

# AFTL:一种基于高级指令的闪存转换层

杨明伟<sup>1,2</sup>,李亚晖<sup>1,2</sup>,赵益民<sup>1,2</sup>,蔡晓乐<sup>1,2</sup>

(1 中航工业 西安航空计算技术研究所,陕西 西安 710068;2 机载弹载计算机航空科技重点实验室,陕西 西安 710068)

**摘 要:** 为了对多层次并行固态盘提供的多种并行性进行高效利用,提出一种基于冷热数据的 FTL 算法:AFTL,通过对高级指令多分组指令(multi-plane)和交错指令(interleave)的使用,达到并行性利用的最大化,冷热数据识别与内部数据迁移指令(copy-back)、多分组擦除指令的结合则大大降低了垃圾回收的成本,两者的组合最终使得整体性能获得进一步提升.仿真结果显示,在多种负载、多种不同通道数的情况下,相对于同样综合利用多种并行性的动态 FTL 算法,写响应时间平均减少 15.3%,相对于 DLOOP 算法,写响应时间平均减少 78.9%.

**关键词:** 闪存;FTL;并行性;高级指令;垃圾回收

**中图分类号:** TP301

**文献标识码:** A

**文章编号:** 1000—7180(2017)01—0015—06

## AFTL: A Flash Translation Layer Based on Advanced Operations

YANG Ming-wei<sup>1,2</sup>,LI Ya-hui<sup>1,2</sup>,ZHAO Yi-min<sup>1,2</sup>,CAI Xiao-le<sup>1,2</sup>

(1 Xi'an an Aeronautics Computing Technique Research Institute, AVIC,Xi'an 710068, China;2 Aviation Key Laboratory of Science and Technology on Airborne and Missile-borne Computer,Xi'an 710068, China)

**Abstract:** In this research, we proposed a new page mapping FTL called AFTL in order to fully take advantage of several parallelism provided by multi-level parallel solid-state disk. Through using multi-plane operations and interleave operations, AFTL achieves the goal to exploit the parallelism effectively, meanwhile the combination of identification of hot-cold data, copy-back operation and multi-plane-erase operation decreases the cost of garbage collection a lot, all the factors make the performance get largely promoted. The experimental results show that under the several real world workloads with different channel number, AFTL decreases the write response time 15.3% and 78.9% respectively compared with existent FTL which also exploit multi-level parallelism and DLOOP.

**Key words:** flash; flash translation layer; parallelism; advanced operations; garbage collection

## 1 引言

基于闪存的固态盘(solid-state drive, SSD)具有更短的启动时间,更快的随机访问速度和通过消除磁盘中的机械开销得到的更低能耗,这些使得其存储性能比磁盘更加优越,成为了磁盘的有力替代品,更加适应于机载环境<sup>[1-2]</sup>.

然而,NAND Flash 存在“异地更新”、擦除次数有限等问题,不能像磁盘一样直接进行使用.闪存转换层(Flash Translation Layer,FTL)的出现,很好地解决了这个问题.FTL 用来向上屏蔽闪存本身的

读写特征,接受上层文件系统传递的读写请求,将读写请求翻译成闪存所能执行的闪存读写命令,同时监视 SSD 中所有闪存的状态,触发垃圾回收,损耗均衡算法.作为一个承上启下的模块,FTL 不仅影响着 SSD 的读写性能,而且决定了 SSD 的耐久性,故 FTL 的优劣至关重要<sup>[3]</sup>.

传统的 FTL 调度算法,如 DFTL、FAST、Super Block 等都是针对闪存本身的读写特性提出的映射算法<sup>[2]</sup>,并没有考虑到最新出现出的多层次并行固态盘的特点,故而未能有效地利用并行性,都不再适应于如今的固态盘.本文在前人研究的基础上,提出

一种新的 FTL 算法: AFTL, 对多种并行性进行高效利用, 并与高级指令进行结合, 使得整体性能得到极大提高。

## 2 多层次并行固态硬盘结构

在多层次并行固态硬盘中, 存在四个层次上的并行性, 分别是通道间并行、芯片间并行、晶圆间并行和分组间并行, 不同的并行性体现在固态硬盘不同的硬件结构中, 利用方式也各不相同<sup>[4]</sup>。以图 1 为例, 说明多层次并行固态硬盘的特性。

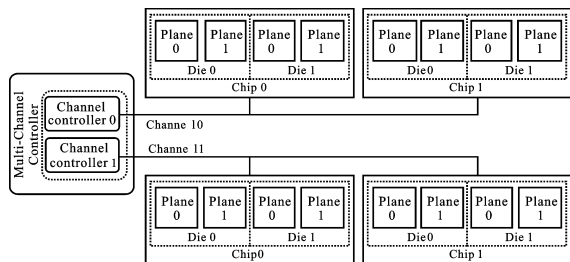


图 1 多层次并行固态硬盘结构

并行固态硬盘中存在多个独立通道, 如图 1 中的 channel 0 和 channel 1, 且各自拥有独立的通道控制器, 所以不会相互影响, 可以并发操作, 即通道间并行。

通道内存在多个芯片, 这些芯片拥有独立的片选信号, 可以独立进行工作, 但共用的数据总线在同一时间内只能为一个芯片提供服务, 所以只能通过流水方式来并行执行。

一个芯片包含多个晶圆, 多个晶圆共用一套外围电路和外部信号线, 与芯片并行类似, 只能以流水的方式并行。

一个晶圆, 由多个分组组成, 每个分组中设置了一个或者多个寄存器, 包含一个页大小的数据寄存器和缓存寄存器, 用来暂存从这个面写入或者读出的物理页数据, 属于同一晶圆的多个分组可以同时执行同一操作, 是闪存层次结构中可以并行的最小粒度<sup>[5-9]</sup>。

## 3 AFTL 设计

AFTL 是一种页级映射方式, 如其他页级映射方式一样, 采用常见的缓存方式, 即只在内存中保留访问较为频繁的映射条目, 来减少映射表对内存的占用, 故而不作详细介绍; 此外, 作为一种动态分配 FTL, 因而读请求对并行性的利用并不是关注的重点, 在读请求服务时进行的多分组读指令和交错读

指令检测, 已能实现对读并行的有效利用。

主要介绍三个部分: 写请求划分、地址映射和垃圾回收。

### 3.1 请求划分与冷热识别

当写请求到来的时候, 首先会对依据写请求的大小进行冷热属性判断, 若是大于所设定的阈值, 则认为该写请求是一个大请求, 将该请求判定为冷请求, 小于则判定为热请求。

这是因为在一般访问模式中, 大请求写入或者更新后, 在接下来很长一段时间内都不会再次进行访问, 相对于小请求而言, 大请求的写入频度特别低, 故而将大请求置为冷数据是完全合理的。在机载环境中, 这种特点更为明显, 大数据如地图、频谱数据库等大数据通常都是一次写入, 长久使用, 而频繁修改的信息如态势信息、飞参数据、系统管理信息等都是小数据, 故而这种冷热数据识别方式能很好适应机载环境下的数据访问模式。

此外, 冷热识别的阈值可以根据使用情况的不同来进行改变, 这就使得本文采用的这种冷热识别方式能很好地适应多种不同情况。

### 3.2 地址映射

地址映射的目标是将 SSD 固有的各种并行性有效地结合, 在不造成页浪费的前提下, 使得性能提升达到最大化。与胡洋提出的并行性利用的顺序不同的是, 本文中并行性利用顺序为: 通道间并行 > 分组间并行 > 晶圆间并行 > 芯片间并行, 以此来执行写调度的分配。

#### 3.2.1 通道间并行

写请求进行分配的时候, 首先会对划分之后的写请求个数进行统计, 得到总写请求个数, 然后将总个数除以总通道数, 得到每个通道应执行的写请求数目, 然后给每个通道传入相应数目的写请求, 同时进行服务大致相同的写请求数目, 实现利用通道间并行性, 且通道间并行性第一的目的。

#### 3.2.2 芯片间并行

通道内部的服务方式是轮转服务, 除非选择的芯片处于忙碌状态, 这样的服务方式使得各个芯片的写入次数趋于均衡。此外, 在对一个芯片进行服务的时候, 自然不能同时对其他芯片进行服务, 直到该芯片释放总线, 由此可见, 芯片内部的并行性利用的优先度是高于芯片间的。

#### 3.2.3 分组间并行

进行芯片内部写请求的分配时, 分组和晶圆之间, 本文优先选择的是分组间并行, 这是因为如果优

先选择晶圆间并行性的话,那么分组间并行性很难得到利用,高级指令中的多分组指令就不能发挥其应有的作用。

在每个分组中,会维持两个写入位置,冷数据写入位置和热数据写入位置,且一个晶圆内的两个分组的冷热位置均保持一致,在不一致的时候,会进行一致化。在分配时为了保证一致性效果,会依次对芯片内的多个晶圆进行查询。

如果该晶圆内两个面的热(冷)数据写入位置相同,且剩余的热(冷)请求数目大于晶圆内分组数目,可执行多分组写操作,并将热(冷)请求数目减去相应值,多分组操作数加 1。

查询之后,如果多分组操作数等于芯片内晶圆数,则可执行多分组交错指令,实现分组、晶圆并行性的同时利用;如果多分组操作数为 1,表明只可进行多分组写操作,实现分组间并行性利用;如果多分组操作数为 0,则进行晶圆间并行性利用。

3.2.4 晶圆间并行

如果多分组操作数为 0,表明当前请求不能执行多分组写指令,存在两种原因:一、写请求数目小于晶圆数;二、多个分组的相应冷热位置不一致,如果强制执行多分组写指令,则会造成页浪费。第一种情况只能使用普通指令,无法利用晶圆间并行性;而第二种情况可以利用交错指令来利用晶圆间并行性。

利用晶圆间并行性时,对芯片内的所有晶圆依次进行如下查询:

如果热(冷)请求数目不为 0,则将热(冷)请求分配到热(冷)写入位置落后的那个面,使得该晶圆内部多个分组热(冷)请求写入位置趋于一致,晶圆操作数加 1,热(冷)请求数减 1。

查询之后,如果晶圆操作数大于芯片内晶圆数,则使用交错指令,利用晶圆间并行性,如果小于,则使用普通指令。

4 实例说明

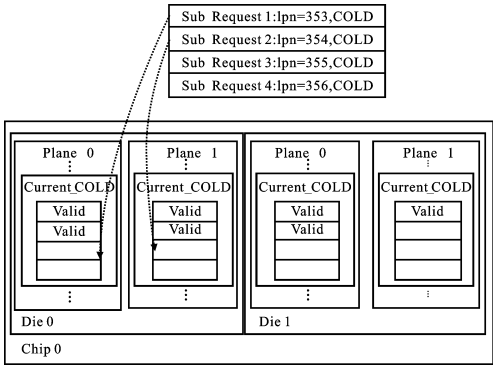
假设存在一个 SSD,其通道数为 2,每个通道存在两个芯片,每个芯片有 2 个晶圆,每个晶圆包含两个分组。当一个地址为 1 412,大小为 32 子页的写请求到来时,将该请求划分为逻辑页地址为 353~360 的 8 个小的写请求,因为其大于设定的冷热阈值,故定义为冷请求,然后将 353~356 分到通道 0,357~360 分到通道 1,进行通道间并行利用。

通道 0 中的芯片 0 空闲,故将 4 个写请求全部

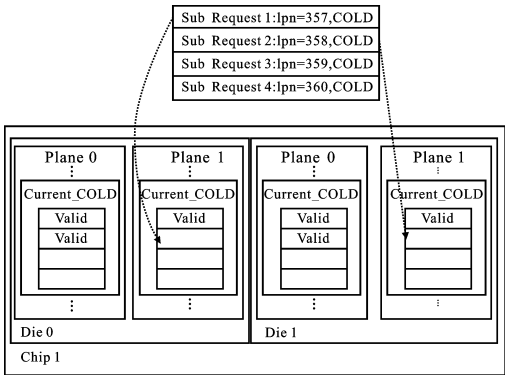
在芯片 0 执行;而通道 1 中芯片 0 属于忙碌状态,故而全部在芯片 1 中执行。

通道 0 的芯片 0 中,对晶圆 0 进行查询,两个分组冷写入位置相同,都为 Current\_COLD 块的第三个页位置,故多分组数变为 1;而晶圆 1 的两个分组冷写入位置不同,故多分组数不变。故芯片 0 的多分组操作数为 1,只将 353 和 354 分配给晶圆 0,执行多分组写指令,剩下的写请求等待下次重新分配通道和芯片,如图 2 中的(a)所示。

通道 1 的芯片 1 中,对晶圆 0 和晶圆 1 进行查询,查询之后,多分组操作数为 0,利用晶圆间并行。对晶圆 0 进行查询,冷请求数不为 0,晶圆操作数加 1;同理,晶圆 1 中,晶圆操作数再加 1。总晶圆操作数为 2,将 357 和 358 依次分配给晶圆 0 的分组 1 和晶圆 1 的分组 1,执行交错指令,剩下的写请求等待下次重新分配,如图 2 中的(b)所示。



(a)分组间并行调度



(b)晶圆间并行调度

图 2 写请求分配

5 垃圾回收

当一个晶圆中的空闲页的数目小于设定比率时,进行垃圾回收操作。这个过程的具体流程如下:

(1)在该晶圆中,选择多个分组中块偏移相同的多个物理块无效页数之和最多的位置,将该位置

的多个物理块同时置为回收块;

(2)计算出多个回收块中的有效页总数目,对该晶圆内的多个分组交替进行有效页复制,在复制过程中,保持多个分组的写入平衡的同时,优先寻找能使用内部数据迁移指令的有效页;

(3)有效页复制完成后,对多个回收块使用多分组擦除指令进行擦除。

假如一个晶圆内有两个分组,每个分组内有 64 个块,每个块内有 4 页,进行垃圾回收时,分组 0 和分组 1 的有效页数分别为 3 和 1,对两个分组交替进行有效页复制,过程如图 3 所示. 首先对分组 1 进行有效页复制(分组 1 写入位置落后),有效页和写入位置奇偶相同且位于同一分组中,故使用内部数据迁移指令完成复制;接下来对分组 0 进行有效页复制,分组 0 中第 0 页不满足内部数据迁移的条件,而第 1 页可以,故选择第一页使用内部数据迁移指令;然后又回到分组 1,此时,分组 1 内的回收块已经不存在有效页,故从分组 0 中选取有效页,如 3 线标明的那样,有效页与写入位置不在同一分组,故只能使用普通指令,通过总线,进行有效页的复制;最后,对分组 0 的最后一个有效页使用内部数据迁移指令,完成整个有效页的复制过程. 有效页的复制完成后,使用多分组擦除指令将两个分组的回收块同时进行擦除,垃圾回收结束。

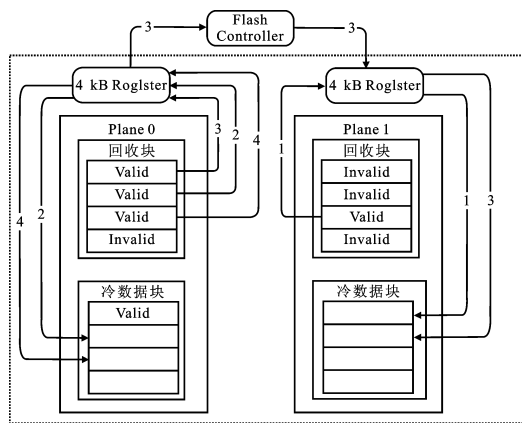


图 3 有效页复制示意图

整个垃圾回收的过程分析如下:(1)回收块中的有效页更倾向于冷数据,将其转移到冷数据块中,这样可以减少下次垃圾回收时有效页的数目,降低开销;(2)分组间交替进行有效页复制的策略保证了两个分组冷写入位置的平衡,方便接下来多分组写指令的使用;(3)冷热数据分开存储,可以在整体上减少热数据块中有效页的数目,而回收块一般都位于

热数据块,故而可以减少垃圾回收的开销;(4)使用多分组擦除指令可以将擦除操作所需时间变为原来的一半,降低垃圾回收的开销。

## 6 性能评估

### 6.1 实验设置及性能参数

本文在 SSDsim 中实现了 AFTL 算法、DLOOP,此外,SSDsim 还自带了胡洋按照最优化并行性利用顺序设计的动态 FTL 和静态 FTL(分别称为 Hu-FTL-D 和 Hu-FTL-S),因而本文主要与这三种利用并行性的 FTL 进行性能上的比较。

实验结果给出了 4 种利用并行性的 FTL 算法在 4 通道、8 通道和 16 通道情况下,平均读响应时间相对于 DLOOP 的比率. 可以看出,AFTL 的读响应时间在大多数情况下是少于 Hu-FTL-D 的,而且随着通道数的增加,这种趋势更加明显,特别在 RADIUS 负载中,平均读响应时间减少了 20% 以上;但是 AFTL 的读响应时间一直高于 Hu-FTL-S,这是因为 AFTL 是一种动态分配,而 Hu-FTL-S 是一种静态分配,在同等条件下,静态分配的读请求响应时间是小于动态分配的;DLOOP 作为一种静态分配,两种动态分配 AFTL 和 Hu-FTL-D 的读响应时间在很多情况下是低于 DLOOP 的,随着通道数的增长,减少的比率越来越高,在 16 通道中,读响应时间平均减少超过了 30%,可见并行性利用方式的不同和高级指令的使用,会对读性能的造成很大影响。

图 4 给出了 4 种利用并行性的 FTL 算法在 4 通道、8 通道和 16 通道情况下,平均写响应时间相对于 DLOOP 的比率. 从中可以看出,AFTL 写响应时间一直低于 Hu-FTL-D,各种情况下写响应时间平均减少 15.3%. 在 4 通道中,写响应时间相对 Hu-FTL-D 平均减少了 19.3%,随着通道数的增加,在 16 通道时降低比率降低到了 11.8%,这是因为 AFTL 在垃圾回收上的开销是低于 Hu-FTL-D 的,而随着通道数的增加,垃圾回收的次数显著降低,故而写性能相对提升效果降低。

动态分配 AFTL 和 Hu-FTL-D 的写响应时间一直低于静态分配 Hu-FTL-S,且远低于 DLOOP, AFTL 写响应时间最低时仅有 DLOOP 的 11%,各种情况平均下来页只有 21.1%;此外同样作为静态分配,Hu-FTL-S 的平均写响应时间不到 DLOOP 的 40%,写性能上存在显著优越性。

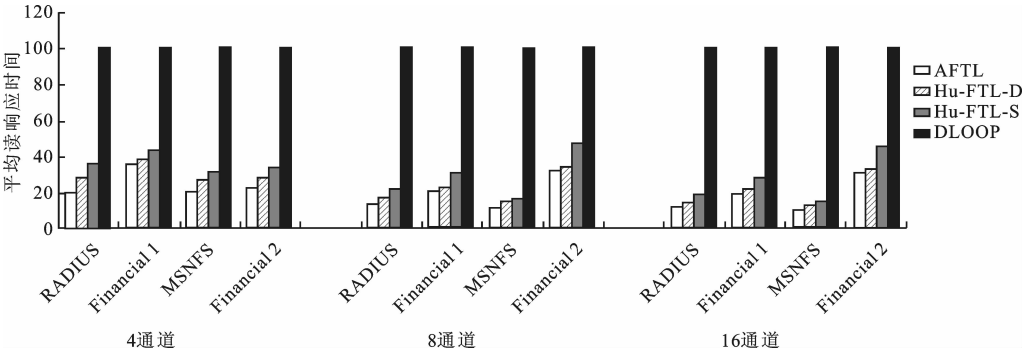


图 4 各种情况下的写响应时间

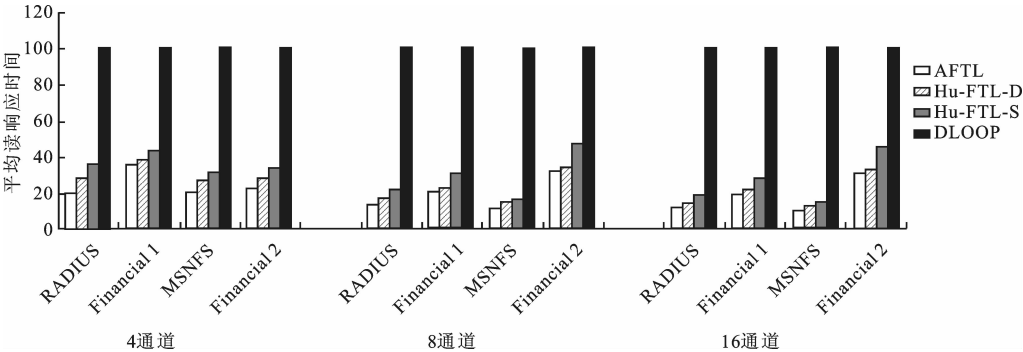


图 5 各种情况下的总擦除次数

图 5 展示了各种情况下各种 FTL 的总擦除次数,可以看到,在大多数情况下,AFTL 的总擦除次数明显少于其他三种情况下的总擦除次数,可见我们所使用的冷热数据识别方式有效地减少了总擦除次数,从而延长了 SSD 的寿命;此外,AFTL 中每个擦除操作都是使用多分组擦除操作的,即一次对两块同时进行擦除,所以实际上 AFTL 擦除操作所消耗的时间是同等擦除数所需时间的一半,而在 Hu-FTL-D 和 Hu-FTL-S 中多分组擦除操作和交错擦除操作只是偶尔使用,比率是远远低于 AFTL 的,故 AFTL 中垃圾回收的总开销明显小于其他三种。

总的来说,AFTL 的综合性能是明显优于其他三种 FTL,达到了综合利用各种并行性和高级指令来提升性能的目的,满足了多层次并行固态硬盘的需求。

7 结束语

提出了一种有效利用 4 种并行性和多种高级指令的闪存转换层 AFTL。AFTL 主要分为两大部分:第一部分是地址映射,通过动态地写请求分配和高级指令的使用,有效地利用了多层次并行固态硬盘提供的并行性,极大地提高了读写性能;第二部分是基于冷热数据的垃圾回收算法,通过和内部数据迁移

指令、多分组擦除指令结合,大大减少了垃圾回收带来的开销,为整体性能的提升和寿命的延长做出了贡献。模拟实验结果表明,与其他 3 种利用并行性的 FTL 算法相比,AFTL 的整体性能最优。

参考文献:

- [1] 胡宁,杨琼,王冬. 基于分组的 NandFlash 块管理方法[J]. 微电子学与计算机,2015,32(3):19-22.
- [2] Cai X, Li Y, Zhang Y, et al. An efficient address mapping algorithm for multichannel structure SSD[C]//2015 11<sup>th</sup> International Conference on Computational Intelligence and Security (CIS). China, Guangzhou, IEEE,2015:121-124.
- [3] Park S, Seo E, Shin Y K, et al. Exploiting internal parallelism of flash-based SSDs[J]. IEEE Computer Architecture Letters, 2010,9(1):9-12.
- [4] 胡洋. 高性能 SSD 的多级并行性及算法研究[D]. 武汉:华中科技大学, 2012.
- [5] Zhou Y, Wu F, Huang P, et al. An efficient page-level FTL to optimize address translation in flash memory [C]// Proceedings of the Tenth European Conference on Computer Systems. USA, New York, ACM, 2015: 12-27
- [6] Narayanan D, Thereska E, Donnelly A, et al. Migrating server storage to SSDs: analysis of tradeoffs[J].

ACM European. Computer Systems, 2009, 87(2): 145-158.

[7] Lee S W, Park D J, Chung T S, et al. FAST: A log-buffer based ftl scheme with fully associative sector translation[J]. Kips Transactions Parta, 2005, 12A(3): 205-214.

[8] Hu Y, Jiang H, Feng D, et al. Performance impact and interplay of SSD parallelism through advanced commands, allocation strategy and data granularity[C] //Proceedings of the 25th International Conference on Supercomputing (ICS'11). Tuscon, AZ, 2011.

[9] Abdurrah A R, Xie T, Wang W. DLOOP: a flash translation layer exploiting plane-level parallelism[C]

// Parallel & Distributed Processing (IPDPS), 2013 IEEE 27th International Symposium on. Cambridge. IEEE, 2013: 908-918.

作者简介:

杨明伟 男, (1992-), 硕士研究生. 研究方向为高安全性操作系统. E-mail: ymwei\_avic@163. com.

李亚晖 男, (1976-), 博士, 高级工程师. 研究方向为嵌入式操作系统、计算机网络、信息安全.

赵益民 男, (1992-), 硕士研究生. 研究方向为计算机应用.

蔡晓乐 男, (1991-), 硕士研究生, 助理工程师. 研究方向为计算机应用.

(上接第 14 页)

参考文献:

[1] 曹世华, 赵方. 2. 4G 低功耗主动式 RFID 识别系统设计与实现[J]. 沈阳师范大学学报, 2008, 26(2): 179-182.

[2] 陈拓明. RFID-SIM 卡芯片集成技术研究及应用[D]. 长春: 长春理工大学, 2013.

[3] Hung Y C. Time-interleaved CMOS chip design of Manchester and Miller encoder for RFID application [J]. Analog Integrated Circuits & Signal Processing, 2012, 71(3): 549-560.

[4] 舒平平, 王小军, 赵嘎. 基于 FPGA 的曼彻斯特编码 DSG 的设计与实现[J]. 试验科学与技术, 2012, 10(4): 6-8.

[5] 孙枫叶, 张科峰, 刘冬生, 等. 高频 RFID 标签曼彻斯特

编码电路的设计[J]. 电子技术应用, 2006, 32(10): 108-110.

[6] 朱恒军, 王发智, 姚仲敏. 基于单片机的数字滤波算法分析与实现[J]. 齐齐哈尔大学学报, 2008, 56(6): 53-54.

作者简介:

陈毅成 男, (1980-), 博士. 研究方向为超大规模集成电路与 SOC 系统设计、RFID 及移动支付. E-mail: chen yicheng@zkxltech. com.

郑朝霞 女, (1975-), 博士, 副教授. 研究方向为超大规模集成电路与 SOC 系统设计、嵌入式及信息安全.

刘政林 男, (1968-), 博士, 教授. 研究方向为超大规模集成电路与 SOC 系统设计、嵌入式及信息安全.