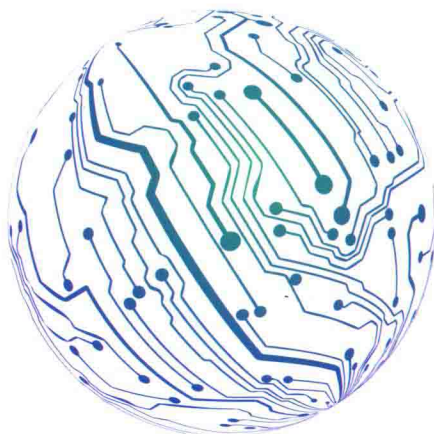




CMOS AND BEYOND

Logic Switches for
Terascale Integrated Circuits

CMOS及其他先导技术 特大规模集成电路设计

[美] 刘金 (Tsu-Jae King Liu) 科林·库恩 (Kelvin Kuhn) 等著 雷鑑铭 等译
加州大学伯克利分校 Intel 公司



机械工业出版社
China Machine Press



电子电气工程师技术丛书

CMOS及其他先导技术

特大规模集成电路设计

CMOS AND BEYOND

Logic Switches for
Terascale Integrated Circuits

[美] 刘金 (Tsu-Jae King Liu) 科林·库恩 (Kelvin Kuhn) 等著 雷鑑铭 等译
加州大学伯克利分校 Intel 公司



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

CMOS 及其他先导技术: 特大规模集成电路设计 / (美) 刘金 (Tsu-Jae King Liu) 等著; 雷鑑铭等译. —北京: 机械工业出版社, 2018.3

(电子电气工程师技术丛书)

书名原文: CMOS and Beyond: Logic Switches for Terascale Integrated Circuits

ISBN 978-7-111-59391-1

I. C… II. ①刘… ②雷… III. CMOS 电路—电路设计 IV. TN432.02

中国版本图书馆 CIP 数据核字 (2018) 第 048850 号

本书版权登记号: 图字 01-2016-7268

This is a Chinese simplified edition of the following title published by Cambridge University Press:
Tsu-Jae King Liu, Kelin Kuhn, CMOS and Beyond: Logic Switches for Terascale Integrated Circuits, 978-1-107-04318-3.

© Cambridge University Press 2015.

This Chinese simplified edition for the People's Republic of China (excluding Hong Kong, Macau and Taiwan) is published by arrangement with the Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

© Cambridge University Press and China Machine Press in 2018.

This Chinese simplified edition is authorized for sale in the People's Republic of China (excluding Hong Kong, Macau and Taiwan) only. Unauthorized export of this simplified Chinese is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of Cambridge University Press and China Machine Press.

本书原版由剑桥大学出版社出版。

本书简体字中文版由剑桥大学出版社与机械工业出版社合作出版。未经出版者预先书面许可, 不得以任何方式复制或抄袭本书的任何部分。

此版本仅限在中华人民共和国境内 (不包括香港、澳门特别行政区及台湾地区) 销售。

本书概述了现代 CMOS 晶体管的技术发展, 并提出了新的设计方法来改善晶体管性能的局限性。本书共四部分。第一部分回顾了芯片设计的注意事项并且基准化了许多替代性的开关器件, 重点论述了具有更大亚阈值摆幅的器件。第二部分涵盖了利用量子力学隧道效应作为开关原理来实现更陡峭亚阈值摆幅的各种器件设计。第三部分涵盖了利用替代方法实现更高效开关性能的器件。第四部分涵盖了利用磁效应或电子自旋携带信息的器件。本书适合作为电子信息类专业与工程类专业的教材, 也可作为相关专业人士的参考书。

CMOS 及其他先导技术 特大规模集成电路设计

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: 谢晓芳

责任校对: 李秋荣

印 刷: 北京市荣盛彩色印刷有限公司

版 次: 2018 年 4 月第 1 版第 1 次印刷

开 本: 186mm × 240mm 1/16

印 张: 21.25

书 号: ISBN 978-7-111-59391-1

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) 88379426 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzit@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

近半个世纪以来，作为数字集成电路中电子开关的主要器件，CMOS（互补金属氧化物半导体）晶体管持续小型化使电子器件的性价比不断提高。“超越摩尔定律”时代将继续推动集成电路行业的创新和发展。因利用光刻技术缩小半导体尺寸的成本日益高昂且难以实施，业界将目光转向于CMOS及其他先导技术。本书针对应用于特大规模集成电路的逻辑开关，重点关注替代CMOS器件的新材料、新器件、新工艺等。为了推动其他先导技术及其新材料、新器件、新工艺、新电路等的研究与发展，以及促使国内更多的研究开发人员与高等院校学生了解超越CMOS的先导技术，受机械工业出版社华章公司的委托，由华中科技大学在微电子集成电路领域长期从事一线科研及教学研究的教师组织并完成了本书翻译工作，将一本关于CMOS和其他先导技术的参考书奉献给读者。

本书采用系统的方法来讲解如何运用新型器件概念，以及用已有深入的、可使用的方法，来克服存在的设计挑战。本书借鉴了工业界和学术界中主要研究人员的专业知识，包括许多开发者的贡献，从一系列不同的观点引入和探讨新的概念，是一本极具前沿性且自成体系的、理论性及实践性较强的著作。本书涵盖了克服晶体管性能已有的局限性且有潜力的前沿技术，如隧道场效应晶体管（TFET）、替代性电荷电子器件、自旋电子器件以及更多先导技术与独特方法，适合微电子科学与工程、集成电路设计与集成系统、电子科学与技术及光电信息科学与工程等领域的学术研究人员和专业工程技术人员阅读，也可作为高等院校相关专业的教师、研究生及高年级本科生的教材和专业参考书。

本书由华中科技大学光学与电子信息学院及武汉国际微电子学院副院长雷鑑铭博士负责组织并完成全书翻译工作，参与本书翻译工作的还有赵于汐及高煜程等。本书在翻译过程中得到了华中科技大学光学与电子信息学院及武汉国际微电子学院邹雪城教授、邹志革副教授和余国义高级工程师的帮助及支持，在此表示感谢。特别感谢文华学院外国语学院英语系肖艳梅老师的审校。

CMOS及其他先导技术涉及的专业面广，鉴于译者水平有限，书中难免有不足及疏漏之处，敬请广大读者批评指正和谅解，在此表示衷心的感谢。

雷鑑铭

前言 Preface

过去四十年以来，作为数字芯片中电子开关的主要类型，CMOS（互补金属氧化物半导体）晶体管持续小型化使电子器件的性价比不断提高。器件的微型化已经造就了信息技术的无所不在，并且对于现代社会生活的方方面面产生了巨大的影响。

CMOS 技术已经趋于成熟，所以持续的晶体管尺寸等比例缩小在未来将不会像过去那样简单可行。这一点从某些方面发展速度的放慢（比如，芯片电源电压等比例减小，晶体管开态漏电流的等比例减小等）就可以看出。很明显，开关设计需要改进，以此来维持下一个十年之后电子行业的发展。很多种类的替代开关设计正在被研究人员讨论，其中许多开关设计用到了与传统 CMOS 晶体管完全不同的工作原理。但是，在这飞速发展的领域中研究人员发表的文章很少具有指导性。因此，很多重要的新信息不能被主流电子领域研究人员所理解。

为了解决以上问题，我们与该研究领域一些公认的专家共同创作了本书，包括：与性能与功耗的折中（激励陡峭的亚阈值摆幅器件）相关的背景信息、隧道效应器件、替代性场效应器件，以及电子自旋（磁性）器件。本书结尾部分论述了这些新型开关设计之间互连存在的挑战。

第一部分回顾了芯片设计的注意事项，并且基准化了许多替代性的开关器件，重点论述了具有更陡峭亚阈值摆幅的器件。第 1 章介绍了过去晶体管尺寸等比例缩小中的基本概念，并且分析了密度、功耗和性能这些推动现代 CMOS 设计要素之间的关键折中。在持续的晶体管尺寸等比例缩放限制的背景下，本书也回顾了诸如电源门控和并行设计等电路设计技术。同时结合具有更陡峭亚阈值摆幅的新型 CMOS 器件的潜在优势，论述了由 $60\text{mV}/10$ 倍频程亚阈值摆幅限制造成的 CMOS 技术中的能量效率限制。第 2 章和第 3 章介绍并基准化了相关研究领域中很多正在探索的替代性器件。这些章节主要关注电子器件（相对于磁性器件），它们包含了可以提高开关性能的新原理和新材料。第 2 章介绍了这些器件的历史和工作原理。第 3 章主要从驱动电流、能量效率、制造成本、复杂程度和存储单元面积等方面来评价这些器件。第 4 章探讨了在 CMOS 晶体管栅叠层中引入铁电层来克服 $60\text{mV}/10$ 倍频程亚阈值摆幅限制的方法。其中展示了理论和近期的实验，用于支持通过小信号负电容来实现 CMOS 晶体管的可能性。

第二部分涵盖了利用量子力学隧道效应作为开关原理来实现更陡峭亚阈值摆幅的各种器件设计。根据同时实现陡峭亚阈值摆幅、大开关电流比和高开态电导的要求,第5章评估了隧道场效应晶体管(TFET)的前景。其中研究了pn结维度的影响,论述了各种设计的折中,以及侧面、垂直及双层实现的优点。根据各种设计要求,对近期的实验数据进行了评价。第6章继续对TFET进行论述,重点关注了Ⅲ~Ⅳ族半导体材料。该章论述了设计同质结相对异质结Ⅲ-V族半导体材料的折中,如何通过p沟道TEFT来实现高性能,以及与Ⅲ~V族半导体材料特别相关的非理想性(比如陷阱、表面粗糙度和混合无序)。第7章通过评估用石墨烯和二维半导体材料制作的TEFT前景进一步探讨了TEFT。该章介绍了面内隧道效应器件和层间隧道效应器件,并结合理论上的理解对近期的实验结果进行论述。第8章介绍了一种新型隧道效应器件,即双层伪自旋场效应晶体管(BiSFET)。BiSFET依赖于实现室温下两个电介质分离的石墨烯层中激子(电子-空穴)超流体凝结的可能性。室温下凝结的形成是BiSFET工作原理的关键所在。该章论述了创造这样一个凝结现象的关键物理条件和挑战。BiSFET的精简模型和电路设计也将论述,同时体现其相对于CMOS的性能优势。

第三部分涵盖了利用替代方法实现更高效开关性能的器件。第9章讨论了使用相关电子材料制作器件的可能性,这种器件可以在绝缘体相和金属相之间转换。其中论述了这种金属-绝缘体转换的物理机制,并着重论述了二氧化钒(VO_2)系统。该章同时论述了Mott FET器件、固态 VO_2 FET器件和液态栅极 VO_2 FET器件,以及使用这些器件的电路结构。第10章介绍了压电晶体管(PET)器件。PET实质上是一个固态继电器,其中压电单元提供了机械力,压阻元件将机械力转化为电子开关。该章同时论述了压电和压阻材料的基本物理原理,以及工艺集成的挑战,也探讨了PET动力学、精简模型和电路设计,以及它们相对于CMOS的性能优势。第11章论述了作为逻辑开关的纳米级机电继电器。继电器用机械运动从物理上缩短或断开两个接触物之间的联系,它有零开态漏电流的理想特征、极大的亚阈值摆幅和低的栅漏。该章还介绍了纳米级继电器特殊的材料要求和工艺集成的挑战,描述了一系列用于更精简的复杂逻辑电路实现的继电器,并且论述了尺寸等比例缩小的方法。

第四部分涵盖了利用磁效应或电子自旋携带信息的器件。这些器件能用于实现纳米磁逻辑(其中小磁体用于构建电路)、电子自旋转矩逻辑和电子自旋波逻辑(其中电子自旋用于表征信息)。第12章论述了利用微小单域磁体制造电路的可能性。该章同时介绍了单域纳米磁体的开关特性和多种同步方案。该章提出了一个与CMOS不同的全加器结构,并回顾了纳米磁逻辑设计中的问题。第13章介绍了利用电子自旋转矩效应来制造大多数逻辑门电路的可能性。这些器件中,结合多种输入的自旋转矩作用将传输充足的转矩用于转换输出的磁化。该章同时回顾了面内和垂直自旋转矩转换的详细仿真。讨论了一全加器电路,并给出了相对于CMOS技术的基础测试结果。第14章分析了用电子自旋波实现逻辑功能的可能性。自旋波是在磁化方向的自旋点阵中自旋振动的集合。该章同时介绍了自旋波器件的物理机制,并讨论了相关实验结果。也

回顾了各种与 CMOS 不同的自旋波电路和体系结构，并给出了相对于 CMOS 技术的基准测试结果。

在讲述 CMOS 其他先导器件的时候，关键却又常常被忽略的问题是互连结构。如果不能与其他有源器件和无源器件互连起来，仅仅构建一个极好的新型开关就没有什么价值。这对于磁性和基于自旋的器件尤其重要，因为它们通常都不能与传统的电子器件直接相连。因此在全书的末尾，第 15 章深入讨论更高级逻辑器件互连的注意事项。其中包括新兴的电荷器件技术互连和自旋技术互连。

相比于主流半导体器件，我们希望这些章节有助于缩小特大规模（万亿级晶体管规模）集成电路中新兴器件研究与实际主流半导体领域研究之间的差距。

译者序

前言

第一部分 CMOS 电路和工艺限制**第 1 章 CMOS 数字电路的能效**

限制 2

1.1 概述 2

1.2 数字电路中的能量-
性能折中 3

1.3 能效设计技术 6

1.4 能量限制和总结 8

参考文献 9

第 2 章 先导工艺晶体管等比例

缩放: 特大规模领域

可替代器件结构 10

2.1 引言 10

2.2 可替代器件结构 10

2.3 总结 22

参考文献 23

第 3 章 基准化特大规模领域

可替代器件结构 30

3.1 引言 30

3.2 可替代器件等比例缩放
潜力 30

3.3 可比器件的缩放潜力 33

3.4 评价指标 35

3.5 基准测试结果 37

3.6 总结 38

参考文献 39

第 4 章 带负电容的扩展

CMOS 44

4.1 引言 44

4.2 直观展示 45

4.3 理论体系 47

4.4 实验工作 51

4.5 负电容晶体管 54

4.6 总结 56

致谢 57

参考文献 57

第二部分 隧道器件

第5章 设计低压高电流隧穿

晶体管	62
5.1 引言	62
5.2 隧穿势垒厚度调制	
陡峭度	63
5.3 能量滤波切换机制	65
5.4 测量电子输运带边陡度	66
5.5 空间非均匀性校正	68
5.6 pn 结维度	68
5.7 建立一个完整的隧穿场效应	
晶体管	80
5.8 栅极效率最大化	84
5.9 避免其他的设计问题	88
5.10 总结	88
致谢	89
参考文献	89

第6章 隧道晶体管

6.1 引言	92
6.2 隧道晶体管概述	93
6.3 材料与掺杂的折中	95
6.4 几何尺寸因素和栅极	
静电	99
6.5 非理想性	103
6.6 实验结果	106
6.7 总结	108
致谢	108
参考文献	108

第7章 石墨烯和二维晶体隧道

晶体管

7.1 什么是低功耗开关	115
7.2 二维晶体材料和器件的概述	116
7.3 碳纳米管和石墨烯纳米带	116
7.4 原子级薄体晶体管	124
7.5 层间隧穿晶体管	130
7.6 内部电荷与电压增益陡峭器件	137
7.7 总结	137
参考文献	137

第8章 双层伪自旋场效应晶体管 ...

8.1 引言	140
8.2 概述	141
8.3 基础物理	145
8.4 BiSFET 设计和集约模型	152
8.5 BiSFET 逻辑电路和仿真结果	157
8.6 工艺	161
8.7 总结	162
致谢	163
参考文献	163

第三部分 可替代场效应器件

第9章 关于相关氧化物中金属-绝缘体转变与相位突变的计算与学习

9.1 引言	166
9.2 二氧化钒中的金属-绝缘体转变 ...	168

9.3 相变场效应器件.....	172	12.2 作为二进制开关单元的单域 纳米磁体	242
9.4 相变两端器件.....	178	12.3 耦合纳米磁体特性	244
9.5 神经电路.....	181	12.4 工程耦合: 逻辑门与级联门 ...	246
9.6 总结.....	182	12.5 磁有序中的错误	248
参考文献	182	12.6 控制磁有序: 同步纳米磁体 ...	250
第10章 压电晶体管	187	12.7 NML 计算系统	252
10.1 概述	187	12.8 垂直磁介质中的纳米磁体 逻辑	255
10.2 工作方式	188	12.9 两个关于电路的案例研究	259
10.3 PET 材料的物理特性	190	12.10 NML 电路建模	260
10.4 PET 动力学	193	12.11 展望: NML 电路的未来	261
10.5 材料与器件制造	200	致谢	261
10.6 性能评价	203	参考文献	262
10.7 讨论	205		
致谢	206	第13章 自旋转矩多数逻辑门 逻辑	267
参考文献	206	13.1 引言	267
第11章 机械开关	209	13.2 面内磁化的 SMG	268
11.1 引言	209	13.3 仿真模型	270
11.2 继电器结构和操作	210	13.4 面内磁化开关的模式	272
11.3 继电器工艺技术	214	13.5 垂直磁化 SMG	276
11.4 数字逻辑用继电器设计优化 ...	220	13.6 垂直磁化开关模式	278
11.5 继电器组合逻辑电路	227	13.7 总结	283
11.6 继电器等比例缩放展望	232	参考文献	284
参考文献	234		
		第14章 自旋波相位逻辑	286
第四部分 自旋器件		14.1 引言	286
第12章 纳米磁逻辑: 从磁 有序到磁计算.....	240	14.2 自旋波的计算	287
12.1 引言与动机	240	14.3 实验验证的自旋波元件及 器件	287

14.4	相位逻辑器件	290
14.5	自旋波逻辑电路与结构	292
14.6	与 CMOS 的比较	297
14.7	总结	299
	参考文献	300

第五部分 关于互连的思考

第 15 章	互连	304
15.1	引言	304

15.2	互连问题	305
15.3	新兴的电荷器件技术的 互连选项	307
15.4	自旋电路中的互连思考	312
15.5	自旋弛豫机制	315
15.6	自旋注入与输运效率	318
15.7	电气互连与半导体自旋电子 互连的比较	320
15.8	总结与展望	324
	参考文献	324

CMOS 电路和工艺限制

- 第 1 章 CMOS 数字电路的能效限制
- 第 2 章 先导工艺晶体管等比例缩放：
特大规模领域可替代器件结构
- 第 3 章 基准化特大规模领域可替代器件结构
- 第 4 章 带负电容的扩展 CMOS

CMOS 数字电路的能效限制

Elad Alon

1.1 概述

在过去的数十年中，CMOS（互补金属氧化物半导体）尺寸的等比例缩小已经与功能、性能和能效的大幅提高联系在一起。特别是，尽管实际的历史趋势并非严格遵循一种器件类型的尺寸缩小比例，但是“登纳德（Dennard）缩放定律”^[1]相对很长一段时间都成立，期间尽管开关频率呈线性增长，晶体管密度的二次方（缩放因子）不断增长，同时单位逻辑门功率的二次方不断减小。所有这些都是靠晶体管光刻尺寸来线性地缩放工作电压而实现的。理论上，这会导致芯片单位面积上持续的能量消耗，从而使芯片设计和制造者相对更容易在固定的芯片面积上增加晶体管密度（也就增加了功率），这样就能在单个芯片上实现更多功能。

但是，就像登纳德预测的那样，因为一些与晶体管相关的本征参数，特别是热电压 kT/q ，不会随着光刻尺寸等比例缩放，这种类型的尺寸缩放在 21 世纪初就终止了。在那时，因为漏电流（和漏能量）本质上是可忽略不计的，因此晶体管的阈值电压已经被当作一个即使减小也没有显著影响的尺寸缩放参数。然而，因为漏电流随着阈值电压呈指数变化，这种类型的尺寸缩放最终必然会停止。

1.2 节会详细讨论，对于当今的设计（大约自从 90nm 工艺技术节点以来），必须根据给定的性能需求选择阈值电压和电源电压，从而平衡泄漏能量和动态能量。这其中的含义就是简单的尺寸缩放在三方面（密度、功耗、性能）都不能带来明显的好处。相反，即便是给出了一个在光刻方面更先进的工艺技术，在能耗和性能之间也必须做出直接的折中。1.2 节将会强调的是，为了实现最优的能量效率，在器件级，晶体管必须实现 $10^4 \sim 10^6$ 级别的开关电流比。1.3 节挑选了一些技术进行讲解，特别是电源门控和并行设计的技术。电路设计和制造者利用这些技术来实现尺寸缩放下 CMOS 技术的能效潜力。最后，1.4 节着重讲述了 CMOS 晶体管有精确定义的最小的每次运行能耗，因此最终并行设计无法成为

限制芯片功率的有效方法。

1.2 数字电路中的能量-性能折中

为了解释为什么必须平衡电源电压和阈值电压来实现高能效比的数字电路，必须先简要讨论典型数字芯片的构成。如图 1.1 所示，处理器（完整的数字芯片设计的典型代表）最大的功耗往往产生的控制/数据通路上。事实上，芯片整体的性能和功耗也取决于这些控制/数据通路。图 1.1 也说明，时钟频率（性能）取决于同步寄存器之间组合逻辑的延时。

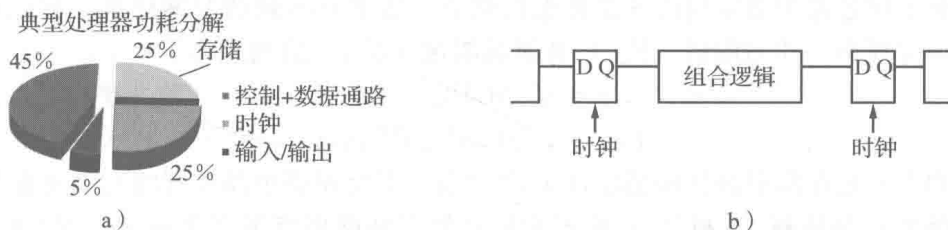


图 1.1 a) 一个典型嵌入式处理器的功耗分解；b) 同步数字电路的概念模型

虽然一个数字芯片中组合逻辑的实际构成明显有极多的种类，但是所有这种逻辑的表现（在功耗和性能方面）都与其级联反相器的表现十分接近。为了分析潜在的折中，可以用如图 1.2 所示的简化模型作为一般数字电路能量和性能的代表。如图 1.2 所示，紧密相关的电路级参数包括：活动因子 α （定义为在任意给定时钟周期下电路转换（也就是改变其状态）中给定节点的平均概率），容性扇出^① f ，每个反相器（门）的电容 C 和逻辑深度 L_d （触发器之间组合逻辑的级数）。

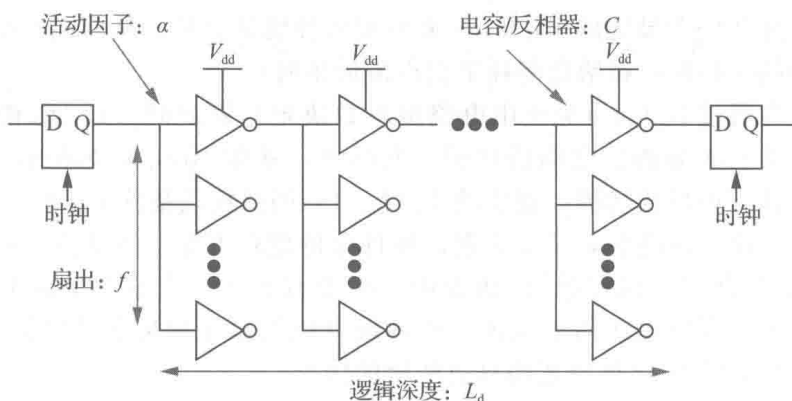


图 1.2 基于反相器的组合逻辑的能量-性能模型

有了这个模型，就很容易理解电路的延时 t_{delay} 由下式确定：

① 模型中扇出逻辑上出现在每一个反相器驱动 f 个相同反相器的情况下，但对于一般的数字电路而言，扇出应当视为容性。如一个给定的门的输入电容与链路中链接的门的输入电容的比。

$$t_{\text{delay}} = \frac{1}{2} \frac{L_d f C V_{\text{dd}}}{I_{\text{on}} (V_{\text{dd}} - V_{\text{th}})} \quad (1.1)$$

式中： V_{dd} 是电路的电源电压； $I_{\text{on}}(V_{\text{dd}} - V_{\text{th}})$ 是处于开态的反相器中晶体管的有效[⊖]漏电流，由给定的电源电压 V_{dd} 和给定的阈值电压 V_{th} 驱动。可以用各种不同的模型来展开 I_{on} 、 V_{dd} 和 V_{th} 之间的函数关系式（例如 α 功率定律^[2]、速度饱和效应^[3]等）。但是我们很快就会发现为了理解这些关键折中的潜在原因没有必要这样做。必须简单地意识到，导通电流会随着 $(V_{\text{dd}} - V_{\text{th}})$ 的增加而增加。

接下来考虑一下单次工作完成期间反相器链所消耗的能量。对于一个设计优秀的数字电路，能量消耗本质上只包括两方面：由电路中的寄生电容充电/放电产生的动态能量以及在整个工作过程中逻辑门的开关开态仍然会导通电流导致的泄漏能量。我们再次回到图 1.2 所示的模型，动态能量（ E_{dyn} ）和泄漏能量（ E_{leak} ）的构成如下：

$$E_{\text{dyn}} = \alpha L_d f C V_{\text{dd}}^2 \quad (1.2a)$$

$$E_{\text{leak}} = L_d f I_{\text{off}}(V_{\text{th}}) V_{\text{dd}} t_{\text{delay}} \quad (1.2b)$$

式中： $I_{\text{off}}(V_{\text{th}})$ 是在给定器件阈值电压 V_{th} [⊖]的条件下反相器中晶体管的有效关态泄漏。

为了强调必须选择 V_{dd} 和 V_{th} ，使得它们在给定性能要求下平衡输出这两个能耗分量，有必要将式（1.1）和式（1.2）相加，得到一个表示每次工作过程的总能量消耗的表达式：

$$\begin{aligned} E_{\text{total}} &= \alpha L_d f C V_{\text{dd}}^2 + L_d f I_{\text{off}}(V_{\text{th}}) V_{\text{dd}} \cdot \frac{1}{2} \frac{L_d f C V_{\text{dd}}}{I_{\text{on}}(V_{\text{dd}} - V_{\text{th}})} \\ &= \alpha L_d f C V_{\text{dd}}^2 \cdot \left[1 + \frac{L_d f}{2\alpha} \cdot \frac{I_{\text{off}}(V_{\text{th}})}{I_{\text{on}}(V_{\text{dd}} - V_{\text{th}})} \right] \end{aligned} \quad (1.3)$$

式（1.3）中最重要的一点是，虽然我们想用更低的 V_{dd} 来降低能耗，但是我们不得不同时降低 V_{th} 来维持同样的性能（也就是 $t_{\text{delay}} \propto C V_{\text{dd}} / I_{\text{on}}$ ），因此泄漏能量也会增加。这当中的重要含义是，存在一个最优的 V_{dd} 和 V_{th} 来平衡两种能量分量，从而能在给定的延时目标下达到最低总能耗（亦即：在给定能耗下获得最低延时）。

也要注意，当通过 $L_d f / \alpha$ （完全由电路级参数决定）缩放时， $I_{\text{on}} / I_{\text{off}}$ 值就能直接表示整个电路动态能量和泄漏能量之间的比例。实际上，就如 Nose 和 Sakurai 在关于超阈值 CMOS 电路的著作^[4]中所描述的，最优的 $I_{\text{on}} / I_{\text{off}}$ （并因此推及最优 V_{dd} 和 V_{th} 以及动态能量和泄漏能量之间的比）直接由 $L_d f / \alpha$ 决定，并且保持相对不变，忽略给定的延时目标。此外，由 Kam 和他的合著者在文献^[5]所做的一项分析表明，该结果实质上适用于任何类 CMOS 的器件技术，同时也适用于任何工作区域（也就是亚阈值和超阈值），甚至是那些漏电流和栅电压比 CMOS 晶体管要明显更陡峭的情况。

基于以上观察和分析，要给最优的 $I_{\text{on}} / I_{\text{off}}$ 提供一个数字化的指导标准，就有必要讨论电路级参数 L_d 、 f 和 α 代表性的值，同时给出选择这些值的理由。从逻辑深度 L_d 开始，这个量通常大约设定为 15~40。很像最优的 V_{dd} 和 V_{th} 折中，随着由冗余的时序元件（也就是

⊖ 在输出变换过程中，器件漏电流实际上不是恒定的，但在大多数情况下，可以很好地近似于一个值。

⊖ 电源电压 V_{dd} 也影响漏电流 I_{off} ，但就本次讨论的目的而言，这种效果不会改变潜在的权衡/结论。

触发器/寄存器)带来的更大开销,通过用更深入的流水设计(也就是减小 L_d)改善时间裕度,来实现平衡^[6]。类似地,扇出 f 通常设定成大于2,从而减小与每个门相关的延时开销。扇出 f 约上调到8是为了保证稳定地工作(具有更大扇出的逻辑门对于噪声和串扰有更强的抗干扰能力)。最后,大部分实际设计中总体上的活动因子 α 是从约10%下降到0.1%;这些相对低的百分比基于如下事实:在大部分复杂逻辑链(在更多存储结构)中,大部分逻辑门的状态在任意时钟周期下都不会改变。

综合考虑并且采用合适的比例系数,广泛适用于各种设计的最优 I_{on}/I_{off} 范围是 $10^4 \sim 10^6$ 。因为对于合理的性能级别,CMOS晶体管实现 $-100\text{mV}/10$ 倍频程的有效反转斜率(也就是 $V_{dd}/\lg(I_{on}/I_{off})$,如文献[5]所定义),实现这种开关电流比例所需的电源电压通常是 $500 \sim 600\text{mV}$ 。必须注意,对电路工作高性能要求越深入,有效总体斜率就越小,因此很多设计工作在 1V 左右的电压,从而实现想要的(峰值)性能。

在进入下一节之前,有必要基于历史和未来CMOS尺寸确认一下上述分析的含义。在传统的(登纳德)尺寸范围,同时降低 V_{dd} 和 V_{th} 会导致从上一代制造工艺到下一代的 I_{on}/I_{off} 比显著减小。然而,结果表明,用这种方式减小 I_{on}/I_{off} 比确实非常令人满意,因为在那个节点阈值已经设定得太高,以至于泄漏能量已经可以忽略。因此减小电源电压并节约动态能量是十分有益的。换句话说,尺寸等比例缩放能够通过这种方式向前发展的原因是,在那个时刻,典型的设计实际上并非工作在能量与延时折中空间的最优点。

为了使这个观点更清晰,图1.3用标记来说明给定制造工艺并且工作在标称电源电压和阈值电压下的设计会在哪里与最优能耗与延时曲线相关。如图1.3a所示,典型的设计大体上工作在最优曲线的右上方,但是随着 V_{dd} 和 V_{th} 的减小,尺寸等比例缩放使这些设计更加接近实际的最优曲线。换句话说,尺寸缩放的能效优势中有相当一部分实际上并不是由于尺寸缩放本身造成的,而是减小次优化的程度带来的结果。

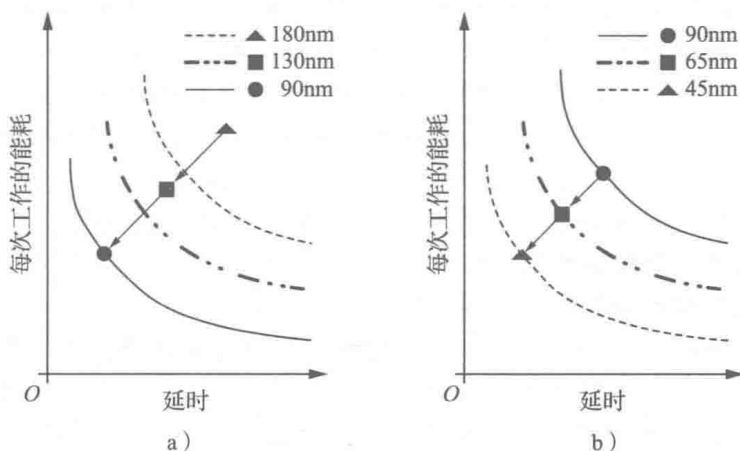


图 1.3 每次工作的能耗与使用标准电源电压和阈值电压的延时空间的设计缩放
a) 传统(登纳德)缩放; b) 现代(约低于90nm)缩放

这当然不是说尺寸缩放对于能耗和延时完全没有益处。简单地说,一旦电路设计基本上工作在曲线的最优部分,如图1.3b所示,单纯的尺寸缩放(保持 V_{dd} 和 V_{th} 固定)所

带来的最好也不过是工作能耗和延时的线性减小，这二者都是由于减小的电容/逻辑门^[7]而引起的。事实上，互连寄生的较弱尺寸缩放和变化的问题将使得电容/逻辑门尺寸缩放相对较差（也就是每逻辑门的最小总电容从一次工作过程到下一次工作过程不会显著减小）。

然而，即便是最好的情况下，简单的尺寸缩放未能提供足够的优势去使得缩放的设计在限定功耗下达到提高性能和功能的目的。尤其是如果电源电压和阈值电压固定，每逻辑门的功率（与 $E_{\text{total}}/t_{\text{delay}}$ 成比例）也就固定了。不过，如果真的能够利用更高的密度在每一次工艺生产过程中集成两倍的逻辑门，芯片的功耗也会翻倍。在绝大部分的应用中，芯片功耗从上一代工艺到下一代工艺必须保持不变（因为热或者电池寿命的限制），所以设计者被迫利用其他方法将尺寸缩放转向可用的先进技术。这些方法中最重要的——也就是并行设计——将会在下一节进一步论述。

1.3 能效设计技术

既然前一节讨论的关于功耗和性能之间的折中都归因于 CMOS 晶体管在应该关闭时会发生泄漏的事实，就自然想到能否用一个电路或系统级的技术来消除或者至少减少这种泄漏。最合适的备选方法是“功耗门控”或者“休眠晶体管”^[8]。图 1.4 描述了把这个概念应用于一个反相器链，其中，核心思路是在已知模块没有工作时将整个模块与电源分离。功率开关本身当然必须由某些种类的晶体管（通常，各种工艺中无论什么开关均可用）来实现，但是如果开关是用具有更高 $I_{\text{on}}/I_{\text{off}}$ 的器件（也就是一个器件具有更高的 V_{th} 或者更大的栅极电压摆幅）来实现的，闭合这个开关就能真正减少在闭合状态时整体电路与原始电路的泄漏。

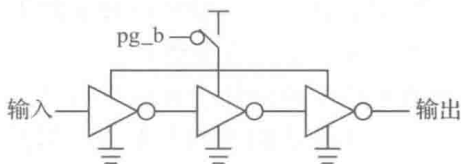


图 1.4 应用于反相器链的功率门控

顺着原始的思路往下走，我们可能会想能否更进一步利用功耗门控这个思想，使每个逻辑门完成有用的工作后立即断开，从而打破或者至少改善之前描述的折中。特别如果逻辑门只在它的输出需要转换时才“唤醒”，活动因子 α 会远大于之前引用的值。然而，这个想法的问题在于，何时断开或者闭合功耗门控开关。在每个独立逻辑门的功耗门控限制下，我们必须复制整个门的功能以计算这个功率门控信号。但是这个复制的逻辑门会遭遇与先前描述几乎完全相同的能耗-性能折中问题。

很明显，试图通过功率门控每个单独的逻辑门不能带来任何益处。然而，即使是用更适中的方法（也就是功率门控独立的子模块），也需要始终记住的关键问题是，不仅仅是功率门控本身会带来能耗-性能开销（由于功率门控器件活跃时的电压降，也由于驱动功率门控器件的寄生电容所消耗的能量），用于计算功率门控是否应该活跃的电路本身也会带来静态的和动态的能耗。因此，功率门控通常只应用于相对较粗粒度的情况，这里非常容易理解（或者由操作系统给出）是否潜在模块正在有效工作。

尽管功率门控不能改善先前描述的基本能耗-性能开销，但对于处理大部分应用中要求的计算具有间断性的情形很有效。比如，当一部手机处于待机模式时，应用程序处理器通