

TURING

图灵计算机科学丛书

[美] John L. Hennessy David A. Patterson 著 贾洪峰 译



ELSEVIER
爱思唯尔

计算机体系结构

量化研究方法 (第5版)

Computer Architecture
A Quantitative Approach

FIFTH EDITION



人民邮电出版社
POSTS & TELECOM PRESS



图灵计算机科学丛书

Computer Architecture
A Quantitative Approach Fifth Edition

计算机体系结构 量化研究方法 (第5版)

[美] John L. Hennessy 著
David A. Patterson

人民邮电出版社
北 京

图书在版编目 (CIP) 数据

计算机体系结构：量化研究方法：第5版 / (美)
亨尼西 (Hennessy, J. L.), (美) 帕特森
(Patterson, D. A.) 著；贾洪峰译. — 北京：人民邮电
出版社, 2013. 1

(图灵计算机科学丛书)

书名原文: Computer Architecture: A Quantitative
Approach, Fifth Edition
ISBN 978-7-115-29765-5

I. ①计… II. ①亨… ②帕… ③贾… III. ①计算机
体系结构 IV. ①TP303

中国版本图书馆CIP数据核字 (2012) 第253568号

内 容 提 要

本书是最权威的计算机体系结构著作，是久负盛名的经典作品。书中系统地介绍了计算机系统的设计基础、指令集系统结构、流水线和指令集并行技术、层次化存储系统与存储设备、互连网络以及多处理器系统等重要内容。在这个最新版中，作者增加了当前炙手可热的云计算和手机客户端技术等相关内容，探讨了在手机、平板电脑、笔记本电脑和其他移动计算设备上云计算的软硬件实现方式。

本书可作为高等院校计算机专业本科生或研究生教材，也可作为从事计算机体系结构或计算机系统设计的工程技术人员的参考书。

图灵计算机科学丛书

计算机体系结构：量化研究方法 (第5版)

-
- ◆ 著 [美] John L. Hennessy David A. Patterson
 - 译 贾洪峰
 - 责任编辑 毛倩倩
 - 执行编辑 丁晓昀
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
 - 邮编 100061 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 北京艺辉印刷有限公司印刷
 - ◆ 开本: 787×1092 1/16
 - 印张: 38.75
 - 字数: 1139千字 2013年1月第1版
 - 印数: 1-6 000册 2013年1月北京第1次印刷
 - 著作权合同登记号 图字: 01-2012-6910号
 - ISBN 978-7-115-29765-5
-

定价: 109.00元

读者服务热线: (010)51095186转604 印装质量热线: (010)67129223

反盗版热线: (010)67171154

对本书的赞誉

“《计算机体系结构：量化研究方法（第5版）》继续发扬传统，为学习计算机体系结构的学生提供了当前计算平台的最新信息，使他们能够洞悉体系结构，便于设计未来系统。这一版的亮点在于大幅修订了数据级并行那一章，用传统的体系结构术语清晰地解读了 GPU 体系结构。”

——Krstě Asanović, 加州大学伯克利分校

“《计算机体系结构：量化研究方法》是一部经典，犹如美酒，历久而弥醇。我在本科毕业时第一次购买了本书，它到现在仍然是我最经常参考的书籍之一。第4版问世时，我发现其中包含了如此之多的新材料，为了跟上这一领域的最新趋势，我必须得再买一本。而当审阅第5版时，我发现 Hennessy 和 Patterson 再现神奇。全书内容都进行了大量更新，对于希望真正理解云和仓库级计算的人们来说，单凭第6章的内容，这个新版本就值得一读。只有 Hennessy 和 Patterson 才可能接触到谷歌、亚马逊、微软等云计算与互联网规模的应用提供商的内部人士，对这一重要领域，业内的介绍材料无出其右。”

——James Hamilton, Amazon Web 服务部

“Hennessy 和 Patterson 撰写本书的第一版时，研究生们是在用 50 000 个晶体管组装计算机。今天，仓库级的计算机集群会包含 50 000 个服务器，每个服务器中包含数十个处理器和数十亿个晶体管。计算机体系结构一直在不停地快速发展，而《计算机体系结构：量化研究方法》紧跟它的步伐，每个版本都准确地解释和分析了这一领域激动人心的最新重要思想。”

——James Larus, 微软研究院

“这一版新增加了一章非常丰富的内容，用来讨论向量、SIMD 和 GPU 体系结构中的数据级并行技术。它解释了应用于大众市场的 GPU 内部的关键体系结构概念，给出这些概念与传统术语的对应关系，并与向量和 SIMD 体系结构进行了对比。这一内容非常及时，与业内转向 GPU

并行计算的潮流相适应。《计算机体系结构：量化研究方法》继续独领风骚，全面地介绍了体系结构方面的重大新进展！”

——John Nickolls, NVIDIA

“本书已经成为一本经典教科书了，这一版突出介绍了各种显式并行技术（数据、线程、请求）的兴起，各用整整一章来描述。数据并行一章尤为夺目：通过向量 SIMD、指令级 SIMD 和 GPU 的对比，避开每种体系结构的专用术语，揭示了这些体系结构之间的相似与区别。”

——Kunle Olukotun, 斯坦福大学

“《计算机体系结构：量化研究方法（第 5 版）》探讨了各种并行概念和它们各自的技术权衡。和过去的几个版本一样，这一新版本中同样涵盖了最新的技术发展趋势。两个重点是个人移动设备（PMD）和仓库级计算（WSC）的爆炸性增长——与原来一味追求性能相比，这里的焦点已经转为更全面地寻求性能与能效之间的平衡。这些趋势刺激了人们不断追求更强劲的处理能力，而这种追求又推动人们在并行道路上走得更远。”

——Andrew N. Sloss, 实施顾问, ARM 公司
ARM System Developer's Guide 一书的作者

作者简介

John L. Hennessy 是斯坦福大学的第 10 任校长，从 1977 年开始在该校电子工程与计算机系任教。Hennessy 是 IEEE 和 ACM 会士，美国国家工程院、国家科学院和美国哲学院院士，美国艺术与科学院院士。他获得过众多奖项，如 2001 年度 Eckert-Mauchly 奖，表彰他对 RISC 技术的贡献；2001 年度 Seymour Cray 计算机工程奖；与 David Patterson 共同获得的 2000 年度约翰·冯·诺依曼奖章。他还拥有 7 个荣誉博士学位。

1981 年，John L. Hennessy 带领几位研究生在斯坦福开始 MIPS 项目的研究。1984 年完成该项目之后，他暂时离开大学，与他人共同筹建 MIPS 计算机系统公司（也就是现在的 MIPS 技术公司），这家公司开发了最早的商用 RISC 微处理器之一。到 2006 年，已经有 20 多亿个 MIPS 微处理器被用于视频游戏、掌上电脑、激光打印机和网络交换机等各种设备中。Hennessy 后来领导了 DASH（Director Architecture for Shared Memory，共享存储器控制体系结构）项目，这一项目设计了第一个可扩展缓存一致性多处理器原型，其中的许多重要思想都在现代多处理器中得到了应用。除了参与科研活动、履行学校职责之外，他仍作为前期顾问和投资者参与了无数的创业项目。

David A. Patterson 自 1977 年进入加州大学伯克利分校执教以来，一直讲授计算机体系结构课程，拥有该校计算机科学 Pardee 讲座教授职位。他因为教学成果显著而荣获了加州大学的杰出教学奖、ACM 的 Karlstrom 奖、IEEE 的 Mulligan 教育奖章和本科生教学奖。因为在 RISC 方面的贡献而获得了 IEEE 技术成就奖和 ACM Eckert-Mauchly 奖，他还因为在 RAID 方面的贡献而分享了 IEEE Johnson 信息存储奖，并与 John Hennessy 共同获得了 IEEE 约翰·冯·诺依曼奖章和 C & C 奖金。和 John Hennessy 相似，Patterson 也是美国艺术与科学院院士、美国计算机历史博物馆院士、ACM 和 IEEE 会士。他还被选入美国国家工程院、美国国家科学院和硅谷工程名人堂。Patterson 身为美国总统信息技术顾问委员会委员，同时也是伯克利电子工程与计算机科学系计算机科学分部主任、计算机研究协会主席和 ACM 主席。这一履历使他荣获了 ACM 和 CRA 颁发的杰出服务奖。

在加州大学伯克利分校，Patterson 领导了 RISC1 的设计与实现工作，这可能是第一台 VLSI 精简指令集计算机，为商业 SPARC 体系结构奠定了基础。他曾是廉价磁盘冗余阵列（Redundant

Arrays of Inexpensive Disks, RAID) 项目的领导者之一, 正是由于这一项目, 才有了后来许多公司出品的可靠存储系统。他还参与了工作站网络 (Network of Workstations, NOW) 项目, 因为这一项目而有了因特网公司使用的集群技术和后来的云计算。这些项目获得了 ACM 颁发的三个论文奖。作为 “算法-机器-人类” (AMP) 实验室和并行计算实验室的主管, 他目前在这里开展自己的研究项目。AMP 实验室的目标是开发可扩展的机器学习算法、适用于仓库级计算机的编程模型、能够快速洞悉云中海量数据的众包 (Crowd-Sourcing) 工具。并行计算实验室的目标是研发先进技术, 为并行个人移动设备提供可扩展、可移植、方便快捷的效率软件。

序 言

Hennessy 和 Patterson 合著的《计算机体系结构：量化研究方法》第 1 版是在我刚上研究生时出版的，因此，我属于第一批在本书指导下学习体系结构的人。要写一篇有用的序言，少不了要有作序者自己的独特观点，而我发现自己在这方面有点欠缺，因为我已深受本书前 4 个版本的影响。还有另外一个不利因素，就是我从学生时代就对这两位计算机科学巨匠心存敬畏，尽管后来我有机会与他们合作，近距离了解他们，但也可能正因如此，所以这种敬畏心现在仍未消失。不过，由于我从第一版开始就一直从事这一领域的研究，有机会看到它不断完善，欣赏它持久不变的实用性，这从一定程度上抵消了上述不利因素。

几年前，英特尔取消其 4GHz 单核 CPU 开发项目，转向多核 CPU 研发，标志着业内对更高 CPU 时钟频率的激烈竞争正式结束，两年之后，本书第 4 版出版。经过两年的充分观察，John 和 Dave 在书中将这一变化明确表述为计算技术在过去 10 年中的一个转折点，而不是一次非常随意的生产线升级。第 4 版对指令级并行（ILP）的强调有所降低，增加了线程级并行的相关内容；第 5 版则更进一步，用整整两章的篇幅来讨论线程级和数据级并行，而将 ILP 的讨论压缩为一章。新增加的第 4 章会让刚刚接触新型图形处理引擎的读者受益匪浅，这一章的重点是数据级并行，解释了一些虽有不同但正在趋于一致的解决方案，这些方案是由通用处理器中的多媒体扩展以及可编程性日益增强的图形处理器提供的。这一章还有一些非常实用的内容：如果你一直被 CUDA 术语搞得晕头转向，可以参考表 4-10。（难题：“共享存储器”实际上是本地存储器，而“全局存储器”更接近于大众认知的共享存储器。）

多核技术仍处在不断变化之中，但第 5 版还介绍了下一代重大技术：云计算。因特网的无处不在和 Web 服务的发展将人们的注意力引向两个极端：一端是超小型设备（智能手机、平板电脑），一端是超大型设备（仓库级计算系统）。第 3 章的“融会贯通”一节（3.13 节）介绍了 ARM Cortex A8，它是智能手机中的一种常用 CPU，而新增加的整个第 6 章则专门结合仓库级计算系统讨论了请求级并行和数据级并行。在这一章中，John 和 Dave 将这些新出现的大型集群看作一类特殊的新型计算机——欢迎广大计算机架构师一同来推动这一新兴领域的发展。将第 3 版中的 Google 集群体系结构与这一版第 6 章中更现代化的具体实现进行对比，读者就可以欣赏到这一领域在过去 10 年的发展过程。

本书的老读者将会再一次领略两位杰出计算机科学家的贡献，他们在整个职业生涯中，将严谨的学术研究和对前沿产品与技术的深刻理解非常完美地结合在一起，形成了一门艺术。如果你见过 Dave 如何实施他一年两度的项目调整，如何精心筹办研讨会，从校企合作之中萃取最多的精华，如果你能回想起 John 在创办 MIPS 时获得的巨大成功，或者曾经在 Google 公司的走廊里碰到他（我偶尔就会在那里遇到他），那你就不会再为两位作者与业界交流何以如鱼得水而感到惊讶了。

新老读者购买本书绝对物超所值，这可能是最重要的一点了。这本书之所以会成为不朽的经典，是因为它的每个版本都不只是更新，而是一次全面修订，力求向读者展示最新信息，以及作者对这个快速变化的迷人领域最独到的见解。对我来说，在从事这个行业 20 多年之后，它让我又一次有机会体验学生时代对这两位杰出师长的敬仰之情。

——Luiz André Barroso, Google 公司

前言

本书的目的

本书到现在已经是第 5 个版本了，我们的目标一直没有改变，就是要阐述那些为未来技术发展奠定基础的基本原理。计算机体系结构的各种发展机遇总是让我们激情澎湃，不曾有丝毫消退。我们在第 1 版中就作出过如下的论述：“这个学科不是令人昏昏欲睡、百无一用的纸版模型。绝对不是！这是一个受到人们热切关注的学科，需要在市场竞争力与成本-性能-能耗之间作好权衡，从事这个学科既可能导致可怕的失败，也可能带来显赫的成功。”

在编写第 1 版时，我们的主要目的是希望改变人们原来学习和研究计算机体系结构的方式。现今，我们感到这一目标依然正确，依然重要。该领域日新月异，在对其进行研究时，必须采用真实计算机上的测量数据和真实示例，而不是去研究一大堆从来都不需要实现的定义和设计。我们不仅热烈欢迎过去与我们结伴而行的老读者，同样也非常欢迎现在刚刚加入我们的新朋友。不管怎样，我们都保证将采用同样的量化方法对真实系统进行分析。

和前几版一样，在编写这个新版本时，我们力争使其既适用于学习高级计算机体系结构与设计课程的学生，也适用于专业的工程师和架构师。与第 1 版类似，这个版本重点介绍新平台（个人移动设备和仓库级计算机）和新体系结构（多核和 GPU）。这一版还秉承了前几版的做法，希望能够通过强调成本、性能、能耗之间的平衡和优秀的工程设计，揭去计算机体系结构的神秘面纱。我们相信这一领域正在日趋成熟，发展成为一门具备严格量化基础的经典理工学科。

关于第 5 版

我们曾经说过，第 4 版可能因为转向讨论多核芯片而成为自第 1 版以来的最重要版本。但我们收到了这样的反馈意见：第 4 版已经失去了第 1 版重点突出的优点，它一视同仁地讨论所有内容，不分重点和场合。我们非常确信，第 5 版不会再有这样的评价了。

我们相信，最令人激动的地方在于计算规模的两个极端：以移动电话和平板电脑之类的个人移动设备（PMD）为客户端，以提供云计算的仓库级计算机为服务器。（具有敏锐观察力的读者可能已经看出本书封面上云计算的寓意。）尽管这两个极端的规模大小不同，但它们在成本、性能和能效方面的共同主题给我们留下了深刻印象。因此，每一章的讨论背景都是 PMD 和仓

库级计算机的计算，第 6 章是全新的一章，专门讨论仓库级计算机。

本书的另一条主线是讨论并行的所有不同形式。我们首先在第 1 章指出了两种应用级别的并行，一个是数据级并行（DLP），它的出现是因为有许多数据项允许同时对其进行操作；另一个是任务级并行（TLP），它的出现是因为创建了一些可以独立执行并在很大程度上并行的工作任务。随后解释 4 种开发 DLP 和 TLP 的体系结构样式，分别是：第 3 章介绍的指令级并行（ILP），第 4 章介绍的向量体系结构和图形处理器（GPU），这一章是第 5 版新增加的内容；第 5 章介绍的线程级并行；第 6 章通过仓库级计算机介绍的需求级并行（RLP），这一章也是第 5 版中新增加的。本书中，我们将存储器层次结构的内容提前到第 2 章，并将存储系统那一章改作附录 D。我们对第 4 章、第 6 章的内容尤为感到自豪，第 4 章对 GPU 的解读是目前最详尽、最清晰的，第 6 章首次公布了 Google 仓库级计算机的最新细节。

与前几版相同，本书前三个附录提供了有关 MIPS 指令集系统、存储器层次结构和流水线的基础知识，如果读者没有读过《计算机组成与设计》之类的书籍，可用作参考。为了在降低成本的同时还能提供一些读者感兴趣的补充材料，我们在网络上提供了另外 9 个附录，网址为：<http://booksite.mkp.com/9780123838728>。这些附录的页数之和比本书还要多呢！

这一版继续发扬“以真实示例演示概念”的传统，并增加了全新的“融会贯通”部分。这一版中的“融会贯通”内容包括以下各服务器的流水线组成与存储器层次结构：ARM Cortex A8 处理器、Intel core i7 处理器、NVIDIA GTX-280 和 GTX-480 GPU，还有 Google 仓库级计算机。

主题的选择与组织

和以前一样，我们在选择主题时采用了一种保守的方法，毕竟这个领域中值得讨论的思想实在太多了，不可能在这样一本主要讨论基本原理的书将其全部涵盖在内。我们没有面面俱到地分析读者可能遇到的所有体系结构，而是将重点放在那些在任何新计算机中都可能涉及的核心概念上。根据一贯坚持的选材标准，本书讨论的思想都经过深入研究并已被成功应用，其内容足以采用量化方法进行讨论。

我们一直重点关注的内容都是无法从其他来源获取的同类资料，因此我们将继续尽可能讨论比较高级的内容。事实上，本书介绍的有些系统，就无法在文献中找到相关描述。如果读者需要了解更为基础的计算机体系结构知识，可以阅读《计算机组成与设计：硬件/软件接口》（*Computer Organization and Design: The Hardware/Software Interface*）一书。

内容概述

这一版对第 1 章进行了补充，其中包括能耗、静态功率、动态功率、集成电路成本、可靠

性和可用性的计算公式。(封二上也列出了这些公式。)在本书后续部分读者能够一直应用这些公式。除了计算机设计与性能测量方面的经典量化原理之外,还对 PIAT 一节进行了升级,采用了新的 SPECPower 基准测试。

我们认为,与 1990 年相比,指令集体系结构扮演的角色有所弱化,所以我们把这一部分内容作为了附录 A。它仍然采用 MIPS64 体系结构。(为便于快速查看,封三汇总了 MIPS ISA 相关信息。)网站上的附录 K 介绍了 10 种 RISC 体系结构、80x86、DEC VAX 和 IBM 360/370,献给 ISA 爱好者们。

随后,我们在第 2 章开始讨论存储器层次结构,这是因为很容易针对这些内容应用成本-性能-功耗原理,而且存储器是其余各章的关键内容。和上一版一样,附录 B 对缓存机制作了概述,以供读者需要时查阅。第 2 章讨论了对缓存的 10 种高级优化方法。这一章还介绍了虚拟机,它便于提供保护、进行软硬件管理,而且在云计算中也扮演着重要角色。除了介绍 SRAM 和 DRAM 技术之外,这一章还包括了闪存的内容。PIAT 示例选择了 PMD 中使用的 ARM Cortex A8 和服务中使用的 Intel Core i7。

第 3 章主要研究高性能处理器中的指令级并行开发,包括超标量执行、分支预测、推理、动态调度和多线程。前面曾经提到,附录 C 是关于流水线的综述,以备随时查阅之用。第 3 章还研究了 ILP 的局限性。和第 2 章一样,PIAT 示例还是 ARM Cortex A8 和 Intel Core i7。第 3 版包括大量有关 Itanium 和 VLIW 的材料,现在这些内容放在网上的附录 H 中,这表明了我们的观点:这种体系结构未能达到过去所宣称的效果。

多媒体应用程序(比如游戏和视频处理)的重要性在提高,因此,开发数据级并行的体系结构也变得更为重要。具体来说,越来越多的人在关注利用图形处理器(GPU)执行的运算,但很少有架构师了解 GPU 到底是如何工作的。我们决定编写新的一章,主要就是为了揭开这种新型计算机体系结构的奥秘。第 4 章开始介绍向量体系结构,对多媒体 SIMD 指令集扩展和 GPU 的解释就是以此为基础的。(网站上的附录 G 深入地讨论了向量体系结构。)GPU 一节是本书最难写的部分,需要多次反复才能给出一个既精确又容易理解的描述。一个重大挑战就是术语。我们决定使用我们自己的术语,然后给出这些术语与 NVIDIA 官方术语之间的对应关系。这一章介绍了 Roofline 性能模型,然后用它来对比 Intel Core i7、NVIDIA GTX 280 和 GTX 480 GPU。这一章还介绍了供 PMD 使用的 Tegra 2 GPU。

第 5 章介绍多核处理器,探讨了对称、分布式存储器体系结构,考查了组织原理和性能。接下来是有关同步和存储器一致性模型的主题,所采用的示例是 Intel Core i7。对片上互连网络感兴趣的读者可以阅读网站上的附录 F,对更大规模多处理器和科学应用感兴趣的读者可以阅读网站上的附录 I。

前面曾经提到,第 6 章介绍了计算机体系结构中的最新主题——仓库级计算机(Warehouse-Scale Computer, WCS)。依靠 Amazon Web 服务部门和 Google 工程师的帮助,本章整合了有关

WSC 设计、成本与性能的详细资料，而以前了解这些内容的架构师寥寥无几。在开始描述 WSC 的体系结构和物理实现（及成本）之前，首先介绍了 MapReduce 编程模型。从成本的角度可以解释为什么会有云计算，以及为何在云中使用 WSC 进行计算的成本要低于在本地数据中心的计算成本。PIAT 实例是对 Google WSC 的描述，有些内容是首次公开的。

接下来就是附录 A 到附录 L。^①附录 A 介绍 ISA 的原理，包括 MIPS64，附录 K 介绍 Alpha、MIPS、PowerPC 和 SPARC 的 64 位版本及其多媒体扩展。其中还包括一些经典体系结构（80x86、VAX 和 IBM 360/370）和流行的嵌入指令集（ARM、Thumb、SuperH、MIPS16 和 Mitsubishi M32R）。附录 H 与其相关，介绍了 VLIW ISA 的体系结构和编译器。

前面曾经提到，附录 B 和附录 C 是缓存与流水线基本概念的教程。建议对缓存不够熟悉的读者在阅读第 2 章之前先阅读附录 B，新接触流水线的读者在阅读第 3 章之前先阅读附录 C。

附录 D “存储系统”包括：进一步讨论可靠性和可用性，以 RAID 6 方案介绍为主体的 RAID 教程，非常珍贵的真实系统故障统计信息。接下来介绍了排队理论和 I/O 性能基准测试。我们评估了一个真实集群 Internet Archive 的成本、性能和可靠性。“融会贯通”部分以 NetApp FAS6000 文件管理程序为例。

附录 E 由 Thomas M. Conte 撰写，汇总了嵌入式系统的相关内容。

附录 F 讨论网络互连，由 Timothy M. Pinkston 和 José Duato 进行了修订。附录 G 最初由 Krste Asanović 撰写，其中详细介绍了向量处理器。就我们所知，这两个附录是其各自相关主题的最好材料。

附录 H 详细介绍了 VLIW 和 EPIC，也就是 Itanium 采用的体系结构。

附录 I 详细介绍了大规模共享存储器多处理方面用到的并行处理应用和一致性协议。附录 J 由 David Goldberg 撰写，详细介绍了计算机算法。

附录 L 将第 3 版每一章中的“历史回顾与参考文献”部分集中在一起。对于各章介绍的思想，它尽量给予一个恰当的评价，并让读者了解这些创造性思想背后的历史。我们希望以此来展现人类在计算机设计方面的戏剧性发展过程。这个附录还提供了一些参考文献，主修体系结构的学生可能会非常喜欢它们。其中提到了本领域的一些经典论文，如果时间允许，建议读者阅读这些论文。直接听原创者讲述他们的思想，在深受教育的同时，也是一种享受。而“历史回顾”是以前版本中最受欢迎的章节之一。

内容导读

所有读者都应当从第 1 章开始阅读，除此之外并不存在什么唯一的最佳顺序。如果你不想阅读全部内容，可以参考下面这些顺序。

^① 本书中文版未收录附录 D 到附录 L，这些内容可在英文书网站 <http://booksite.mkp.com/9780123838728> 获取。

——编者注

- 存储器层次结构：附录 B、第 2 章、附录 D。
- 指令级并行：附录 C、第 3 章、附录 H。
- 数据级并行：第 4 章、第 6 章、附录 G。
- 线程级并行：第 5 章、附录 F、附录 I。
- 请求级并行：第 6 章。
- ISA：附录 A、附录 K。

附录 E 可以随时阅读，但在 ISA 和缓存序列之后阅读，效果可能会更好一些。附录 J 可以在涉及运算时阅读。附录 L 的各部分内容应当在读完正文中相应章节后阅读。

章节安排

我们根据一种统一的框架安排内容，使各章在结构方面保持一致。首先会介绍一章的主题思想，然后是“交叉问题”部分，说明本章介绍的思想与其他各章有什么相互关系。接下来是“融会贯通”部分，通过展示如何在实际计算机中应用这些思想，将它们串在一起。

再下面是“谬论与易犯错误”，让读者从他人的错误中汲取教训。我们将举例说明一些常见误解与体系结构陷阱，要避免犯错是非常困难的，哪怕你明明知道它们就在前面等着你。“谬论与易犯错误”部分是本书最受欢迎的内容。每一章都以一个“结语”节结束。

案例研究与练习

每一章的最后都有案例研究和练习。这些案例研究由业内和学术界的专家编撰而成，通过难度逐渐增大的练习来探讨该章的关键概念，检验读者的理解程度。教师们会发现这些案例研究都非常详尽和完善，完全可以针对它们设计出一些练习。

每个练习中用尖括号括起的内容（<章节>）指明了做这道题应该阅读哪部分正文内容。我们这样做的目的，一方面是为了提供复习内容，另一方面是希望帮助读者避免在还没有阅读相应正文的情况下去做一些练习。为了使读者大致了解完成一道题需要多长时间，我们为这些练习划定了不同等级：

- [10] 短于 5 分钟（阅读和理解时间）；
- [15] 5~15 分钟给出完整答案；
- [20] 15~20 分钟给出完整答案；
- [25] 在 1 小时内给出完整的书面答案；
- [30] 小型编程项目：时间短于 1 整天；
- [40] 大型编程项目：耗时 2 周；
- [讨论] 与他人一起讨论的主题。

在 textbooks.elsevier.com 注册的老师可以得到案例研究与习题的解答。

补充材料

我们还通过网络提供了多种资料，网址为 <http://booksite.mkp.com/9780123838728/>，内容包括：

- 参考附录——涵盖了一系列高级主题，由相关领域的专家撰写；
- 历史材料，考察了正文各章所介绍的关键思想的发展形成过程；
- 供老师使用的 PowerPoint 幻灯片；
- PDF、EPS 和 PPT 格式的书中插图；
- 网上相关材料的链接；
- 勘误表。

我们会定期补充新材料和网上其他可用资源的链接。

帮助改进本书

如果你阅读后面的“致谢”部分，将会看到我们已经下了很大的功夫来纠正错误。由于一本书会进行多次印刷，所以我们有机会进行更多的校订。如果你发现了任何遗留错误，请通过电子邮件联系出版商（ca5bugs@mkp.com）。^①

我们欢迎你对本书给出其他意见，请将它们发送到另一个电子信箱：ca5comments@mkp.com。

结语

本书仍然是一本真正的合著作品，我们每人编写的章节和附录各占一半。如果没有对方完成另一半工作，如果没有对方在任务似乎无望完成时给予鼓励，如果没有对方点透某个难以表述的复杂概念，如果没有对方花费周末时间来审阅书稿，又如果没有对方在自己因为其他繁重职责而难以提笔时给予宽慰（从简历可以看出，这些职责是随着本书的版本号以指数形式增加的），我们无法想象这本书要花费多长时间才能完成。当然，对于你将要读到的内容，其中若有不当之处，我们也负有同等责任。

John Hennessy

David Patterson

^① 读者可以到图灵社区本书主面（www.ituring.com.cn/book/888）提交中译本勘误。——编者注

致 谢

尽管本书仅正式发布了 5 个版本，但我们实际上已经写出过 10 个不同版本：第 1 版有 3 个版本（alpha 版、beta 版和最终版），第 2、3、4 版各有 2 个版本（beta 版和最终版）。一路走来，我们得到了数百位审阅者和用户的帮助，他们每一位都让这本书变得更好。因此，我们决定列出所有对本书各版本作出贡献的人员名单。

第 5 版的贡献者

和前几版一样，第 5 版也是一个有许多志愿者参与的集体成果。没有这些志愿者的帮助，这一版就不可能保持一贯的品质。

审阅者

南卡罗来纳大学的 Jason D. Bakos、加州大学圣巴巴拉分校的 Diana Franklin、HP 实验室的 Norman P. Jouppi、田纳西大学的 Gregory Peterson、HP 实验室的 Parthasarathy Ranganathan、克莱姆森大学的 Mark Smotherman、威斯康星大学麦迪逊分校的 Gurindar Sohi、西班牙加泰罗尼亚理工大学的 Mateo Valero 以及新泽西理工学院的 Sotirios G. Ziavras。

加州大学伯克利分校并行实验室和 RAD 实验室的成员（他们多次审阅第 1、4、6 章，并使我们对 GPU 和 WSC 的解释部分得以成形）：Krste Asanović、Michael Armbrust、Scott Beamer、Sarah Bird、Bryan Catanzaro、Jike Chong、Henry Cook、Derrick Coetzee、Randy Katz、Yunsup Lee、Leo Meyervich、Mark Murphy、Zhangxi Tan、Vasily Volkov 以及 Andrew Waterman。

顾问团

Google 公司的 Luiz André Barroso、R&E Colwell & Assoc. 公司的 Robert P. Colwell、ARM 公司研发副总裁 Krisztian Flautner、宾州州立大学的 Mary Jane Irwin、NVIDIA 公司的 David Kirk、Tensilica 首席科学家 Grant Martin、威斯康星大学麦迪逊分校的 Gurindar Sohi 以及西班牙加泰罗尼亚理工大学的 Mateo Valero。

附录

加州大学伯克利分校的 Krste Asanović (附录 G)、北卡罗来纳州立大学的 Thomas M. Conte (附录 E)、西班牙巴伦西亚理工大学的 José Duato (附录 F)、施乐帕洛阿尔托研究中心的 David Goldberg (附录 J)、南加州大学的 Timothy M. Pinkston (附录 F)。

瓦伦西亚理工大学的 José Flich 为附录 F 的更新作出了重大贡献。

案例研究与练习

南卡罗来纳大学的 Jason D. Bakos (第 3 章和第 4 章)、加州大学圣巴巴拉分校的 Diana Franklin (第 1 章和附录 C)、HP 实验室的 Norman P. Jouppi (第 2 章)、HP 实验室的 Naveen Muralimanohar (第 2 章)、田纳西大学的 Gregory Peterson (附录 A)、HP 实验室的 Parthasarathy Ranganathan (第 6 章)、圣克拉拉大学的 Amr Zaky (第 5 章和附录 B)。

Jichuan Chang、Kevin Lim 和 Justin Meza 帮助制定和测试了第 6 章的案例研究与练习。

补充材料

NVIDIA 公司的 John Nickolls、Steve Keckler 和 Michael Toksvig (第 4 章 NVIDIA GPU), Intel 公司的 Victor Lee (第 4 章 Core i7 与 GPU 的对比), 美国劳伦斯伯克利国家实验室 (LBNL) 的 John Shalf (第 4 章最新向量体系结构), LBNL 的 Sam Williams (第 4 章中的 Roofline 计算机模型), 澳大利亚国立大学的 Steve Blackburn 和得克萨斯大学奥斯汀分校的 Kathryn McKinley (第 5 章的 Intel 性能与功耗测量), Google 公司的 Luiz Barroso、Urs Hölzle、Jimmy Clidaris、Bob Felderman 和 Chris Johnson (第 6 章的 Google WSC), Amazon Web 服务部的 James Hamilton (第 6 章的功率分配与成本模型)。

南卡罗来纳大学的 Jason D. Bakos 为这一版制作了新的授课幻灯片。

最后, 要再次特别感谢克莱姆森大学的 Mark Smotherman, 他对我的最终手稿进行了技术审阅, 发现了大量错误和含糊不清的地方, 使本书的表述清晰了许多。

当然, 没有出版商, 本书也不可能出版, 所以我们要感谢 Morgan Kaufmann/Elsevier 全体员工的努力和支持。关于第 5 版, 我们特别要感谢本书的编辑 Nate McFadden 和 Todd Green, 他们协调了从开展调查、组织顾问团、制定案例研究与练习、组成焦点小组、手稿审读到附录更新的整个过程。

我们还得感谢我们学校的员工 Margaret Rowland 和 Roxana Infante, 在编写本书期间, 她们为我们处理了无数的快递邮件, 帮助我们化解了斯坦福和伯克利的许多紧急情况。

最后要感谢我们的妻子, 感谢她们容忍我们越来越早起进行阅读、思考和写作。