

国内首部片上网络领域学术专著
国防科技大学王志英教授领衔撰写，全景呈现其科研团队
国际领先的研究方法和研究成果

片上网络 原理与设计

王志英◎主编

马胜 黄立波 赖明澈 石伟 王鹏◎著

The Principle
and Design
of Networks on Chip



机械工业出版社
China Machine Press

片上网络

原理与设计

王志英◎主编

马胜 黄立波 赖明澈 石伟 王鹏◎著

The Principle and Design
of Networks on Chip



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

片上网络原理与设计 / 王志英主编. —北京: 机械工业出版社, 2016.12
(计算机学科前沿技术系列)

ISBN 978-7-111-55516-2

I. 片… II. 王… III. 计算机体系结构 IV. TP303

中国版本图书馆 CIP 数据核字 (2016) 第 295926 号

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 曲 熠

责任校对: 殷 虹

印 刷: 三河市宏图印务有限公司

版 次: 2017 年 1 月第 1 版第 1 次印刷

开 本: 186mm×240mm 1/16

印 张: 22.25

书 号: ISBN 978-7-111-55516-2

定 价: 99.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

前 言

P r e f a c e

半导体工艺的进步使得集成电路的晶体管数目不断增加,设计复杂单核处理器面临的功耗急剧上升和收益逐渐下降等问题导致研究人员转向设计多核处理器。当前处理器已经集成了数十甚至上百个计算核,业界已进入众核处理器时代,但是众核处理器设计依然面临着从底层逻辑实现到高层编程模式等方面的关键挑战。众核处理器的正确性、性能和效率都取决于处理器内部的通信机制,有效解决众核处理器面临的众多挑战需要采用以通信为核心的跨层次优化方法。

传统的片上总线通信机制存在带宽较低、延迟较大、功耗较高、可扩展性较差等缺陷。为避免这些缺陷,片上网络将报文交换思想引入芯片内部通信机制中,由于它具备许多优良的特性,因此迅速成为系统结构领域的一个重要研究方向。尽管片上网络与片外网络具有一定的相似性,但是两者在延迟、功耗和面积等方面有着本质区别。片上网络要与计算核一起竞争芯片宝贵的功耗和面积资源,这使得片上网络只能利用有限的功耗和面积资源。为了在有限的功耗和面积开销下获得较高的性能,设计者应该更加关注片上网络的优化,包括底层逻辑实现的优化、网络层路由和流控的优化以及面向高层编程模式的优化。

我们课题组从事计算机系统结构前沿领域的研究,在片上网络方向已有十余年的耕耘,在许多国际知名期刊和会议上发表了数十篇具有影响力的论文。本书对课题组发表的较具影响力的片上网络相关论文进行了回顾和总结,包括在顶级和权威会议 ISCA、HPCA、DAC、ICCD、ASAP、ASP-DAC 发表的 7 篇论文,以及在顶级和权威期刊《IEEE Transactions on Computers》《IEEE Transactions on Parallel and Distributed Systems》《ACM Transactions on Architecture and Code Optimization》《Microprocessors and Microsystems》发表的 6 篇论文。

本书并不是对课题组研究成果和设计经验的简单堆叠,而是对内容进行了精心组织,自底向上涵盖了从底层路由器、缓存和拓扑结构的实现,到网络层路由算法

和流控机制的设计,再到片上网络与高层并行编程模式的协同优化等内容。本书包括五个部分,第一部分对本书内容进行简介,第五部分对全书内容进行总结并展望了未来工作,中间三部分是本书的主体部分。

第二部分关注底层逻辑实现,由第2~4章组成。第2章讨论了一种基于翼通道的单周期路由器结构,它能显著降低网络延迟。第3章研究了两种具备拥塞感知能力的动态虚通道结构,这些结构能够在不同虚通道或者不同物理端口之间共享缓存资源。第4章介绍了一种扩展虚拟总线的片上网络拓扑,这种拓扑结构能高效支持重要的多播和广播通信。

第三部分讨论路由算法和流控机制,包括第5~8章。第5章设计面向负载整合模式的路由算法,该路由算法能为多个同时运行的应用程序提供较高适应性和动态隔离性。第6章提出了一种面向完全自适应路由算法的流控机制,它能显著提升缓存资源的利用率。第7章讨论了面向Torus网络的流控机制,所提出的切片气泡流控在低开销条件下获得了较高的性能。第8章研究高效能和公平性流控机制,提出的基于当值缓存的流控机制可有效支持对缓存进行电源门控,同时提出了考虑报文传输公平性的流控机制。

第四部分研究片上网络和高层并行编程模式的协同优化,包括第9~11章。第9章为cache一致性协议中的聚合通信提供硬件支持以防止这些通信成为性能瓶颈。第10章面向消息传递编程模式定制片上网络,为消息传递基本原语提供了专用硬件实现。第11章研究了一种自适应消息传递通信协议,该协议兼具缓存通信模式和同步通信模式的优点。

总之,基于以通信为核心的跨层次优化方法,本书内容显著推动了片上网络和多核处理器研究领域的进步。本书以一种自底向上的方式全面探索了片上网络设计空间,研究成果涵盖片上网络领域的大量研究主题,不仅提升了众核处理器通信层的性能并降低了硬件开销,还有效缓解了众核处理器逻辑实现层和并行编程模式层面临的问题。

本书由王志英主编、策划和统筹,写作分工如下:第1章、第5章、第6章、第7章、第9章和第12章由马胜撰写,第2章由赖明澈撰写,第3章由石伟和赖明澈撰写,第4章、第10章和第11章由黄立波撰写,第8章由王鹏撰写。

本书的编写得到了机械工业出版社的大力支持,并得到了如下项目的资助:国家自然科学基金课题61572508、61272144、61672526、61303065、61472435、61572509和61402497,国家重点研发计划课题2016YFB0200203,国家863计划课题2015AA015302。

本书是国内第一部片上网络领域的学术专著,可用作研究人员的科研参考书,

也可作为计算机科学和微电子专业高年级本科生和研究生教材。

由于作者的能力和知识面有限，书中难免存在错误和缺陷，恳请读者批评指正。

王志英

国防科技大学

zywang@nudt.edu.cn

2016年11月于湖南长沙

目 录

Contents

前言

第一部分 序言

第 1 章 绪论	2
1.1 众核处理器时代	2
1.2 以通信为核心的跨层次优化	3
1.3 片上网络简介	5
1.3.1 拓扑结构	6
1.3.2 路由算法	7
1.3.3 流控机制	8
1.3.4 路由器微结构	10
1.3.5 性能评价指标	13
1.4 片上网络研究现状	14
1.4.1 拓扑结构的研究	14
1.4.2 单播通信路由的研究	15
1.4.3 聚合通信路由的研究	16
1.4.4 流控机制的研究	17
1.4.5 路由器微结构的研究	18
1.5 真实处理器的片上网络	19
1.5.1 MIT RAW 处理器	19
1.5.2 Tiler TILE64 处理器	21
1.5.3 Sony/Toshiba/IBM Cell 处理器	23

1.5.4 U. T. Austion TRIPS 处理器	24
1.5.5 Intel Teraflops 处理器	26
1.5.6 Intel SCC 处理器	27
1.5.7 Intel Larrabee 处理器	29
1.5.8 Intel Knights Corner 处理器	30
1.5.9 真实处理器片上网络特性总结	32
1.6 全书内容概述	34
1.7 参考文献	36

第二部分 逻辑层实现

第 2 章 单周期翼通道路由器结构

2.1 引言	48
2.2 翼通道路由器体系结构	50
2.2.1 翼通道单周期路由器总体结构	50
2.2.2 翼通道工作原理	55
2.3 路由器微体系结构设计	58
2.3.1 通道分配部件	58
2.3.2 快速仲裁部件	60
2.3.3 SIG 管理单元和 SIG 控制单元	61
2.4 实验评估	62
2.4.1 模拟环境	62
2.4.2 流水线延迟分析	63
2.4.3 延迟与吞吐率	64
2.4.4 面积与功耗	67
2.5 本章小结	68
2.6 参考文献	69

第 3 章 动态虚通道路由器

3.1 引言	71
3.2 拥塞感知的动态虚通道结构	72
3.2.1 动态虚通道	72
3.2.2 拥塞缓解策略	74

3.3	拥塞感知的多端口共享缓冲结构	75
3.3.1	多端口共享缓冲的动态虚通道	75
3.3.2	拥塞缓解策略	78
3.4	DVC 路由器微结构	78
3.4.1	虚通道控制部件	79
3.4.2	拥塞缓解电路	81
3.4.3	虚通道仲裁部件与开关仲裁部件	82
3.5	HiBB 路由器微结构	84
3.5.1	虚通道控制部件	85
3.5.2	虚通道仲裁部件与输出端口仲裁部件	86
3.5.3	虚通道调整结构	88
3.6	实验评估	89
3.6.1	DVC 路由器评估	89
3.6.2	HiBB 路由器评估	92
3.7	本章小结	95
3.8	参考文献	96
第 4 章	虚拟总线拓扑结构	98
4.1	引言	98
4.2	相关研究	99
4.3	研究动机	100
4.3.1	基本片上通信网络	100
4.3.2	片上网络问题分析	101
4.3.3	基于事务的总线通信优势	103
4.4	虚拟总线片上网络	103
4.4.1	拓扑结构	103
4.4.2	虚拟总线机制	105
4.4.3	饿死与死锁避免	111
4.4.4	VBON 路由器结构	111
4.5	实验评估	112
4.5.1	模拟框架	113
4.5.2	合成流量评估	115
4.5.3	真实应用评估	118

4.5.4 功耗分析	119
4.5.5 开销分析	119
4.6 本章小结	120
4.7 参考文献	120

第三部分 网络层路由和流控设计

第 5 章 区域隔离路由算法	124
5.1 引言	124
5.2 相关研究	126
5.3 研究动机	127
5.3.1 局部自适应算法的局限性	128
5.3.2 应用程序内部的干扰	129
5.3.3 应用程序之间的干扰	130
5.4 区域隔离路由算法	131
5.4.1 拥塞信息传播网络	132
5.4.2 DBAR 路由器微结构	134
5.4.3 路由函数设计	136
5.5 实验评估	139
5.5.1 路由函数评估	140
5.5.2 单区域性能	142
5.5.3 多区域性能	145
5.5.4 集中型网格网络性能	148
5.6 硬件开销讨论	151
5.6.1 连线资源	151
5.6.2 路由器开销	151
5.6.3 功耗和能量延迟积	151
5.7 进一步讨论	152
5.7.1 拥塞信息传播网络带宽	152
5.7.2 DBAR 的可扩展性	153
5.7.3 拥塞信息传播延迟	153
5.8 本章小结	153
5.9 参考文献	153

第 6 章 完全自适应路由算法的流控机制	158
6.1 引言	158
6.2 相关研究	161
6.2.1 死锁避免理论	161
6.2.2 完全自适应路由算法设计	162
6.3 研究动机	162
6.3.1 虚通道分配策略	162
6.3.2 路由灵活性	163
6.4 流控和路由设计	165
6.4.1 全报文发送	165
6.4.2 逃逸虚通道的积极分配策略	168
6.4.3 完全自适应路由算法	171
6.4.4 路由器微结构	171
6.5 合成流量模式评测	173
6.5.1 合成流量模式结果	174
6.5.2 路由算法的缓存利用率	176
6.5.3 敏感性分析	178
6.6 真实应用程序评测	181
6.6.1 实验方法和实验配置	182
6.6.2 PARSEC 测试集结果	182
6.7 流控机制的详细分析	183
6.7.1 缓存利用率的详细分析	183
6.7.2 流控机制的公平性分析	187
6.8 进一步讨论	189
6.8.1 报文长度和虚通道深度	189
6.8.2 DAMQ 和混合流控机制	190
6.9 本章小结	190
6.10 参考文献	190
第 7 章 切片气泡流控机制	195
7.1 引言	195
7.2 传统设计的局限	197
7.2.1 dateline	197

7.2.2	本地气泡策略	198
7.2.3	关键气泡策略	198
7.2.4	处理变长报文的低效性	199
7.3	切片气泡流控机制及策略	200
7.3.1	理论描述	200
7.3.2	本地切片气泡策略	202
7.3.3	关键切片气泡策略	202
7.3.4	饿死现象	203
7.4	路由器流水线和微结构	205
7.4.1	FBFC 路由器	205
7.4.2	VCT 路由器	206
7.5	实验方法	207
7.6	一维 Torus 网络性能评测	208
7.6.1	性能	208
7.6.2	缓存利用率	210
7.6.3	短报文和长报文的传输延迟	211
7.7	二维 Torus 网络性能评测	212
7.7.1	4×4 Torus 网络性能	212
7.7.2	单切片报文比例敏感性分析	214
7.7.3	缓存数量敏感性分析	215
7.7.4	8×8 Torus 网络可扩展性分析	216
7.7.5	饿死现象分析	217
7.7.6	PARSEC 测试集实验结果	219
7.7.7	大规模系统和消息传递编程模式	220
7.8	开销: 功耗和面积	221
7.8.1	测试方法学	222
7.8.2	功耗	222
7.8.3	面积	225
7.8.4	与网格网络的比较	226
7.9	进一步讨论和相关工作	229
7.9.1	进一步讨论	229
7.9.2	相关工作	229
7.10	本章小结	230

7.11 参考文献	231
-----------------	-----

第 8 章 高效能与公平性流控

8.1 基于当值缓存高效能路由器设计	235
8.1.1 引言	235
8.1.2 高效能路由器	236
8.1.3 实验设计	240
8.1.4 实验评估	241
8.1.5 小结	243
8.2 基于报文剩余跳数的自适应流控	244
8.2.1 引言	244
8.2.2 报文之间的相互作用	245
8.2.3 公平性流控	246
8.2.4 实验评估	249
8.2.5 小结	252
8.3 参考文献	253

第四部分 软件层编程模式支持

第 9 章 一致性协议聚合通信支持

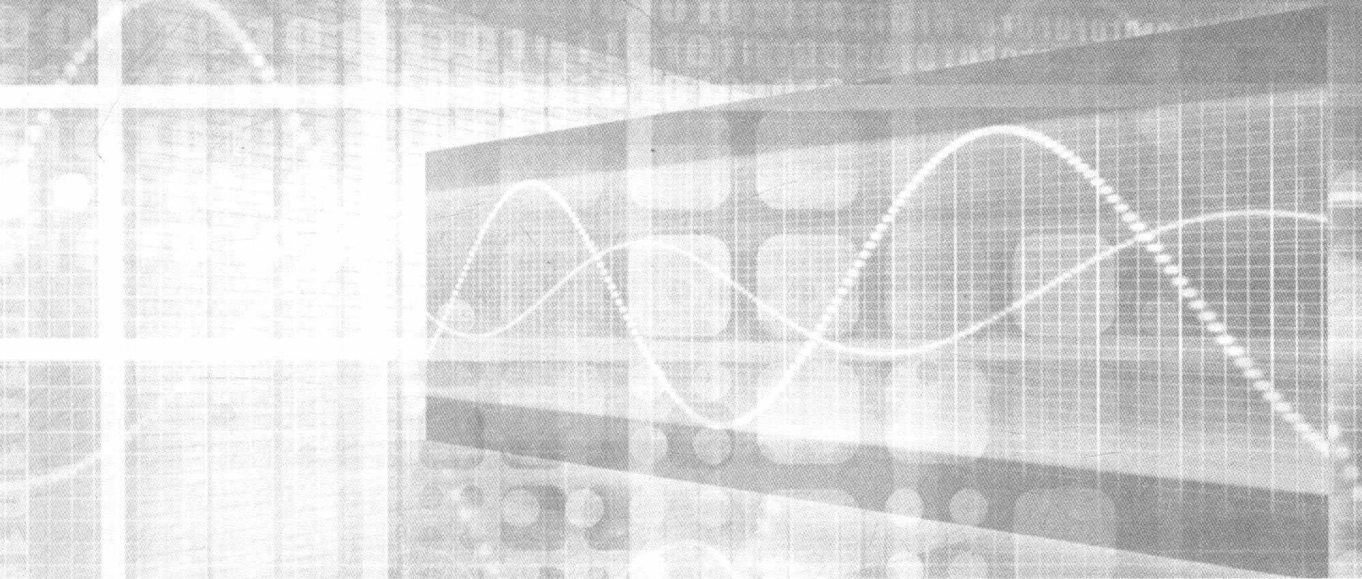
9.1 引言	256
9.2 归约消息组合框架	259
9.2.1 消息组合表格式	260
9.2.2 消息组合实例	260
9.2.3 消息组合表项的不足	262
9.3 均衡自适应多播路由算法	262
9.4 路由器流水线和微结构	264
9.5 实验评估	266
9.5.1 性能	267
9.5.2 BAM 和 RPM 多播虚拟网络性能	271
9.5.3 消息组合表大小	272
9.5.4 敏感性分析	274
9.6 功耗和能量延迟积分析	276

9.7	相关研究	277
9.7.1	消息组合	277
9.7.2	片上网络多播路由算法	278
9.8	本章小结	278
9.9	参考文献	279
第 10 章	MPI 原语的定制通信架构	283
10.1	引言	283
10.2	研究背景	284
10.3	研究动机	286
10.3.1	在 NoC 上实现 MPI	286
10.3.2	MPI 函数优化	286
10.4	定制通信架构	287
10.4.1	体系结构概述	287
10.4.2	定制的片上网络设计: VBON	288
10.4.3	MPI 原语实现: MU	288
10.5	实验评估	295
10.5.1	方法	295
10.5.2	实验结果	297
10.6	本章小结	303
10.7	参考文献	304
第 11 章	MPI 通信协议优化	308
11.1	引言	308
11.2	研究背景	309
11.2.1	MPI 的通信协议	309
11.2.2	现存问题	310
11.2.3	相关工作	313
11.3	研究动机	314
11.4	自适应通信协议	315
11.4.1	目标与方法	315
11.4.2	基本 MPI 加速的片上网络设计	316
11.4.3	ADCM 结构的支持	318

11.4.4	与理想协议的比较	323
11.5	实验评估	324
11.5.1	方法	324
11.5.2	合成流量结果	326
11.5.3	真实应用程序	328
11.5.4	敏感性分析	330
11.5.5	硬件开销	331
11.6	本章小结	331
11.7	参考文献	332

第五部分 后记

第 12 章	结语与展望	336
12.1	结语	336
12.2	展望	338



第一部分

Part I

序 言

■ 第1章 绪论

第 1 章

Chapter 1

绪 论

1.1 众核处理器时代

半导体工业在过去半个多世纪一直遵循“摩尔定律”发展，为维持产品的竞争力，集成电路单位面积的晶体管数目每隔约两年翻一番^[108]。目前，单块芯片上已经集成了数十亿只晶体管^[114, 128]，摩尔定律在未来十年内将继续有效^[131]。芯片上持续增长的晶体管数目一方面允许体系结构研究者采用许多新颖技术不断提升处理器的性能^[54]，另一方面，数量众多的晶体管也给处理器设计带来了许多问题，包括功耗增加^[42, 111]、资源利用率减少^[38]和可靠性降低^[14]等，这些问题使整个体系结构界面临着严峻的挑战。

直到 21 世纪初，体系结构界主要通过提高单核处理器的串行处理能力来提升性能^[110]。自从第一块处理器诞生起，其串行处理性价比已经提升了 100 亿次^[117]。一些能显著提升性能的技术包括采用超深流水线提高处理器主频^[56, 60]、使用先进的超标量技术挖掘指令级并行^[73, 83]和配置大容量片上高速缓存挖掘程序的时空局部性^[70, 121]。然而，由于处理器的动态功耗与主频成正比，先进的超标量技术需使用复杂的逻辑，大容量高速缓存带来了大量的存储阵列，因此这些技术都导致了处理器复杂度和功耗的急剧上升^[3, 134]。此外，已经被挖掘的指令级并行和时空局部性已接近串行程序的上限，因而提高处理器串行处理能力能获得的收益在逐步下降^[120, 143]。急剧上升的功耗和逐步下降的性能收益都成为单核处理器设计中不可逾越的障碍，这些障碍引发了处理器体系结构设计的深刻变革。

为有效利用单芯片上的数十亿只晶体管，计算机体系结构界开始设计多核处理器^[50]。与单核处理器通过提升主频或使用超标量技术提高串行性能不同，多核处理