根据我们多年来从事并行计算研究的实践，发现并行计算已经从原来的阳春白雪、曲高和寡的局面越来越渗透并普及到各个领域和方面，并行计算的硬件支持也从原来只有国家或大公司才能负担的超级计算机发展到现在一个小的实验室就可以组建的“个人超级计算机”——网络工作站或网络 PC。并行计算是一种强大的计算工具，但是，真正需要并行计算的非计算机专业人员却缺乏这方面的指导，他们迫切需要一本较为通俗的、可以较快理解并掌握并行计算这一工具的参考书，而不是过多侧重并行计算理论的专业著作。这是促使本书问世的主要原因。

MPI 是目前最重要的并行编程工具，它具有移植性好、功能强大、效率高等多种优点，而且有多种不同的免费、高效、实用的实现版本，几乎所有的并行计算机厂商都提供对它的支持，这是其他的并行编程环境所无法比拟的。MPI 于 1994 年产生，虽然产生时间相对较晚，但由于它吸收了其他多种并行环境的优点，同时兼顾性能、功能、移植性等特点，在短短的几年内便迅速普及，成为消息传递并行编程模式的标准，这也从一个方面说明了 MPI 的生命力和优越性。

MPI 其实就是一个“库”，共有上百个函数调用接口，在 FORTRAN 77 和 C 语言中可以直接对这些函数进行调用。MPI 提供的调用虽然很多，但最常使用的只有 6 个，只要会使用 FORTRAN 77 或者是 C 语言，就可以比较容易地掌握 MPI 的基本功能，只需通过使用这 6 个函数就可以完成几乎所有的通信功能。由于 Fortran 90 和 C ++ 语言的使用也十分广泛，MPI 后来又进一步提供对 Fortran 90 和 C ++ 的调用接口，这更提高了 MPI 的适用性。

根据我们的经验，只要是掌握了 FORTRAN 77 语言和 C 语言的程序员，哪怕一开始对 MPI 并行程序设计一无所知，只要通过一定的培训和练习，是可以很快掌握 MPI 的基本功能并使用它来编写并行程序的。但是要全面掌握 MPI，则必须通过大量的练习并需要更长的时间。有鉴于此，本书的编写也是按照从易到难、从简单到复杂、从低级到高级的顺序进行的。书中第一部分简单介绍了并行程序设计的基本知识；第二部分介绍基本的 MPI 并行程序设计方法，它虽然基本，但是却非常重要，因为通过这部分介绍的功能，可以实现几乎所有的通信功能；第三部分是在第二部分的基础上，介绍高级、复杂的 MPI 并行程序设计，使用

高级的 MPI 调用可以提高并行程序的通用性和移植性，对提高并行程序的开发效率、可读性以及并行程序的执行效率等都有好处；最后一部分介绍 MPI 的最新扩展 MPI-2，着重对动态进程管理、远程存储访问和并行 I/O 进行讲解。

据我们所知，目前国内还没有一本专门介绍 MPI 并行程序设计的著作，希望本书的问世能够对那些迫切需要这方面参考资料的读者有所帮助，并对并行计算在我国的推广和普及作出贡献。

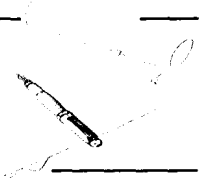
在本书的筹划过程中，清华大学出版社的李幼哲老师根据他多年从事计算机类图书出版的经验，对我进行了指导并提出了大量的建设性意见，我感谢他！本教研组的老师和同学也对本书提出了许多宝贵的建议，我感谢他们！

希望读者在阅读此书时，将发现的错误或建议告诉我（duzh@ tirc. cs. tsinghua. edu. cn），以便在再版时进一步提高本书的质量，我将非常感谢！

作者

2001 年 2 月 1 日于清华大学

目录



第三部分　高级 MPI 并行程序设计

第四部分　MPI 的最新发展 MPI-2

第一部分

并行程序设计基础

本部分包括如下内容：并行计算机、并行编程模型与并行语言、并行算法。

通过本部分的介绍，使读者对并行计算和并行程序设计有一个基本的概念，为后续几章具体讲解 MPI 并行程序设计方法提供基础知识。

第1章

并行计算机

本章给出了并行计算机的基本划分方法和与之相关的体系结构，是宏观的总体的论述而不是具体的细节，这主要是考虑到读者是并行程序设计人员而不是并行机的设计与开发人员。对并行机有一个总体上的了解可以帮助编程者设计出更高效的并行程序。

本章还给出了一个物理问题是如何一步步在并行机上得到解决的，以及并行程序设计在这一过程中所起的作用。

1.1 并行计算机的分类

为什么要采用并行计算？这是因为：(1) 它可以加快速度，即在更短的时间内解决相同的问题或在相同的时间内解决更多更复杂的问题，特别是对一些新出现的巨大挑战问题，不使用并行计算是根本无法解决的；(2) 节省投入，并行计算可以以较低的投入完成串行计算的任务；(3) 物理极限的约束，光速是不可逾越的速度极限，设备和材料也不可能做得无限小，只有通过并行才能够不断提高速度。

并行计算机即能在同一时间内执行多条指令(或处理多个数据)的计算机，并行计算机是并行计算的物理载体。通过下面对并行计算机的不同分类方式，可以对它有一个总体上的了解，为并行程序设计奠定基础。

1.1.1 指令与数据

根据一个并行计算机能够同时执行的指令与处理的数据的多少，可以把并行计算机分为 SIMD(single-instruction multiple-data，单指令多数据并行计算机)和 MIMD(multiple-instruction multiple-data，多指令多数据并行计算机)两种，如图 1 所示。

SIMD 计算机同时用相同的指令对不同的数据进行操作。比如数组赋值运算

$$A = A + 1$$

在 SIMD 并行机上可以用加法指令同时对数组 A 的所有元素实现加 1。即数组(或向量)运算特别适合在 SIMD 并行计算机上执行，SIMD 并行机可以对这种运算形式提供直接地支持