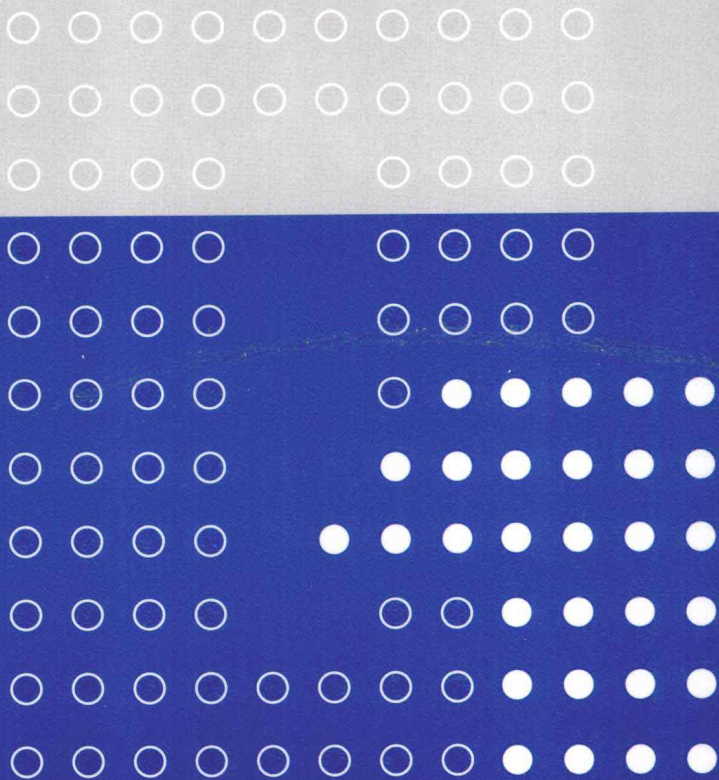


计算机系列教材

OpenMP

编译原理及实现技术



罗秋明 明仲 刘刚 毛睿 著



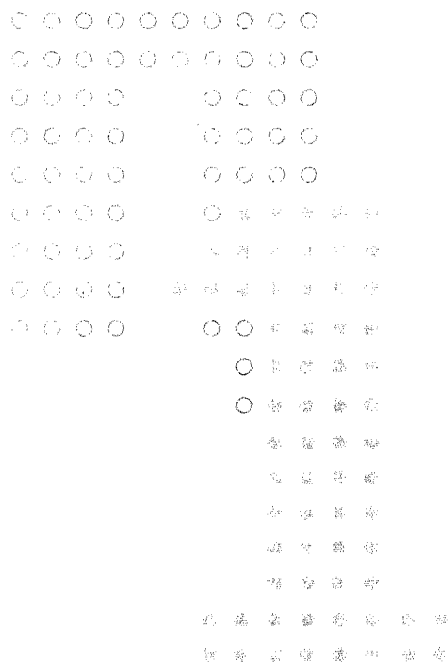
清华大学出版社

计算机系列教材

罗秋明 明仲 刘刚 毛睿 著

OpenMP

编译原理及实现技术



清华大学出版社
北京

内 容 简 介

本书是一本学习 OpenMP 编译原理和实现技术的入门级教材。全书分成三篇,第一篇是并行计算及 OpenMP 编程的基础内容,第二篇是 OpenMP 编译及其运行环境,第三篇是实践内容。在第二篇中,以一般编译器常见结构为主线,通过结合详细的 OMPi 源代码分析向读者介绍 OpenMP 编译器的工作原理及其实现技术,具体包括词法分析、语法分析、AST 的结构、AST 的生成及相关操作、OpenMP 编译制导指令的代码变换、OpenMP 线程与 OS 线程库的接口、运行环境等细节。OpenMP 编译制导指令的变换是 OpenMP 编译的核心内容,需要将 OpenMP 编译制导指令的语义功能利用操作系统的线程库来实现,分成并行域管理问题、任务分担和同步问题、变量数据环境问题三个核心内容。第三篇的四章内容分别给出了常见编译器、性能测试工具以及 OMPi 源代码的框架分析。

本书是国内第一本对 OpenMP 编译器工作原理和实现细节进行阐述的图书。本书是研究 OpenMP 编译技术的研究人员和高校师生的入门教材,也是研究生和高年级本科生学习并行语言编译技术相关课程的参考书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

OpenMP 编译原理及实现技术/罗秋明等著. —北京:清华大学出版社,2012.5

(计算机系列教材)

ISBN 978-7-302-27298-4

I. ①O… II. ①罗… III. ①并行程序—程序设计—教材 IV. ①TP311.11

中国版本图书馆 CIP 数据核字(2011)第 233253 号

责任编辑:龙啟铭 李玮琪

封面设计:常雪影

责任校对:梁 毅

责任印制:杨 艳

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京密云胶印厂

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:185mm×260mm

印 张:20.5

字 数:480千字

版 次:2012年5月第1版

印 次:2012年5月第1次印刷

印 数:1~3000

定 价:29.50元

产品编号:041766-01



随着多核处理器在台式机、笔记本以及服务器中的广泛使用，我们迎来了全面的并行计算时代。无论是系统软件还是应用软件，也不论是编程人员还是用户都必须作好准备迎接这次并行浪潮，因此高校的教学和教材方面也需要紧跟技术的进步进行扩展。

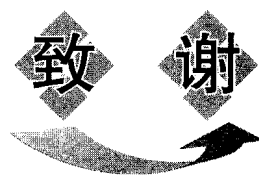
多核处理器可以构建出 SMP、NUMA 等架构的并行平台，这类共享内存的计算机上的一个主流编程技术是 OpenMP。虽然当前关于编译原理和 OpenMP 编程的教材图书已有不少，但是还未有 OpenMP 编译方面的专著或教材。

传统的编译原理教材主要针对串行语言中的语法规则、有限自动机、语法分析方法等，并在优化部分讨论了并行性问题，例如在经典教材——Alfred 的《Compilers: Principles, Techniques, & Tools》中讨论了 SMP 和 NUMA 等共享内存平台上并行编译的有关迭代空间、数据空间和处理器空间上的仿射变换等问题，但一般不涉及具体语言和应用技术。本教材则可以衔接传统的编译原理课程，在掌握编译器的并行性相关原理后，将串行语言的编译扩展到并行语言的编译。针对具体的 OpenMP，本书讨论如何将 OpenMP 程序翻译转换到标准 C 语言代码，并利用操作系统的线程或进程来实现具体的并行语义行为，书中内容非常贴近实践并有大量代码的解释说明。了解 OpenMP 编译也有利于其编程，通过对 OpenMP 编译原理和实现技术的学习，可以更好地了解程序和相关的开销，编写出性能更好的代码。

该书有选择地介绍了大多数读者可以接触到的并行计算的相关知识，然后结合样例代码对 OpenMP 的编程要素和具体方法进行讲解。在讲述词法分析和语法分析以及 AST 的创建时，着重讲解 Lex、Yacc 工具的应用以及 AST 的数据结构和相关操作，加强了编译原理实践性内容。作者还使用了大量图表来辅助原理的说明，并有丰富的源代码分析。该书还分析了 OpenMP 开源编译器、基准测试工具以及 OMPi 编译器的源码框架。对于有志研究 OpenMP 编译器的读者，这是难得的好书。

陈国良

2011-10-20



本书得以顺利完稿，是许多人的共同努力！

在此首先要感谢陈国良院士的支持和帮助。陈院士于2009年到深圳大学主持计算机与软件学院的教学科研的全面工作，作者作为其高性能团队成员有幸参与了众多高性能计算的科研工作，包括参与全国产万亿次个人高性能计算机KD-60的研制、作为核心人员参与SD-1 PHPC研制以及正在进行基于龙芯处理器的SD-30十万亿次全国产化高性能计算机的研制等。在这些工作中，作者萌发了编写OpenMP编译方面书籍的想法，陈院士表示认可和支持并给出了非常有价值的意见和建议，这就是本书得以编写并顺利完稿的最初原因。其次需要感谢明仲教授，在本书的最初构思、规划和资料整理的初期，明仲教授不仅阅读了初步的稿件材料并给出了大量的指导和修改意见，对后期稿件也提出了许多建设性意见。

刘刚老师编写了第1章的部分内容和第2章的大部分内容，毛睿老师参与了多个章节的编写和订正工作，陆克中老师阅读了初期的稿件并给出了有益的意见。正因为有了这几位老师的贡献，使本书的编写质量和水平得以提高。另有两位研究生参与了本书的编写，孔畅同学完成了第11章的所有实验并编写了该章内容，刘成健同学是该书的第一位真正意义上的读者，该生与庞宇杰同学参与了稿件查错订正工作。

对于深圳大学高性能计算团队中给本书编写工作提供了各种帮助的老师 and 同学，不能一一尽数，在此一并表示衷心的感谢！



编写本教材是深圳大学“计算机科学与技术国家特色专业建设点”的建设内容之一。该教材和相应课程的设计目的有三点：衔接本科《编译原理》课程、扩展 OpenMP 并行语言编译的知识、增强学生的动手实践和编程能力，书中以 OpenMP 的一个开源编译器——OMP_i 作为分析对象，作到理论与实践紧密结合，为进一步学习和研究打下必要的基础。

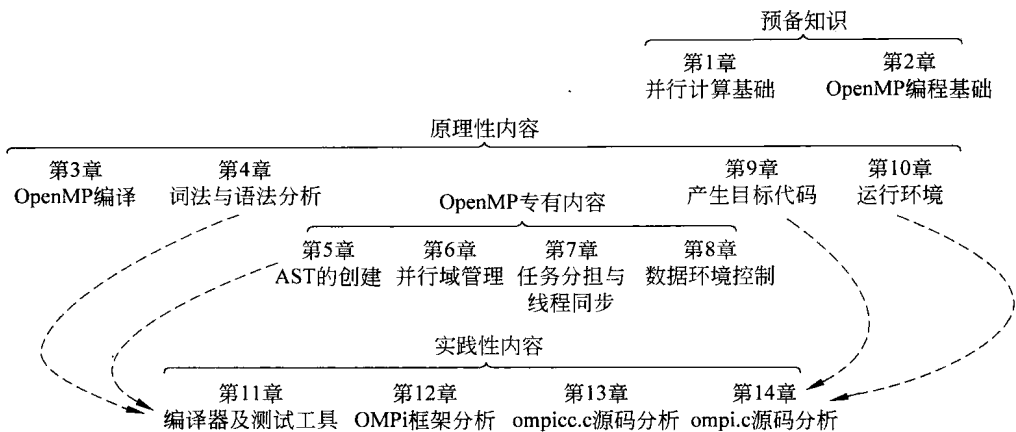
本书虽然不要求读者对编译原理有深入理解，但是需要先对编译有初步认识、具备基本概念。对于关心实现技术的读者，建议下载 OMP_i 1.0.0 的源码并进行同步阅读。

书中第一篇基础部分共两章，分别讲述并行平台和 OpenMP，C 编程作为预备知识。对于没有接触过并行计算技术的读者可以作为一个补充阅读，第 2 章的 OpenMP 编程基本上将相关的语言要素作了全面的介绍，没有特殊需要一般不再需要去阅读 OpenMP 的标准。

第二篇编译部分共八章，主要讲述 OpenMP 的编译器以及运行环境。第 3 章介绍 OpenMP 编译器基本框架。第 4 章介绍词法分析和语法分析，主要以 Lex 和 Yacc 工具实现 OpenMP 词法和语法分析为主要内容。第 5 章介绍 OpenMP 编译中使用的 AST 中间表示。第 6 章、第 7 章和第 8 章分别介绍 OpenMP 编译中的并行域管理、任务分担和同步、变量数据环境三大问题，这是 OpenMP 编译的核心所在。第 9 章讲述目标代码生成，主要是如何利用“框架”来实现 OpenMP 翻译的技术。第 10 章以 OMP_i 运行环境为例讲述相应的运行环境。

第三篇实践部分共四章。第 11 章给出了几种常见的 OpenMP 编译器以及性能测试工具，用于读者测试自己设计的或修改的编译器性能。第 12 章、第 13 章和第 14 章是关于 OMP_i 编译器工作流程和框架的内容，并有两个主要的上层源文件的代码分析，如果希望在 OMP_i 的基础之上进行增强改进，那么这三章的内容将有所帮助。

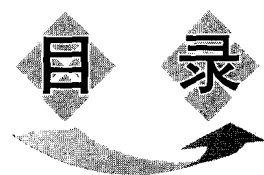
对学习或课程的安排可以分成不同流程，如根据教学偏向于原理性还是实践性、是否具有并行计算基础等情况，可选取书中部分章节作为参考材料，书中各章倾向性情况安排如下：



虽然书中有大量的代码，但是读者在第一遍阅读时可以只作粗略浏览，第二遍阅读时再仔细阅读。由于源代码阅读往往需要参考不同章节的内容，因此书中有大量的交叉引用说明。如果对 OpenMP 编译有一定的了解，并希望掌握 OMPi 的具体编码实现技术，可以从第 12 章开始阅读，将整体框架看完，再根据需要返回第二篇阅读编译细节。

限于作者的学识水平，书中难免有不少错误和不足，恳请读者批评指正。作者在编写本书时深感“抛砖引玉”一词不仅仅是场面上的客套话，更是作者的内心体会和期望。

编者
2011 年 12 月



第一篇 基 础

第 1 章 并行计算基础 /3

1.1	基本概念	3
1.2	并行计算平台	4
1.2.1	典型结构	4
1.2.2	SMP	6
1.2.3	NUMA	8
1.2.4	GPU	10
1.2.5	Cluster	11
1.3	并行程序设计技术	14
1.3.1	并行程序设计	14
1.3.2	OpenMP	17
1.3.3	MPI	17
1.3.4	CUDA	18
1.3.5	HPF	19
1.4	本章小结	19
	习题	19

第 2 章 OpenMP 编程基础 /20

2.1	OpenMP 基本概念	20
2.1.1	执行模式	20
2.1.2	OpenMP 编程要素	21
2.2	OpenMP 编程	24
2.2.1	并行域管理	24
2.2.2	任务分担	25
2.2.3	同步	34
2.2.4	数据环境控制	41
2.3	本章小结	49
	习题	49

第二篇 OpenMP 编译

第 3 章 OpenMP 编译 /53

3.1	OpenMP 编译系统	53
3.1.1	编译系统	53
3.1.2	目标语言	54
3.2	OpenMP 编译器结构	55
3.2.1	功能模块	56
3.2.2	工作流程	59
3.3	编译优化	60
3.4	本章小结	60
	习题	60

第 4 章 词法与语法分析 /61

4.1	Lex 工具	61
4.1.1	Lex 的正则表达式	62
4.1.2	Lex 的使用方法	64
4.2	OpenMP/C 的词法分析	66
4.2.1	C 语言单词	66
4.2.2	OpenMP 单词	66
4.2.3	OpenMP 与 C 语言公用单词	67
4.3	scanner.l	68
4.3.1	全局声明段	68
4.3.2	模式匹配规则段	70
4.3.3	补充函数段	75
4.3.4	scanner.c	78
4.3.5	scanner.h	78
4.4	Yacc 工具	79
4.4.1	Yacc	79
4.4.2	Yacc 文件实例	81
4.5	OpenMP/C 语法分析	85
4.6	本章小结	88
	习题	88

第 5 章 AST 的创建 /90

5.1	中间表示	90
5.1.1	两种中间表示形式	90

5.1.2	中间表示的选择.....	91
5.2	AST 节点数据结构.....	92
5.2.1	语句节点.....	92
5.2.2	类型说明节点.....	95
5.2.3	声明节点.....	96
5.2.4	表达式节点.....	97
5.2.5	OpenMP 制导节点.....	98
5.3	AST 节点维护函数.....	103
5.4	AST 的创建.....	104
5.4.1	语法制导翻译.....	105
5.4.2	例 1 OpenMP 的 for 节点.....	107
5.4.3	例 2 C 语言 while 语句.....	108
5.4.4	Helloworld.c 的 AST.....	109
5.5	符号表.....	111
5.5.1	字符串表.....	112
5.5.2	符号表.....	112
5.5.3	符号表操作.....	114
5.5.4	作用域管理.....	115
5.6	本章小结.....	115
	习题.....	116

第 6 章 并行域管理 /117

6.1	并行域及其嵌套.....	117
6.2	并行域管理.....	119
6.2.1	线程无关接口.....	119
6.2.2	线程的供给.....	120
6.2.3	线程层次关系.....	120
6.2.4	并行域代码封装与标识.....	122
6.2.5	任务分担问题.....	122
6.3	目标代码形式.....	123
6.4	OMP _i 的并行域管理.....	124
6.4.1	ORT 统一界面.....	124
6.4.2	并行域代码变换.....	126
6.4.3	线程管理与控制.....	127
6.4.4	总览.....	132
6.5	本章小结.....	133
	习题.....	133



第7章 任务分担与线程同步 /134

7.1	for 制导指令	134
7.1.1	for 任务分担	134
7.1.2	循环变量分解原则	135
7.1.3	目标代码功能	136
7.1.4	目标代码形式	137
7.1.5	OMP _i 的 for 制导指令	138
7.2	sections 制导指令	144
7.2.1	sections 任务分担描述	144
7.2.2	section 划分原则	145
7.2.3	目标代码功能	145
7.2.4	目标代码形式	145
7.2.5	OMP _i 的 sections 制导指令	147
7.3	single 制导指令	149
7.4	nowait 问题	149
7.5	归约操作	151
7.6	线程同步	152
7.6.1	atomic	152
7.6.2	critical	153
7.6.3	master	154
7.6.4	ordered	154
7.6.5	nowait	155
7.6.6	flush	155
7.6.7	barrier	155
7.7	本章小结	155
	习题	156

第8章 数据环境控制 /157

8.1	共享与私有	157
8.1.1	非全局变量的共享	158
8.1.2	变量的私有化	159
8.1.3	threadprivate 子句	159
8.1.4	基于进程的问题	160
8.2	并行域边界处理	160
8.2.1	private 变量	160
8.2.2	threadprivate 变量	160

8.3	OMP <i>i</i> 数据环境控制	161
8.3.1	共享变量	161
8.3.2	私有变量	163
8.3.3	线程专有变量	166
8.3.4	归约变量	169
8.4	本章小结	170
	习题	170
第 9 章 产生目标代码 /171		
9.1	源代码变换	171
9.1.1	变换流程	171
9.1.2	支撑函数	173
9.2	AST 变换	173
9.2.1	拼接及创建函数	173
9.2.2	变换函数集	174
9.2.3	OpenMP 节点变换	177
9.2.4	parallel 变换	180
9.2.5	for 变换	183
9.2.6	sections 变换	185
9.2.7	数据环境的处理	186
9.3	代码优化	188
9.4	AST 输出	189
9.4.1	OMP <i>i</i> 的 AST 输出	189
9.4.2	OpenMP 节点输出	189
9.5	本章小结	192
第 10 章 运行环境 /193		
10.1	重要数据结构	193
10.1.1	ORT	193
10.1.2	线程池与 EECB	194
10.1.3	任务分担结构	195
10.1.4	共享变量结构	197
10.2	初始化与退出	198
10.2.1	ORT 初始化	199
10.2.2	EELIB 初始化	202
10.3	并行支撑函数	204
10.3.1	线程状态管理	204
10.3.2	并行域管理	206

10.3.3	任务分担.....	211
10.3.4	同步.....	220
10.3.5	变量的数据环境.....	221
10.4	OpenMP 的 API.....	222
10.4.1	API 函数.....	222
10.4.2	ICV 变量.....	224
10.4.3	引用与链接.....	225
10.5	环境变量.....	225
10.6	本章小结.....	226

第三篇 实践篇

第 11 章 编译器及测试工具 /229

11.1	常见 OpenMP 编译器.....	229
11.1.1	GCC 编译器.....	229
11.1.2	OMP <i>i</i> 编译器.....	231
11.1.3	Omni 编译器.....	233
11.2	性能测试工具.....	235
11.2.1	EPCC Microbenchmark.....	236
11.2.2	NSA Parallel Benchmark.....	242
11.2.3	SPEC OMP2001.....	243
11.2.4	LLNL.....	243
11.3	本章小结.....	243

第 12 章 OMP*i* 框架分析 /245

12.1	工作流程.....	245
12.2	OMP <i>i</i> 的处理步骤.....	246
12.3	代码转换.....	250
12.4	进程问题.....	252
12.4.1	全局变量.....	252
12.4.2	非全局共享变量.....	253
12.5	运行环境.....	254
12.5.1	初始化.....	254
12.5.2	并行域的处理.....	254
12.5.3	任务分担.....	254
12.5.4	同步.....	256
12.5.5	线程专有变量.....	256

12.5.6	与 EELIB 的接口	258
12.6	源代码文档结构	258
12.7	后续阅读建议	259
12.8	本章小结	259
第 13 章	ompicc.c 源码分析 /260	
13.1	ompicc 工作流程	260
13.2	变量声明及参数处理	261
13.3	编译部分	271
13.3.1	文件名处理	272
13.3.2	预处理	273
13.3.3	代码变换	274
13.3.4	C 编译	275
13.4	链接部分	276
13.5	主函数部分	278
13.5.1	参数及配置函数	278
13.5.2	main 函数	281
13.6	配置文件	283
13.7	运行参数与选项	287
13.7.1	环境变量	288
13.7.2	命令行参数	288
13.7.3	配置文件	291
13.8	本章小结	292
第 14 章	ompi.c 源码分析 /293	
14.1	ompi 工作流程	293
14.2	ompi.c	294
14.2.1	变量声明及辅助函数	294
14.2.2	main() 函数	297
14.2.3	错误处理	303
14.3	ort.defs	303
14.4	ompi.h	307
14.5	小结	308



第一篇 基础

第一篇的内容是并行计算和 OpenMP 编程的基础。由于并行计算和 OpenMP 编程在大多数高校的本科课程中都没有被讲授，因此基础篇中选择这两个部分作为内容，而其他如编译原理、操作系统、C 语言编程等预备性基础内容在本科课程都有讲授，不需要在本书中重复。

第 1 章包括并行软硬件平台和相应的模型，关于并行计算平台的内容首先介绍了典型结构并略为详细地介绍了 SMP、NUMA、GPU 和 Cluster 四种常见并行计算机，关于并行程序设计技术则介绍了相关的计算模型、编程模型和 PCAM 设计方法，并略微详细地介绍了 OpenMP、MPI、CUDA 和 HPF 四种常见编程技术。

第 2 章比较全面地介绍了 OpenMP 编程的基础知识，包括 OpenMP 的基本元素和所有的制导指令方面的内容。关于基本元素分成了制导指令、API 函数和环境变量三个部分，制导指令首先介绍的是并行域方面的内容，然后是任务分担及同步的内容，最后是数据环境控制的内容。

