

高性能微处理器：技术与结构

张民选 王永文 编著

国防科技大学出版社

国防科技大学学术
专著专项经费资助出版

高性能微处理器： 技术与结构

张民选 王永文 编著

TP332
248

国防科技大学出版社
湖南·长沙

图书在版编目(CIP)数据

高性能微处理器:技术与结构/张民选,王永文编著. —长沙:国防科技大学出版社,2004.3

ISBN 7-81099-061-6

I. 高… II. 张… III. 微处理器 IV. TP332

中国版本图书馆 CIP 数据核字(2004)第 017816 号

国防科技大学出版社出版发行

电话:(0731)4572640 邮政编码:410073

E-mail: gfkdebs@public.cs.hn.cn

责任编辑:肖滨 罗青 责任校对:罗青 肖滨

新华书店总店北京发行所经销

国防科技大学印刷厂印装

*

开本:850×1168 1/32 印张:16.375 字数:411千

2004年3月第1版第1次印刷 印数:1-2000册

ISBN 7-81099-061-6/TP·5

定价:43.00元

前言

微处理器(Microprocessor)是仅由一片超大规模集成电路组成的中央处理器。在三十年的发展过程中,微处理器的性能以每年近 30% 的速度增长。伴随着性能的提高,其应用在各个领域取得了巨大的成功和发展,产生了良好的经济效益和社会效益。应用的发展要求微处理器提供更高的性能,推动了微处理器技术与结构的研究不断深入。

《高性能微处理器:技术与结构》一书就是在这样一个背景下编写的。

目的

本书希望能全面地讨论高性能微处理器的关键技术和体系结构,以总结微处理器发展的一般规律,并分析未来的设计方向。

近年来,伴随着微型计算机的普及,微处理器引起了广大科技人员的关注。国内对微处理器的介绍和评论也越来越多,它们对普及微处理器的基本知识起到了一定的作用,然而对专业研究人员来讲,却不够全面和深入,甚至影响了初学者对微处理器的整体认识。本书编写的第一个目的,就是希望能够全面地讨论高性能微处理器的方方面面,给读者提供一个微处理器的整体概念。

进入 20 世纪 90 年代,随着 RISC 技术的流行,国内出版了很多关于 RISC 微处理器的著作。然而微处理器的发展已进入后 RISC 时期,更多、更新、更好的技术和产品不断涌现,微处理器的研究进入了新的阶段。本书编写的第二个目的就是消化和吸收最新的研究成果。

作者在计算机体系结构和微处理器设计领域从事教学和科研

工作已经有了很长的一段时间,其间遇到了很多的困难,也积累了一些资料和经验,希望能够总结一下近些年来研究成果和经验教训,也为感兴趣的读者提供一些有价值的参考资料。这是本书编写的第三个目的。

主题

《高性能微处理器:技术与结构》的书名就是本书的主题。所谓技术,是指在劳动生产方面的经验、知识和技巧,也泛指其他方面的操作技巧。而结构,是指组成整体的各部分的搭配和安排。

我们这里讨论的技术是指提高微处理器性能的经验 and 技巧,它们大多是为了解决某一个特定问题而提出的,比如分支处理技术、指令调度技术、降低延迟技术等。当然,每一个问题都有多种解决方法,我们把相同目的的技术放在一起介绍,并予以分析讨论。

我们这里讨论的结构是指微处理器的体系结构。经典的关于“计算机体系结构”的定义是 1964 年 C.M.Amdahl 在介绍 IBM360 系统时提出的:计算机体系结构是程序员所看到的计算机的属性,即概念性结构与功能特性。微处理器作为计算机的心脏,其体系结构包含两个层次:即指令集体系结构和微体系结构。前者指的是微处理器的指令系统、数据表示、寄存器定义等,这是硬件与软件分界的界面。后者是指微处理器的实现结构,比如超标量结构、流水线结构等。本书把指令集体系结构作为微处理器设计的基础,而重点讨论微体系结构。

内容

本书共 12 章,分为三大部分。

第一部分为基础篇,介绍高性能微处理器的基础知识和基本原理,分为 4 章。

第 1 章微处理器发展概论:概述了微处理器发展的历程,总结了微处理器发展的动力、目标和手段。

第 2 章微处理器性能评价:介绍微处理器性能评价的一般模

型和基准程序等知识。

第3章微处理器中的并行性:介绍微处理器中各种并行性以及实现要素。

第4章微处理器指令集系统:介绍指令系统设计的一般原则和相关知识。

第二部分为技术篇,介绍提高微处理器性能的经验 and 技巧,分为3章。

第5章指令调度:介绍调度指令,开发指令级并行性的软硬件方法。

第6章分支处理:介绍克服分支开销的软硬件技术。

第7章存储器访问延迟处理:介绍降低访存延迟,提高程序执行效率的软硬件方法。

第三部分为结构篇,介绍微处理器的微体系结构,分为5章。

第8章流水线结构:介绍一般流水线和超流水线的结构和实例。

第9章向量处理器与 SIMD 处理器:介绍向量处理器和 SIMD 处理器的典型结构和通用微处理器的扩展。

第10章 ILP 处理器:介绍超标量、超长指令字、显式并行指令计算的典型结构和实例。

第11章多线程与多处理器芯片:介绍开发更高级并行性的典型结构和实例。

第12章微处理器发展展望:分析微处理器的发展方向。

本书最后列出了参考文献和术语表。

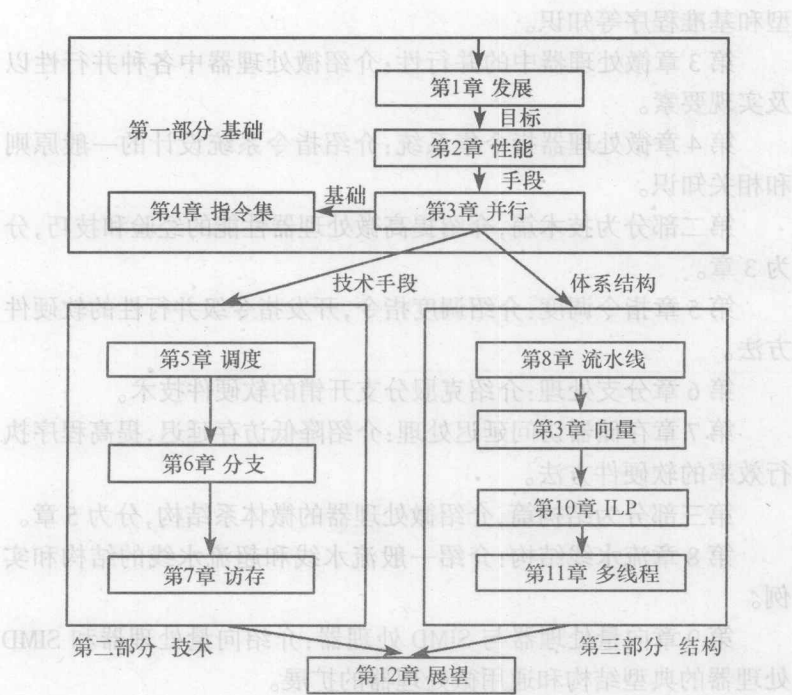


图 0-1 全书导读图

读者

本书适用于具备计算机高级语言程序设计、计算机组成原理、计算机体系结构和编译原理的基础知识,并希望对微处理器做进一步了解的读者。不具备上述知识的读者可以先阅读第一部分,但在阅读第二、三部分时可能会遇到困难,请参看相关资料。图 0-1 给出了本书的导读。

本书的编写得到了欧国东、张承义、孙彩霞、高军、李瞰、王蕾、鲁健壮、马卓、邓晴莺、张明、屈双喜、焦永、宋鑫等的帮助,在此表示深切谢意。

本书的编写得到国家自然科学基金(60273069)、国家 863 计划(2002AA11101)的资助。

由于作者的水平和经验有限,书中必有错误和不当之处,恳请读者批评指正。

作者
2003 年 9 月

目 录

I 基础篇

第一章 微处理器发展概论

1.1	简介	(3)
1.1.1	微处理器的分类	(3)
1.1.2	微处理器的分代	(5)
1.2	微处理器出现以前	(7)
1.2.1	算盘	(7)
1.2.2	分析机	(9)
1.2.3	ENIAC	(10)
1.2.4	冯·诺依曼结构	(11)
1.2.5	UNIVAC - I	(12)
1.3	CISC 时期	(12)
1.3.1	微处理器的诞生	(13)
1.3.2	CISC 微处理器的繁荣	(16)
1.3.3	CISC 设计思想	(19)
1.3.4	指令使用的研究统计	(21)
1.4	RISC 时期	(22)
1.4.1	RISC 的提出	(22)
1.4.2	高性能 RISC 微处理器	(23)
1.4.3	RISC 的设计思想	(27)
1.4.4	CISC 的继续发展	(29)

1.5	后 RISC	(31)
1.5.1	后 RISC 时期的特点	(31)
1.5.2	后 RISC 时期的新型结构	(33)
1.5.3	后 RISC 时期的 CISC 与 RISC	(35)
1.6	微处理器带来的冲击	(36)
1.6.1	微型计算机的繁荣	(36)
1.6.2	工作站的发展	(37)
1.6.3	用商品化微处理器构造超级计算机	(38)
1.6.4	微机和 workstation 互连集群	(39)
1.7	微处理器发展的动力	(40)
1.7.1	需求动力	(40)
1.7.2	技术动力	(41)
1.7.3	产业动力	(41)
1.8	微处理器体系结构设计的重要思想	(42)
1.8.1	微处理器设计的影响因素	(42)
1.8.2	系列化思想	(43)
1.8.3	分层思想	(44)
1.8.4	子集与超集的思想	(47)
1.8.5	几个对立的观点	(49)
1.9	小结	(52)

第二章 微处理器性能评价

2.1	简介	(53)
2.1.1	时间度量	(54)
2.1.2	性能评估和测试的方法	(55)
2.1.3	性能测试策略	(56)
2.2	简单的性能模型	(56)

(101)	2.2.1	计算机的性能比较	(56)
(101)	2.2.2	加速比	(57)
(101)	2.2.3	CPU 性能公式	(59)
(101)	2.2.4	考虑访存延迟的 CPU 性能公式	(61)
(101)	2.3	基准程序	(63)
(101)	2.3.1	MIPS 和 MFLOPS 指标	(64)
(101)	2.3.2	基准程序的由来	(66)
(101)	2.3.3	基准程序的分类	(68)
(101)	2.3.4	基准程序的用途	(71)
(101)	2.4	SPEC 基准程序组	(74)
(101)	2.4.1	SPEC 概述	(74)
(101)	2.4.2	SPEC CPU 基准程序组	(76)
(101)	2.4.3	实际应用基准测试	(90)
(101)	2.5	模拟器	(92)
(101)	2.5.1	常用模拟器	(93)
(101)	2.5.2	SimpleScalar 简介	(94)
(101)	2.5.3	SimpleScalar 的组成	(95)
(101)	2.6	小结	(97)

第三章 微处理器中的并行性

(101)	3.1	简介	(98)
(101)	3.1.1	时间并行和空间并行	(98)
(101)	3.1.2	控制并行和数据并行	(99)
(101)	3.1.3	并行性的实现要素	(100)
	3.1.4	并行的级别	(100)
	3.2	比特级并行	(101)
(101)	3.2.1	字长	(101)
(101)	3.2.2	64 位高性能微处理器	(102)

3.3	标量级并行	(103)
3.3.1	向量处理方式	(103)
3.3.2	标量级并行处理器	(104)
3.4	指令级并行	(105)
3.4.1	基本原理	(105)
3.4.2	指令并行度和处理器并行度	(106)
3.4.3	ILP 体系结构	(107)
3.5	线程级并行	(109)
3.5.1	多线程原理	(109)
3.5.2	线程级并行体系结构	(110)
3.6	程序中可用的并行性	(110)
3.6.1	相关性的限制	(110)
3.6.2	不同程序的 ILP 的差异	(114)
3.6.3	ILP 极限研究	(115)
3.7	并行性的析取	(118)
3.7.1	建立指令窗口	(119)
3.7.2	确定和消除相关	(120)
3.7.3	调度指令	(121)
3.8	微处理器的硬件支持	(122)
3.8.1	传统机制	(122)
3.8.2	寄存器文件	(123)
3.8.3	指令发射部件	(124)
3.8.4	动态前瞻	(125)
3.9	小结	(126)

第四章 微处理器指令集结构

4.1	指令的构成	(128)
4.1.1	操作数	(129)

4.1.2	操作码	(130)
4.1.3	寻址技术	(131)
4.2	指令集分类	(132)
4.2.1	分类依据	(132)
4.2.2	通用寄存器型指令集结构	(133)
4.2.3	通用寄存器型指令集结构的分类	(134)
4.3	指令行为研究	(136)
4.3.1	指令使用频率	(136)
4.3.2	数据类型的访问统计	(137)
4.3.3	控制指令	(138)
4.3.4	寻址方式的研究	(139)
4.3.5	立即数的范围	(140)
4.3.6	偏移量的统计	(142)
4.4	指令兼容技术	(143)
4.4.1	典型二进制兼容技术	(144)
4.4.2	兼容的指令集体系结构	(145)
4.4.3	软件解释	(145)
4.4.4	微代码仿真器	(145)
4.4.5	二进制翻译	(146)
4.4.6	本地编译器	(146)
4.5	多媒体指令扩展	(147)
4.5.1	指令扩展概述	(147)
4.5.2	子字算术指令	(148)
4.5.3	数据重整指令	(149)
4.5.4	格式化指令	(150)
4.5.5	条件指令	(152)
4.5.6	复杂指令	(152)
4.5.7	存储器指令	(153)

4.6	字节顺序与对齐	(153)
4.6.1	字节顺序问题	(154)
4.6.2	对齐问题	(158)
4.7	IA-64 指令系统	(159)
4.7.1	指令类型	(159)
4.7.2	指令功能	(160)
4.7.3	指令模板	(163)
4.8	小结	(164)

II 技术篇

第五章 指令调度

5.1	概述	(167)
5.1.1	调度的作用	(168)
5.1.2	调度分类	(171)
5.2	程序的表述和执行	(171)
5.2.1	程序的表示	(171)
5.2.2	程序表述顺序	(172)
5.2.3	控制驱动表述	(173)
5.2.4	数据驱动表述	(174)
5.2.5	控制驱动任务中的数据驱动表述	(175)
5.2.6	程序的执行顺序	(175)
5.3	调度所需的程序信息	(176)
5.3.1	程序基本块划分	(177)
5.3.2	基本块的数据流图	(179)
5.3.3	优先图	(180)
5.3.4	关键路径的概念	(182)
5.3.5	资源保留表	(183)

5.4	基本块内的软件调度算法	(184)
5.4.1	最优调度的开销	(184)
5.4.2	有向无环图 DAG 指令调度	(185)
5.4.3	线性压缩算法指令调度	(186)
5.4.4	搜索树算法	(187)
5.4.5	列表调度	(187)
5.4.6	调度顺序的影响	(192)
5.4.7	其他的调度方法	(194)
5.5	跨越基本块的软件调度方法	(196)
5.5.1	循环展开	(196)
5.5.2	踪迹调度	(197)
5.5.3	过滤调度	(199)
5.5.4	超块调度	(200)
5.5.5	甚块调度	(200)
5.5.6	软件流水	(200)
5.5.7	推进	(203)
5.6	调度器和编译器的关系	(204)
5.6.1	寄存器分配和调度的交互	(204)
5.6.2	编译时调度和编译后调度	(206)
5.7	判断并消除相关性	(208)
5.7.1	静态消除存储器地址歧义	(209)
5.7.2	软件寄存器重命名	(210)
5.7.3	归约变量扩展	(210)
5.8	记分牌	(211)
5.8.1	记分牌的结构	(212)
5.8.2	指令执行的步骤	(213)
5.8.3	数据结构	(214)
5.8.4	记分牌控制	(215)

5.8.5	记分牌评价	(216)
5.9	Tomasulo 算法	(218)
5.9.1	Tomasulo 算法的思想	(218)
5.9.2	Tomasulo 算法的流水线	(220)
5.9.3	保留站的内容	(222)
5.9.4	Tomasulo 算法的评价	(224)
5.10	硬件前瞻执行	(226)
5.10.1	重定序缓冲区	(226)
5.10.2	前瞻执行的步骤	(228)
5.10.3	硬件前瞻评价	(230)
5.11	其他硬件调度方法	(231)
5.11.1	寄存器更新部件	(231)
5.11.2	分派栈	(232)
5.11.3	DRIS	(232)
5.11.4	分离执行	(233)
5.12	小结	(234)

第六章 分支处理

6.1	概述	(236)
6.1.1	分支对单指令流水线的影响	(237)
6.1.2	分支对 ILP 的影响	(237)
6.1.3	程序中分支的行为特点	(238)
6.1.4	如何处理分支	(240)
6.2	分支消除	(241)
6.2.1	循环展开	(242)
6.2.2	循环剥落	(242)
6.2.3	函数内联	(242)
6.2.4	条件执行	(243)

6.3	分支延迟	(243)
6.3.1	分支延迟槽	(244)
6.3.2	快速比较	(246)
6.3.3	分支压缩	(248)
6.3.4	分支扩散	(249)
6.3.5	分支折叠	(250)
6.4	静态分支预测	(251)
6.4.1	预测分支成功	(252)
6.4.2	预测分支失败	(252)
6.4.3	根据操作码预测	(252)
6.5	动态分支预测	(253)
6.5.1	分支标志预测	(253)
6.5.2	一级分支预测	(234)
6.5.3	两级分支预测	(259)
6.5.4	两级自适应分支预测的改进	(264)
6.5.5	前瞻(Look-Ahead)	(265)
6.6	分支目标缓冲区	(265)
6.7	其他分支处理技术	(271)
6.7.1	循环缓冲区	(271)
6.7.2	共享流水线多处理器	(272)
6.7.3	预取分支目标	(272)
6.7.4	数据预取目标	(272)
6.7.5	准备分支	(273)
6.7.6	多指令流	(273)
6.8	Itanium 处理器的分支处理器机制	(274)
6.8.1	推测执行	(274)
6.8.2	动态预测	(277)
6.8.3	静态分支预测	(280)