

信息科学技术学术著作丛书

低功耗集成电路

张 锋 沈海华 陈铨颖 著



科学出版社

低功耗集成电路

张 锋 沈海华 陈铨颖 著

科学出版社

北 京

内 容 简 介

本书从多个角度对超大规模集成电路 VLSI 的低功耗设计方法进行介绍。首先,从 SoC 芯片的角度出发介绍大规模集成电路的低功耗来源、发展趋势及功耗的评估和验证,这部分对于理论和内容都是从数字电路系统级角度出发,针对当前大规模 SoC 芯片的最新技术和成果进行论述。其次,从微电子的固态电路设计角度出发,较为详细地介绍亚阈值晶体管、低功耗低噪声放大器、低功耗 Sigma-Delta 模数转换器等晶体管电路低功耗设计原理与趋势。然后,主要介绍半导体领域的一个重要分支——存储器的高性能低功耗设计,并重点论述静态随机存储器和阻变存储器两个特征明显的存储器。最后对芯片未来的低功耗技术发展趋势进行阐述。

本书对 CMOS 集成电路设计和研究的读者,特别是高等院校中从事计算机芯片系统、电子信息领域、微电子专业学生,以及进行集成电路设计的工程师,都会起到有益的帮助。

图书在版编目(CIP)数据

低功耗集成电路/张锋,沈海华,陈铨颖著. —北京:科学出版社,2016
(信息科学技术学术著作丛书)

ISBN 978-7-03-050042-7

I. 低… II. ①张…②沈…③陈… III. 低功耗集成电路 IV. TN4

中国版本图书馆 CIP 数据核字(2016)第 232852 号

责任编辑:魏英杰 / 责任校对:郭瑞芝

责任印制:张 倩 / 封面设计:陈 敬

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京教图印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 9 月第 一 版 开本:720×1000 B5

2016 年 9 月第一次印刷 印张:15 1/4

字数:306 000

定价:90.00 元

(如有印装质量问题,我社负责调换)

《信息科学技术学术著作丛书》序

21 世纪是信息科学技术发生深刻变革的时代,一场以网络科学、高性能计算和仿真、智能科学、计算思维为特征的信息科学革命正在兴起。信息科学技术正在逐步融入各个应用领域,并与生物、纳米、认知等交织在一起,悄然改变着我们的生活方式。信息科学技术已经成为人类社会进步过程中发展最快、交叉渗透性最强、应用面最广的关键技术。

如何进一步推动我国信息科学技术的研究与发展;如何将信息技术发展的新理论、新方法与研究成果转化为社会发展的新动力;如何抓住信息技术深刻发展变革的机遇,提升我国自主创新和可持续发展的能力? 这些问题的解答都离不开我国科技工作者和工程技术人员的求索和艰辛付出。为这些科技工作者和工程技术人员提供一个良好的出版环境和平台,将这些科技成就迅速转化为智力成果,将对我国信息科学技术的发展起到重要的推动作用。

《信息科学技术学术著作丛书》是科学出版社在广泛征求专家意见的基础上,经过长期考察、反复论证之后组织出版的。这套丛书旨在传播网络科学和未来网络技术,微电子、光电子和量子信息技术、超级计算机、软件和信息存储技术,数据知识化和基于知识处理的未来信息服务业,低成本信息化和用信息技术提升传统产业,智能与认知科学、生物信息学、社会信息学等前沿交叉科学,信息科学基础理论,信息安全等几个未来信息科学技术重点发展领域的优秀科研成果。丛书力争起点高、内容新、导向性强,具有一定的原创性;体现出科学出版社“高层次、高质量、高水平”的特色和“严肃、严密、严格”的优良作风。

希望这套丛书的出版,能为我国信息科学技术的发展、创新和突破带来一些启迪和帮助。同时,欢迎广大读者提出好的建议,以促进和完善丛书的出版工作。

中国工程院院士

原中国科学院计算技术研究所所长



前 言

信息时代发展的一个重要标志是集成电路晶体管的发明,信息技术的革命性发展是建立在集成电路技术的不断革新的基础之上的。如今,几乎任何一个电子产品中都会有集成电路(integrated circuit, IC)产品,它每时每刻地在改变着人们的生活习惯,使人们通信更加便捷,生活更加方便,它的重要性已经不言而喻。

中国的集成电路产业发展到今天,已不再是世界集成电路的配角。中国的集成电路企业正在由过去从事简单、低端的集成电路产品向利润更高、高端的集成电路产品过渡,而科研界也慢慢的向世界一流的学术领域过渡,中国的学者在集成电路领域发表的文章已经逐渐跻身世界前列。因此,对于集成电路领域的一些前沿课题非常值得深入的学习和探讨。本书从集成电路低功耗的角度出发,对集成电路芯片设计中数字 SoC 低功耗芯片、模拟集成电路低功耗设计、存储器的低功耗设计等关键部分展开深入的分析和介绍,还在其中穿插了一定的工程实例,以供对集成电路低功耗感兴趣的读者学习之用。

本书主要包括 9 章。

第 1 章~第 3 章分别介绍集成电路低功耗的来源、低功耗技术的设计和实现、低功耗技术的评估和验证。

第 1 章从三个部分介绍集成电路低功耗来源,分别是动态切换功耗、瞬时短路功耗以及静态功耗。第 2 章从集成电路设计方法出发,自顶向下的从系统级到算法级、体系结构级、RTL 级、门级、电路级直到最终工艺级详细介绍低功耗的设计方法。第 3 章从多个角度介绍功耗评估的方法。

第 4 章~第 6 章,从微电子的固态电路设计角度出发,介绍基于晶体管电路设计的低功耗技术。

第 4 章介绍弱反型区 MOSFET 器件模型知识,重点讨论弱反型区中 MOS 晶体管的 PVT(电源、电压、温度)变量、失配效应,以及器件噪声等设计问题。第 5 章主要介绍放大器电路噪声的来源以及电路解决方法,同时介绍个别工程案例。第 6 章介绍 Sigma-Delta 模数转换器的基础知识,包括 Sigma-Delta 模数转换器的基本结构、调制器电路,以及性能参数等,并着重讨论低电源电压环境下 Sigma-Delta 调制器的设计思路和挑战。

第 7 章和第 8 章,介绍挥发性存储器的代表静态随机存储器(SRAM)和非挥发存储器的代表阻变存储器(RRAM)。这是两种典型存储器的高性能低功耗技术。

第7章对存储器的发展历史进行介绍,并详述SRAM的基本结构、特征电路,以及性能参数等。着重讨论SRAM的低功耗实现技术及仿真技术。第8章介绍阻变存储器的由来,并对其结构类型、材料组成、基本指标及低功耗实现技术提出了若干解决方案。

第9章,重点介绍集成电路低功耗技术的发展趋势及未来的发展方向。

本书理论性分析内容丰富,主要由张锋、沈海华、陈铖颖一同完成。其中张锋主要完成了第7、8、9章的编写,沈海华主要完成了第1、2、3章的编写,陈铖颖完成了第3、4、5章的编写。感谢项中元、马文龙等同事、同学在查找资料和文档整理方面付出的辛勤劳动,正是有了大家的共同努力,本书才得以顺利完成。

感谢863课题“RRAM外围电路关键技术研究”(2011AA010403)对本书的大力支持。

由于本书涉及知识面较广,限于水平,书中难免存在不足和局限,恳请读者批评指正。

作者

2016年3月

目 录

《信息科学技术学术著作丛书》序

前言

第 1 章 集成电路功耗来源	1
1.1 动态切换功耗	1
1.2 瞬时短路功耗	2
1.3 静态功耗	4
第 2 章 低功耗的设计与实现	6
2.1 系统级实现	6
2.1.1 动态电压/频率调节技术介绍	6
2.1.2 分块耗能控制的自动 DVFS 在能量受限的 NoC 通信模块上的应用	16
2.1.3 在线学习进行系统级能量控制	18
2.1.4 带有 DVFS 的多分区的内存结构	20
2.1.5 多时钟域处理器中的集成 CPU 高速缓存功耗管理	22
2.2 算法级实现(通过较少总线上比特翻转的次数减少功耗).....	26
2.2.1 编译码算法	27
2.2.2 Markov 模型	28
2.2.3 减少比特翻转次数的算法	29
2.3 结构级低功耗设计方法.....	30
2.3.1 总线的低功耗设计.....	30
2.3.2 存储器优化	31
2.3.3 预运算技术	31
2.3.4 并行技术	31
2.3.5 流水线技术	32
2.4 寄存器传输级(RTL)和门级(Gate-level)低功耗设计	37
2.4.1 时钟门控	37
2.4.2 动态频率调整(DFS)技术	39
2.4.3 电源门控技术	40
2.4.4 信号门控	41
2.5 电路级.....	41

2.5.1	电荷循环总线结构	41
2.5.2	多米诺逻辑	47
2.6	工艺级	50
2.6.1	多阈值电压	50
2.6.2	多电压技术	54
2.6.3	Gate Sizing	59
2.6.4	面积优化技术	63
第3章	功耗评估	67
3.1	基于模拟方法的 Fractal 算法低功耗估计	67
3.2	混合级别功率估计	68
3.3	存储器的功率估计	70
第4章	亚阈值 MOS 晶体管	72
4.1	MOS 工艺概述	72
4.2	MOS 器件模型	73
4.2.1	MOS 管 I/V 特性	73
4.2.2	二阶效应	76
4.3	亚阈区设计考虑	78
4.3.1	PVT 变量	78
4.3.2	匹配性	79
4.3.3	噪声	81
4.4	极低功耗亚阈值 MOS 晶体管电路设计	82
4.4.1	MOS 晶体管泄漏机理	83
4.4.2	MOS 晶体管泄漏降低技术	87
4.5	亚阈值 CMOS 逻辑中的参数变化影响	89
4.5.1	噪声裕度	90
4.5.2	能耗	92
4.6	小结	94
第5章	低功耗、低噪声放大器	95
5.1	芯片中的噪声	95
5.2	低频噪声及失调电压消除技术	98
5.3	斩波调制放大器设计	104
5.3.1	传统斩波调制放大器设计	104
5.3.2	低阻结点斩波调制放大器设计	107
5.4	亚阈值心电放大器设计	110
5.5	小结	115

第 6 章 低功耗 Sigma-Delta 模数转换器	116
6.1 Sigma-Delta 模数转换器基础	116
6.2 Sigma-Delta 模数转换器结构	119
6.2.1 单环调制器结构	119
6.2.2 多级噪声整形调制器结构	122
6.2.3 多位量化调制器结构	123
6.3 Sigma-Delta 调制器的性能参数	125
6.4 低功耗 Sigma-Delta 调制器电路设计	126
6.4.1 前馈 Sigma-Delta 调制器结构	127
6.4.2 采样开关运算放大器	129
6.4.3 低功耗运算放大器	132
6.4.4 低功耗比较器	136
6.5 小结	138
第 7 章 低功耗高速静态随机存储器	139
7.1 存储器说明	139
7.2 SRAM 的设计基础	141
7.2.1 基于 CMOS 工艺 SRAM 的电路结构	141
7.2.2 SRAM 的性能指标	144
7.3 SRAM 的高速低功耗设计技术	150
7.3.1 SRAM 的译码电路功耗	151
7.3.2 数据通路的高速低功耗设计技术	164
7.3.3 SRAM 的低功耗结构优化技术	174
7.4 小结	186
第 8 章 低功耗阻变存储器	187
8.1 阻变存储器说明	187
8.2 阻变存储器的低功耗操作	190
8.2.1 RRAM 的低功耗高可靠写入操作	194
8.2.2 RRAM 的低功耗高可靠读出操作	198
8.3 阻变存储器的热效应	203
8.4 小结	210
第 9 章 低功耗集成电路发展趋势分析	211
9.1 低功耗 SoC 的技术发展趋势分析	211
9.2 低功耗混合信号集成电路发展方向分析	211
9.3 低功耗存储器电路的未来发展方向	214
参考文献	216

第 1 章 集成电路功耗来源

进行低功耗设计的前提是应该详细地了解电路的功耗来源,以便量化分析影响集成电路功耗的各个因素,进而指导低功耗设计并有针对性地采取相关措施。

CMOS 数字集成电路的功耗由三部分组成,即

$$P = P_{\text{switch}} + P_{\text{leak}} + P_{\text{short}} \\ = \alpha \cdot C_L \cdot V_{\text{DD}}^2 \cdot f_{\text{clk}} + I_{\text{leak}} \cdot V_{\text{DD}} + I_{\text{sc}} \cdot V_{\text{DD}} \quad (1-1)$$

其中, P 是总体功耗; P_{switch} 是电路翻转时产生的动态切换功耗; P_{leak} 是由主要由泄漏电流引起的静态功耗(又称泄漏电流功耗); P_{short} 是 P 管和 N 管同时导通引起的瞬时短路功耗。

P_{switch} 和 P_{short} 又可统一归为动态功耗。对于不同的工艺,每一部分的功耗所占的比重一般是不同的。下面分别对这些功耗的原理进行简单介绍。

1.1 动态切换功耗

动态切换功耗来源于对负载电容的充放电。在 90nm 及以上工艺中,动态切换功耗占据主要地位,例如在 0.18 μm 中占 80%~90%。下面我们用一个反相器翻转过程来简单分析 CMOS 电路切换功耗的产生机理。

考虑该反相器的一个完整的充放电过程,如图 1-1 所示。图中 C_{load} 表示电路的负载电容。

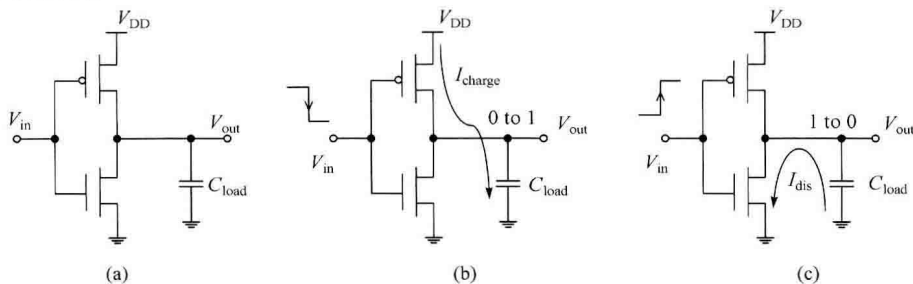


图 1-1 反相器动态切换过程

首先,假设输入信号 V_{in} 波形具有为零的上升时间和下降时间,或者说 PMOS 和 NMOS 器件决不会同时导通。对于如图 1-1(a) 所示的反相器,当输入信号 V_{in} 从高电平变为低电平时, N 管截止而 P 管导通,产生一个充电电流 I_{charge} 对负载电容 C_{load} 充电(图 1-1(b)),从 0 电平充至电源电压值 V_{DD} ,所需要的能量为 $E = C_{\text{load}} \cdot V_{\text{DD}}^2$ 。

但是,存储在 C_{load} 上的能量仅为 $C_{\text{load}} \cdot V_{\text{DD}}^2/2$,这是由于 MOS 管电阻的存在,在电流 I_{charge} 对负载电容 C_{load} 充电时,电源提供的另一半耗散能量在上拉电阻(PMOS 管的电阻)。注意到,这一能耗与 PMOS 管的尺寸无关(因此也与电阻无关)。

当输入信号 V_{in} 从低电平变到高电平时,P 管截止而 N 管导通,产生一个泄放电流 I_{dis} 对负载电容 C_{load} 放电,如图 1-1(c) 所示。前面充电时,存储在负载电容上的能量由于 I_{dis} 流过下拉电阻(NMOS 管的电阻)而被耗散掉,理想情况下此时供电电源没有提供新的能量,并且同样这一部分能耗与 NMOS 器件的尺寸无关。

可见,CMOS 电路完成一次完整的负载电容充放电所消耗的能量为一个固定值 $E=C_{\text{load}} \cdot V_{\text{DD}}^2$ 。若系统充放电的频率为 f ,则功耗为

$$P_{\text{switch}} = E \cdot f = C_{\text{load}} \cdot V_{\text{DD}}^2 \cdot f \quad (1-2)$$

考虑到实际电路中的结点并不是每个时钟周期都是活动的,而是以一定的概率翻转,设某结点的翻转概率(或者称为开关活动性因子)为 α ,则动态功耗为

$$P_s = \alpha \cdot C_{\text{load}} \cdot V_{\text{DD}}^2 \cdot f \quad (1-3)$$

对于一个有 N 个结点的电路系统,其总体动态功耗为

$$P_{\text{switch}} = \sum_{i=1}^N P_{s_i} = V_{\text{DD}}^2 \cdot f \cdot \sum_{i=1}^N \alpha_i \cdot C_{\text{load}_i} \quad (1-4)$$

其中, V_{DD} 为电源电压; f 为系统时钟频率; α 为结点的翻转概率; C_{load} 为负载电容。

式(1-4)为 CMOS 数字集成电路中最重要的功耗计算公式,由该式可知,动态功耗与晶体管的电阻、充放电电流,以及充放电时间等参量无关,仅取决于电源电压、电路工作频率、电路负载电容,以及结点翻转概率。要降低 CMOS 电路系统的动态功耗,必须从这几方面进行考虑,采取措施降低电压、工作频率、负载电容和翻转概率等。

1.2 瞬时短路功耗

瞬时短路功耗产生原因是 CMOS 电路输入信号的非理想性带来的。前面我们假设输入信号波形的上升时间和下降时间为零,这仅是理想的情况。输入信号不为无穷大的斜率造成 CMOS 门电路电平发生翻转时,即电平从“0”跳变到“1”或者从“1”跳变到“0”的过程中,会有一个很短的时间间隔。在该间隔内,NMOS 下拉网络和 PMOS 上拉网络同时导通出现一条直流通路,产生一个窄的从 V_{DD} 到 V_{ss} (或者地)的电流脉冲。下面以图 1-2 所示的反相器来简单说明。

如图 1-2(a) 所示为一个输入信号上升和下降的过程。在图 1-2(b) 中,反相器工作在电源电压 V_{DD} 下,N 管和 P 管的阈值电压分别为 V_{tn} 和 V_{tp} ,如传输曲线所示。当输入电平高于 V_{tn} 时,N 管打开;当输入电平低于 V_{DD} 一个 V_{tp} 时,P 管打开;当输入电平 V_{in} 跳变而处于 V_{tn} 和 $(V_{\text{DD}} - V_{\text{tp}})$ 之间时,P 管和 N 管均导通,此时 V_{DD}

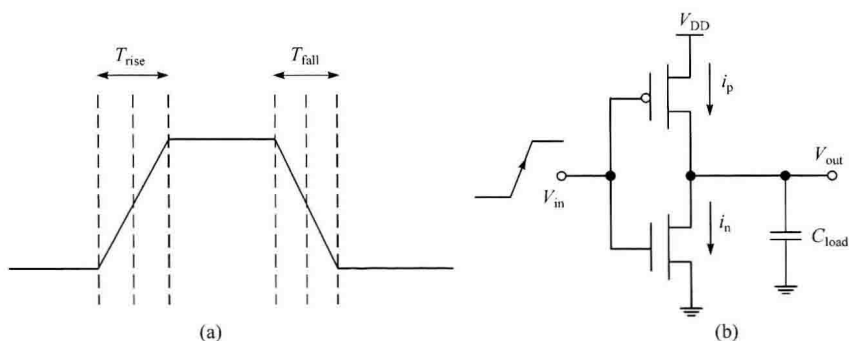


图 1-2 瞬时短路功耗的产生

和地之间有一个短路电流,能量以热能形式在 P 管和 N 管上耗散。

瞬时短路电流与输入电压 V_{in} 的关系如图 1-3 所示。在恒定的电源电压下,将 V_{DD} 与短路电流相乘就可以得到瞬时短路功耗,即 $P_{leak} = V_{DD} \cdot I_{short}$ 。短路功耗大小取决于 V_{DD} 、转换时间、门的设计,以及负载电容等。对于转换时间很短的电路,短路电流功耗所占比例比较小,对于转换比较慢的电路,则可占到总功耗的 30% 甚至更多;平均情况则占到 10% 左右。

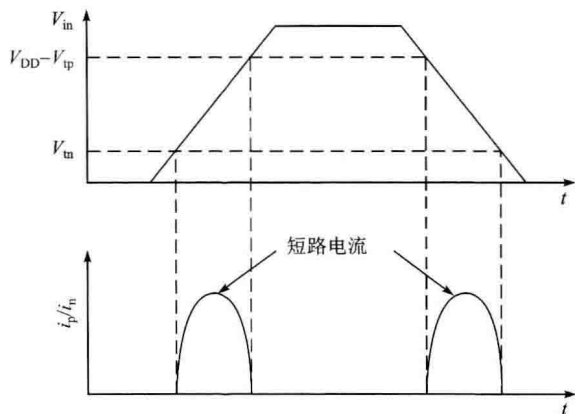


图 1-3 CMOS 反相器瞬时短路电流与输入电压的关系

一般而言,瞬态短路功耗与动态翻转功耗相比都很小,对其考虑并不是很多,设计时只要适当注意即可。

① 当输入端的 rise/fall 时间远大于输出端时,短路电流很大。

② 为减小总的平均短路电流值,应使输入输出的跳变时间相等,此时短路电流小于总的动态电流的 10%。

③ 如果供电电压小于晶体管阈值电压之和,即 $V_{DD} < V_{tn} + V_{tp}$, 短路电流将不存在,因为此时对任何输入,两个器件(NMOS 和 PMOS)不会同时导通。

1.3 静态功耗

静态功耗又称泄漏电流功耗,主要分布形式是漏电损失。在理论上,CMOS 电路不会有静态功耗,因为电路处于稳态时从电源到地没有直接的路径,但实际上由于晶体管并不理想,总有漏电流的存在。

在当前工艺的 CMOS 电路中,最主要的两种漏电流是亚阈值电流 I_{sub} 和反偏 PN 结漏电流 I_{junction} ,如表 1-1 所示。

表 1-1 纳米工艺下的主要漏电流机制

工艺/nm	年份	I_{sub}	I_{junction}
90	2004	840pA	25pA
50	2010	21nA	3.0nA
25	2016	260nA	120nA

反偏 PN 结漏电流为流过位于晶体管源(或者漏)与衬底之间的反向偏置二极管的漏电流。根据半导体物理的分析,该电流为

$$I_{\text{junction}} = I_s (e^{\frac{V}{V_{\text{th}}}} - 1) \quad (1-5)$$

其中, I_s 为 PN 反偏饱和电流,取决于制造工艺和 PN 结面积; V_{th} 为热电压,即

$$V_{\text{th}} = \frac{kT}{q} \quad (1-6)$$

式中, $k = 1.38 \times 10^{-23}$ Joule/K, 为玻尔兹曼常数; $q = 1.60 \times 10^{-19}$ C, 为单个电子量; T 是器件的工作温度。

在室温情况下, $T = 300\text{K}$, 则 $V_{\text{th}} = 25.9\text{mV}$ 。式(1-5)中的电压为负值(反向偏置), 由于室温下 V_{th} 仅为 25.9mV , 所以一般情况下都满足 $|V| \gg V_{\text{th}}$, 即 $e^{\frac{V}{V_{\text{th}}}} \approx -I_s$ 。因此, 在实际电路中, 反偏 PN 结漏电流基本与工作电压无关。在目前工艺条件下, I_s 大约为 $1\text{pA}/\mu\text{m}^2$, 与动态功耗相比, 该值非常低, 几乎可以忽略不计。

亚阈值电流 I_{sub} 存在于 MOS 器件的沟道之中, 即使 MOS 处于逻辑关断状态, 其沟道中仍然存在一定的电流, 称为亚阈值(沟道)电流。该电流不仅与器件的尺寸、制造工艺有关, 还取决于栅极电压 V_{gs} 及工作温度。亚阈值电流表达式为

$$I_{\text{sub}} = I_0 e^{(V_{\text{gs}} - V_{\text{T}})/(\alpha \cdot V_{\text{th}})} \quad (1-7)$$

其中, V_{T} 为器件的阈值电压; V_{th} 为热电压; I_0 是 $V_{\text{gs}} = V_{\text{T}}$ 时的电流; α 是常数, 取决于制造工艺, 并在 $1.0 \sim 2.5$ 变化。

由于 $(V_{\text{gs}} - V_{\text{T}})$ 小于 0, 因此随着 V_{gs} 的减小, I_{sub} 呈指数下降, 并且随着阈值电压 V_{T} 的降低, I_{sub} 呈指数增长。

前面提到,为了降低动态切换功耗,降低系统电源电压是一个有效的方法,但是降低电源电压的同时保持阈值不变会造成性能的严重损失。解决这一问题的方法是同时降低器件的阈值电压,但是降低阈值电压会造成亚阈值漏电流的迅速增大。因此,要在性能、动态功耗和静态功耗之间进行权衡取舍。

在 $0.18\mu\text{m}$ 工艺水平下,静态电流功耗所占的比重一般不大于10%;当工艺进入90nm结点以下,泄漏电流急剧增加,在一些45nm的设计中,泄漏电流甚至与动态电流相当。特别是,在采用深亚微米工艺由电池供电的便携式设备中,泄漏是一个很重要的问题。此外,如果整个电路系统可能长时间处于休眠状态,泄漏电流功耗就更加需要进行认真考虑。为了解决这个问题,设计者在设计的各个阶段采用各种方法,从软件到结构再到版图实现。这些方法包括用电源门控将不用的模块关断,以及用多阈值库来折中泄漏电流和速度。

第 2 章 低功耗的设计与实现

现在的集成电路设计都采用自顶向下的设计方法,从系统级到算法级、体系结构级、RTL 级、门级、电路级直到最终工艺级。在以上各个层次中都可以进行功耗的优化,但是效率不同。研究经验表明,在越高的层次进行优化获得的效果越好,所耗费的时间也越少。

2.1 系统级实现

对于系统级,现在在这一层应用最广效果最好的是动态电压频率管理技术(dynamic voltage frequency scaling, DVFS)。这项技术的主要思想是根据系统的任务繁重状态(workload)来动态的调节电路的工作电压和频率,甚至彻底关断某些模块,以达到节省功耗的目的。下面首先介绍 DVFS 的基本原理,再通过针对不同领域的应用来介绍此技术的工作原理及优势。

2.1.1 动态电压/频率调节技术介绍

动态电压/频率调节技术是近年来微处理器低功耗设计的一个突破,允许动态地调节微处理器工作电压/频率,同时保持系统正常工作。现代处理器在设计上一般采用 CMOS 电路,CMOS 电路的功耗由静态功耗与动态功耗两部分组成。静态功耗主要由泄漏电流和开启电流产生,其功耗非常小,可以忽略。由电容的充电与放电形成的动态功耗是 CMOS 电路的主要功耗来源,其功耗的计算公式为 $P = \alpha C_{sw} V^2 f$,其中 α 是比例因子, V 是工作电压, C_{sw} 是负载电容, f 是时钟频率。对一个给定的任务,若执行时间为 t ,则功耗 $E = Pt = \alpha C_{sw} V^2 ft$ 。可以看出,降低频率可以降低功率,但是单纯地降低频率并不能节省能量。因为对于一个给定的任务, ft 是一个常量,只有在降低频率的同时降低电压,才能真正地降低能量的消耗。实际上,对 CMOS 电路而言,降低工作电压 V 的同时会降低时钟频率,因为 CMOS 电路的延时 $T_D \propto V / (V - V_T)^\alpha$,其中 V_T 是阈值电压, α 是与工艺相关的参数。为简便,下面以处理器运行速度或是电压来表征处理器的功耗。

DVFS 技术可以分为动态技术和静态技术。动态技术是根据芯片所运行的应用程序对计算能力的不同需要,动态调节芯片的运行频率和电压。

人们通过观察发现,对许多实时系统,计算需求能被量化成许多具有时间指标的周期性或非周期性的任务,而这些时间指标是事先知道的。静态技术利用这一

事实,脱机决策一种合适的 schedule 来保证所有任务的时间要求都能被满足的前提下,大大减小了系统对运算资源的冗余需求。静态技术包括不同的低功耗模式,芯片内部组件的时钟或电源的按需开关等,在芯片设计阶段(如编译和综合)实现。

1. 静态 DVFS

实时系统环境下通用的 DVFS 模型可以形式化地表述如下。

输入条件:存在一组互不依赖的任务集合 $J = \{J_i = (r_i, d_i, c_i) \mid i = 1, 2, \dots, N\}$, 每个任务用参数 r_i, d_i, c_i 来表征,其中 r_i 表示任务 J_i 已经准备就绪的时间; d_i 表示任务 J_i 必须被执行完成的最迟时间; c_i 表示任务 J_i 的执行长度,即无中断情况下完成任务所需要的 CPU 最大时钟周期数。为简便,假设处理器的电压/频率可以瞬时和连续变化,同时忽略泄漏电流功耗。对任务集 J ,优化电压调度被定义在一系列时间间隔上,每个间隔的起点和终点都是任务的发布时间或截止时间,并且在每个时间间隔内处理器以恒定频率工作。因此,只需要在某个任务的发布时间或截止时间调节处理器的频率,这些时间点被称为调度点。

优先级驱动抢先调度是许多实时系统中最常用的调度,而不同的优先级分配方式展示出完全不同的特性和调度行为。两种最有名的优先级分配机制是最早截止期优先调度(EDF)和固定优先级调度(FP)。

最早截止期优先调度根据任务的绝对截止时间来分配优先级。关于最早截止期优先调度算法,Hong 等在 1998,Pillai 和 Shin,以及 Aydin 等在 2001 年都发表过相关论文,对该算法的研究已经非常成熟。例如,Yao 等在 1995 年提出的算法能在二项式时间内计算出最佳调度方案。下面对该算法作简要介绍(图 2-1)。

定义时间间隔 $[t_a, t_b]$ 的强度 $I(t_a, t_b) = \frac{\sum_i c_i}{t_a - t_b} [t_a, t_b]$, 其中 $[r_i, d_i] \subseteq [t_a, t_b]$ 。

强度值最大的时间间隔 $[t_s, t_f]$ 称为临界间隔。最早截止期优先调度的优化算法便是处理器在临界间隔 $[t_s, t_f]$ 工作在 $I(t_s, t_f)$ 。具体实现方法是对任意任务集,在所有时间间隔内找出第一个临界间隔,从而确定该间隔内处理器工作速度;然后移去该间隔内的所有任务,同时其他任务的发布时间和截止时间作相应调整;在新的任务集上找出临界间隔,并反复重复以上过程直到所有任务都从原任务集中移去。该算法的复杂度是 $O(N^3)$,若采用更加巧妙的数据结构和实现方法可以将复杂度降至 $O(N \log^2 N)$ 。

```

1: LPEDF( $\mathcal{J}_1, S_{opt}$ )
2: Input: A real-time job set  $\mathcal{J} = \{J_1, \dots, J_N\}$ 
3: Output: The optimal voltage schedule  $S_{opt}$ .
4: while  $\mathcal{J}$  is not empty do
5:   Identify the critical interval  $I^* = [t_b, t_f]$ , such that  $\bar{S}(t_i, t_f) = \frac{\sum J_i C_2}{t_f - t_b}$  is the
      maximum, for any  $J_i[r_i, d_i] \subseteq [t_b, t_f]$ .
6:   Add  $I^*$  in  $S_{opt}$ ;
7:   for any job  $J_i \in \mathcal{J}$  do
8:     if  $[r_i, d_i] \subseteq I^*$  then
9:       Remove  $J_i$  from  $\mathcal{J}$ ;
10:    end if
11:    if  $r_i \in I^*$  then
12:       $r_i = t_s$ ;
13:    else if  $r_i > t_f$  then
14:       $r_i = r_i - |I^*|$ ;
15:    end if
16:    if  $d_i \in I^*$  then
17:       $d_i = t_s$ ;
18:    else if  $d_i > t_f$  then
19:       $d_i = d_i - |I^*|$ ;
20:    end if
21:  end for
22: end while

```

图 2-1 EDF 最佳电压调度算法

固定优先级调度因其低额外功耗,易实现和高预测性,在实时系统中得到广泛应用。在学术界,人们对固定优先级调度算法做了大量研究。Yun 和 Kim 严格证明找到 FP 调度的最优解是 NP-hard 问题,从而预示最优的多项式时间算法不可能存在。Shin 等建议通过在最坏的情况下响应时间分析来寻找最优调度,但这种方法仅对起始时间相同的周期任务比较有效,而且那些不是最坏的情况下响应时间的任务处理器的功效可以进一步降低。Manzak 和 Chakrabarti 认为,可以用拉格朗日乘数法求解那些依 RMS 安排的周期性任务。Quan 和 Hu 研究成果显著,在 2001 年提出多项式时间算法,即著名的多项式时间试探 (polynomial-time heuristic),虽然有效却不能保证该算法的效率。2002 年,他们又提出利用穷举算法来寻找优化电压调度,基本思想是将复杂的 FP 问题简化,即采用穷举法寻找 FP 的初级任务集并采用 LPEDF 方法计算相应的功耗,而最低功耗对应 FP 问题的最佳电压方案。遗憾的是,这种算法的计算成本太高(最坏情况下复杂度随任务数目成指数集增长),因此只应用在任务数比较少的实时系统中。Yun 和 Kim 在总结前人成果,特别是对 Quan 和 Hu 的算法作了深入研究后于 2003 年提出完全多项式时间近似方案(FPTAS),该近似算法引入一个额外的输入 $\epsilon (> 0)$,得到的调度方案消耗能量最多是最优解的 $(\epsilon + 1)$ 倍,而运行时间限制在任务数和 $\frac{1}{\epsilon}$ 的多项式以内。由于固定优先级调度是 NP-hard 问题,也许 FPTAS 算是最好的答案,以在多项式时间内找到近似于最优解的调度方案,并且通过改变 ϵ 的值,可以找到无限接