前　言

1998 年上半年，国家高性能计算中心（上海）在复旦大学物理楼设立，使我第一次与并行计算机有了近距离的接触。此前的 1996 年和 1997 年，虽然在某地区计算中心也曾经使用过 IBM-SP2 并行机（82 个节点），但只是作串行计算。当时，并行计算在我心中还是比较神秘的。

该中心成立之际，开设了一天并行计算机使用讲座。在一个下午被灌输了大约 50、60 种 MPI 并行函数的使用方法后，就开始了并行计算的实践。该中心配置的一台仅有 16 个节点的曙光 1000 一时用者如潮，深感不便，就自己动手，将自己和自己的研究生使用的、连在校园网上的两台 Linux 微机设置成了本人第一台微机集群。虽然当时复旦大学校园网带宽只有 10M，连接的还是集线器，但在这个集群上调试程序已经绰绰有余。大约化了一个月左右的时间，将自己最常用的一个串行程序——第一性原理混合基赝势方法从头计算程序包（约三万多行）——在这台集群上改成了并行计算程序。在随后的访问香港工作期间，这个并行计算程序充分发挥了它的效力。短短三个月，用它完成了一个相当有意义的工作——第一次接触就领略了并行计算速度威力的巨大冲击。

浅尝而不能停止，并行计算就是有如此魅力，自己动手组建微机集群的实践就此起步，时间是 1998 年 5 月。当时自建集群的网络带宽只有 10M。自己没有意识到，这种微机集群有朝一日竟会发展到足以与商业计算机大公司的传统超级计算机竞争的地步；更没想到，随着网络技术的飞速发展，100M 带宽网络成为主流后，自己也会组建一台速度可以进入世界超级计算机 500 强的微机集群。

这已经是 2001 年的事了。在复旦大学计算凝聚态物理 985 计划 100 万元人民币的资助下，自己开始着手试制、组建大型微机集群。经过一年左右的样机（8 个 K7 节点）试制后，在 2002 年 5 月底终于完成了一台大型微机集群（96 个 P4 节点）。该集群采用了网卡捆绑和网络启动两项当时国内率先采用的技术。用世界超级计算机 500 强排序所采用的 High Performance Linpack（HPL）标准程序测试，该集群的 Linpack 速度达每秒 1417 亿次，可以排当时（2002 年 6 月底公布）的世界超级计算机 500 强第 468 位。

为什么不是现在更时髦的网格计算，而是并行计算？回答是：应用方便。理论上网格计算能更有效地利用现成的计算机资源。但它的应用程序，比如现有的串行程序如何移植成适合网格计算的程序？这个问题不能容易地解决，网格计算只适合有限的特有的应用；而对于并行计算本书将介绍，最简单地只需掌握一个函数，就能将现有的串行程序方便地移植成并行计算程序，而掌握这个函数又不比编程中掌握使用 sin 和 cos 函数难多少。

为什么不购买现成的超级计算机，而是组建微机集群？回答是：经费匮乏！微机集群性能价格比决定了它是一种符合我国国情的超级计算机，也由此有了大力推广它的最初设想。将我的这些经验写出来，让更多的人分享，则是在开课讲授“自己动手组建超级计算机”之后。

2002/2003 学期，我为复旦大学物理系的学生开设《Linux 系统基础》一课。该课程原是为培养复合型人才而设，说白了，是为了让学生毕业后更顺利地求职而设。现在流行自

己动手做（DIY），不谋而合，该课程就演变成了 DIY 超级计算机的课程，大受学生欢迎。由此想到，也许国内众多玩家也会广有兴趣。实际上，一个大学生，甚至一个高中生，即使不懂 Linux，也可以实践 DIY 超级计算机，复旦大学物理系 1999 级和 2000 级那批学生就是明证——他们都是为了选修《Linux 系统基础》课程而来的，原先当然连 Linux 也没有学过。宿舍里、家里的几台微机也成了他们 DIY 超级计算机的对象。他们从该课程中所学到的基本知识，一旦有机会，完全可以用来组建大型的微机集群——超级计算机。

可是目前，无论国内还是国外，都还没有一本这方面的合适教材。国外虽然已经有一些关于微机集群的著作出版，但大多着重集群结构原理、并行计算方法、网络技术理论等等。读者对象大概是熟悉计算机结构、网络技术和 Linux 操作系统的学者。至今尚无一本详细地讲解如何自己动手组建超级计算机这类入门书。在同学的要求下，在我系王迅院士的鼓励和推动下，有了写出自己在这方面经验的冲动，遂演绎成本书。

这是一本这样的书——教读者如何花费相对较少的钱，用市场上可以买到的现成的微机，交换机以及网上可以免费下载的 Linux 操作系统，自己动手组建一台超级计算机——同样计算速度的传统超级计算机的价格在它的十倍以上。

这不是一本关于并行计算机结构理论或并行算法的书。我在本书中将尽可能地手把手地教读者如何组建、如何实践、如何管理这类微机集群超级计算机。目的是使读者了解，实现并行计算在并行计算机结构上有什么要求，这样的结构在硬件和软件上如何实现。即使读者连 Linux 也没有学过，也可以象复旦大学物理系 1999 级和 2000 级的那批学生一样基本学会如何自己动手组建超级计算机——学会这些组建微机集群超级计算机必不可少的知识。

本书既然讲述自己动手组建超级计算机，就需要对超级计算机下一个定义。也许，读者概念中的超级计算机与科幻影片中的庞大的机柜，飞速旋转的磁带机，不停地闪烁的 LED 联系在一起的。由于计算机技术的飞速发展，超级计算机的概念实际上一直在变。超级计算机就是速度最快、性能最强的计算机。按 Linpack 每秒浮点运算次数排序的世界超级计算机 500 强可以作为一个标准——能排进此行列的计算机应该可以被称为超级计算机。

本书各章内容安排如下：

第1章、 微机集群的由来、现状和趋势；市场决定微机集群是发展方向，因为它符合通用，开放，兼容的标准。微机集群的最大的优点是性价比；最大的缺点是网络速度不及传统超级计算机，并且难以管理。

第2章、 介绍目前流行的消息传递并行计算对微机集群的网络结构有什么要求；所需的网络功能；微机集群网络结构；为自己动手组建微机集群做准备。

第3章、 根据微机集群所需的网络结构和网络功能进行对微机集群服务器和第一台节点机进行安装和设置；介绍 SuSE Linux 的强大的管理工具 YaST；根据微机集群的需要对内核进行重新编译以及 MPI 并行平台的安装。

第4章、 只需一根网线就可将服务器和一台节点机连接成集群；如何进行网络连接；网

络功能调试；速度测试；以及并行计算平台可能碰到的问题和应对方法。

第5章、　针对微机集群的一些难点讲解：网络唤醒和网络启动；网络自动管理、配置；网络通讯瓶颈的性价比较好的解决方案，网卡捆绑。

第6章、　针对微机集群的另一管理难点讲解：任务排队管理系统的安装和设置。

附录A、　组建微机集群必须知道的 Linux 基础知识，供对 Linux 不太熟悉的读者参考。

附录B、　组建大型微机集群时的一些选购硬件的经验，供读者参考。

本书写作分工如下：附录 A 由洪峰撰写，附录 B 由赵坚撰写，其余章节由车静光撰写。

对读者的建议如下：如果没有接触过 Linux 的读者，可先阅读附录 A 的“ Linux 系统基础 ”或其他 Linux 的入门著作。只想过把组建微机集群瘾的玩家，浏览第 2、3、4 章内容足矣。这三章足以帮助读者从头至尾组建、调试和测试一台个人微机集群。两台微机一根线，就可以实践自己动手装超级计算机。如果要组建一台可靠的微机集群，还需了解一些硬件性能，可参阅附录 B 的“ 微机集群的硬件选购 ”。想少花钱，多办事的读者，就请在第 5 章里寻找有哪些东西能为你所用。集群规模大于 16 个节点以上，管理工作就占据了突出的地位，就需要掌握第 6 章的内容。

本人从事物理研究，只是在研究工作需要、研究经费又捉襟见肘的情况下，才走上了这条“ 自己动手组建超级计算机 ”的道路。本书是自己安装大型微机集群过程中的经验总结：有些问题的答案是通过咨询好友得到的（有关网站、Linux 自带的 Howto 等也是最好的朋友）；有些问题的答案属于自己的探索和理解（希望能自圆其说，不致太过谬误）。本人教学、研究工作繁忙，成书时间仓促，错误在所难免，非常欢迎计算机专家、学者和广大 DIY 发烧友、玩家批评指正。

值此脱稿之际，感谢我的硕士导师、谢希德院士和张开明教授，是她们引导我走上了艰难而又有意义的表面物理研究道路；感谢王迅院士，他的鼓励和鞭策是本书写作启动的第一推动力；感谢清华大学李家明院士，他的关心和指导是我不断尝试微机集群新技术的源泉；感谢好友香港科大冯永嘉博士的无私帮助，否则我还需在实践中摸索更长的时间；感谢我的博士导师、德国明斯特大学的 J. Pollmann 教授，我的计算物理学（包括计算机编程）方面的严格训练是在那个特有的环境下完成的；感谢同事资剑教授的支持和信任，使我得以完成大型微机集群的实践。最后，感谢妻子顾青和女儿车逸文，她们的理解、支持和牺牲是如期完成本书的保证。

车静光

2003 年 2 月于复旦大学物理系