

基于LDPC读延迟的刷新和副本结合策略优化方案

张耀方, 李培轩, 谢平

引用本文

张耀方, 李培轩, 谢平. [基于LDPC读延迟的刷新和副本结合策略优化方案](#)[J]. 计算机科学, 2023, 50(7): 38-45.

ZHANG Yaofang, LI Peixuan, XIE Ping. [Policy Optimization Scheme of Refresh and Duplication Combination Based on LDPC Read Delay](#) [J]. Computer Science, 2023, 50(7): 38-45.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于机器学习的微服务负载均衡算法研究](#)

Study on Load Balancing Algorithm of Microservices Based on Machine Learning

计算机科学, 2023, 50(5): 313-321. <https://doi.org/10.11896/jsjcx.220400019>

[一种存储介质优化的大规模图遍历方法研究](#)

Study on Big Graph Traversals for Storage Medium Optimization

计算机科学, 2023, 50(1): 34-40. <https://doi.org/10.11896/jsjcx.211100049>

[基于机器学习的剩余使用寿命预测实证研究](#)

Empirical Research on Remaining Useful Life Prediction Based on Machine Learning

计算机科学, 2022, 49(11A): 211100285-9. <https://doi.org/10.11896/jsjcx.211100285>

[基于Stacking多模型融合的IGBT器件寿命的机器学习预测算法研究](#)

Study on Machine Learning Algorithms for Life Prediction of IGBT Devices Based on Stacking Multi-model Fusion

计算机科学, 2022, 49(6A): 784-789. <https://doi.org/10.11896/jsjcx.210400030>

[一种基于有向感知区域调整的强栅栏构建算法](#)

Strong Barrier Construction Algorithm Based on Adjustment of Directional Sensing Area

计算机科学, 2022, 49(6A): 612-618. <https://doi.org/10.11896/jsjcx.210300291>

基于 LDPC 读延迟的刷新和副本结合策略优化方案

张耀方^{1,2,3,4} 李培轩^{1,2,3,4} 谢 平^{1,2,3,4,5}

- 1 青海师范大学计算机学院 西宁 810016
2 省部共建藏语智能信息处理及应用国家重点实验室 西宁 810008
3 青海省物联网重点实验室 西宁 810016
4 高原科学与可持续发展研究院 西宁 810016
5 青海师范大学网络信息管理中心 西宁 810006
(1253906861@qq.com)

摘 要 针对闪存存储器的密度增大和容量增加导致可靠性下降的问题,提出了一种基于 LDPC 读延迟的刷新和副本结合策略优化方案。通常,原始策略是在闪存存储器上加一个 LDPC 码模块,使用硬解码和软解码对数据进行纠错。而传统的刷新策略是在原始策略的基础上,在 LDPC 软解码未能纠错时,使用刷新策略对其进行纠错。本方案基于 LDPC 软解码 7 个量化级别的特性,并以此为判定条件,采用分析对比的方法确定刷新的条件是量化级别到达 3,副本的条件是级别达到 5,将两种方式合理地应用在 LDPC 软解码方式中。与前两种策略相比,所提方法缩短了闪存存储器的平均响应时间,在一定程度上提升了闪存存储器的读取性能。在模拟器 disksim+ssd 的扩展平台上进行仿真,实验结果表明,相比原始策略,该方案平均响应时间均值降低了 10%;与传统的刷新策略相比,延长了闪存存储器的寿命。

关键词: 闪存存储器;低密奇偶校验码;刷新;副本;响应时间;寿命

中图法分类号 TP393

Policy Optimization Scheme of Refresh and Duplication Combination Based on LDPC Read Delay

ZHANG Yaofang^{1,2,3,4}, LI Peixuan^{1,2,3,4} and XIE Ping^{1,2,3,4,5}

- 1 The College of Computer, Qinghai Normal University, Xining 810016, China
2 The State Key Laboratory of Tibetan Intelligent Information Processing and Application, Xining 810008, China
3 The Key Laboratory of Internet of Things of Qinghai Province, Xining 810016, China
4 Academy of Plateau Science and Sustainability, Xining 810016, China
5 Network Information Management Center of Qinghai Normal University, Xining 810006, China

Abstract Aiming at the problem of reliability degradation caused by the increase of the density and capacity of flash memory, an optimization scheme of refresh and copy combination strategy based on LDPC read delay is proposed. In general, the original strategy is to add LDPC code module to flash memory and use hard and soft decoding to correct data errors. The traditional refresh strategy is based on the original strategy, when the LDPC soft decoding fails to correct the error, the refresh strategy is used to correct the error. The scheme is based on the characteristics of LDPC soft decoding 7 quantitative levels, and takes this as the judgment condition, using the method of analysis and comparison to determine that the condition of refresh is level 3, the condition of the copy is level 5, and the two methods are reasonably applied in the LDPC soft decoding mode. Compared with the previous two strategies, the average response time of flash memory is reduced, and the read performance of flash memory is improved to a certain extent. Simulation is performed on the extended platform of the simulator disksim+ssd, and experimental results show that, the average response time of this scheme is 10% shorter than that of the original strategy, and its flash memory lifetime is prolonged compared to traditional refresh strategies.

Keywords Flash memory, Low-density parity-check codes, Refresh, Duplication, Response time, Lifetime

到稿日期:2022-09-19 返修日期:2023-02-27

基金项目:国家自然科学基金(61762075);青海省物联网重点实验室项目(2022-ZJ-Y21)

This work was supported by the National Natural Science Foundation of China(61762075) and Key Laboratory of IoT of Qinghai(2022-ZJ-Y21).

通信作者:谢平(xieping@qhnu.edu.cn)

1 引言

闪存作为一种非易失性存储器,具有读取速度快、体积小、重量轻等优点。过去,它被用作 DRAM 和主存储器之间的高速缓存,以提高 I/O 传输速度。NAND 闪存每个单元仅存储一位,称为单级单元闪存 (Single-Level Cell, SLC)。现在,开发者已经开发出高密度闪存,例如多级单元闪存 (Multi-Level Cell, MLC)^[1]、三级单元闪存 (Trinary-Level Cell, TLC)^[2] 和 3D NAND 闪存^[3]。其中,SLC,MLC 和 TLC 的区别如图 1 所示。闪存的成本和价格逐渐下降,容量逐渐增加。高性能闪存也被企业所接受,并广泛用于移动存储设备和笔记本电脑^[4],但这会导致许多问题。存储密度越来越高,同时,存储器件的尺寸也在变小^[5-6],存储单元的电器耦合性变得愈加复杂,例如氧化层变薄,读取位的读电压控制操作变得更精密等,这就导致 NAND flash 更容易出错,闪存设备的可靠性变得越来越低。信息的爆炸式增长也给信息的存储造成了很大的压力,这种压力来自于快速增长的信息量对存储容量的要求越来越高,以及数据本身的价值越来越高,有些甚至远远超过了信息存储介质的价值,因此对闪存的可靠性要求也越来越高。

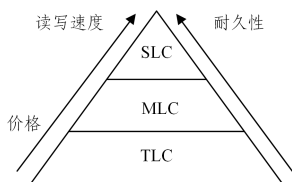


图 1 SLC,MLC,TLC 对比图

Fig. 1 Comparison of SLC,MLC and TLC

随着人们对闪存存储器的可靠性提出了新的要求,原有的各种编码技术开始捉襟见肘。光盘和磁介质硬盘中的里德·索罗门码 (Reed-Solomon Codes, RS)、单层单元闪存存储器中的循环检验码 (Bose-Chaudhuri-Hocquenghem Codes, BCH) 已经到达或接近其极限^[7],迫切需要一种新的编码。低密度奇偶校验码 (Low-density Parity-check Codes, LDPC) 虽然有着已知最接近香农限的纠错能力,但其因编译码过程比较复杂,在其发现初期并未进入实际应用领域。而随着近年来半导体技术的飞速发展,人们已经可以实现基于超大规模集成电路的 LDPC 码编译码工作,因此继实际应用于通信领域后,LDPC 码必将在存储领域得到广泛应用。基于此,本文对 LDPC 码在闪存存储器中的应用展开了前瞻性研究。

随着闪存存储器的存储容量越来越大,数据差错率也越来越高,LDPC 码凭借强大的纠错能力在闪存存储器中的作用也越来越突出。针对闪存寿命和可靠度存在的问题,采用多元 LDPC 硬判决技术^[8],提高了闪存存储器中的译码可靠性和擦除寿命。而 LDPC 码的编译码复杂度较高,直接使用 LDPC 码对数据进行纠错会造成大量的时间延迟,因此需要提出新的方案来优化其读取性能。为了在不影响闪存的纠错性能的前提下提升闪存的读性能,已有的许多工作都致力于根据 LDPC 解码的特点来提出解决方案。例如,文献[9-10]通过研究闪存的出错特征并将其融合到 LDPC 的解码过程

中来加快 LDPC 的解码。为了降低或者解决 LDPC 码在闪存中纠错带来的读延迟增加问题,研究从 LDPC 码的解码方式着手。LDPC 解码方式分为硬解码和软解码,软解码的解码过程中会涉及 7 个量化级别,并且解码能力会随着量化级别的提升而增大,但是造成的延迟也会越来越大。鉴于此,根据量化级别的特点提出了一种将刷新和副本结合应用到 LDPC 解码方式的方案,该方案在一定程度上提升了闪存存储器的读取性能并且延长了其寿命。

2 研究背景和动机

本节将从闪存和低密度奇偶校验码的基础知识出发,介绍其研究背景和动机。

2.1 闪存的基础简介

闪存作为目前发展速度最快、适用范围最广的存储设备,已经被广泛应用于嵌入式系统、个人计算机以及高性能计算机。作为存储设备,闪存具有诸多优点,比如随机访问性能好、功耗低、无振动以及尺寸小等。相对于传统的磁盘,基于闪存的存储系统有很好的性能,其访问速度比磁盘快一个数量级,功耗只有磁盘的六分之一。目前,闪存存在嵌入式系统以及个人电脑中已经开始替代磁盘成为新的存储设备,而在高性能计算机中,闪存存储器的应用也得到了深入的发展。

闪存通过在存储元中存储一定量的电荷来表示数据,在每个存储元存储 n 比特数据的闪存,存储元共有 $2n$ 个不同的电压状态。图 2 为每个存储单元存储 2 比特信息的 MLC 的电压分布,由 4 个状态 ((11, 10, 01, 00) 来表示 2 比特数据。图中电压状态的分布和状态之间的间隔直接影响闪存中数据的可靠性。为了满足对存储容量日益增加的需求,目前嵌入式系统中闪存的发展主要依赖于存储元尺寸的缩小(近年来闪存的尺寸从 90nm 级别发展成为 10 nm 级别)、存储密度的增大(单个存储元比特数目的增加,从每存储元 1 比特发展到现在的 6 比特),以及存储元多层堆叠(3D 闪存)的方式来实现。这些发展趋势使得单个存储元在更小的尺寸情况下,需要对其电压做更细粒度的区分,电压状态之间的容错间隙变小。3D 的出现使得存储元之间的连接更为复杂,因此闪存存储元更容易受到各种错误源的影响,数据更容易出错,闪存的可靠性随之降低。

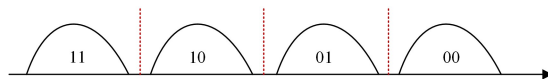


图 2 MLC 阈值电压分布

Fig. 2 MLC threshold voltage distribution

闪存的写操作通过向存储元中充一定量的电荷来表示数据,而读取操作通过判定存储元中所存储的电荷量来确定其中存放的数据。在整个过程中,闪存存储元中的电量会由于多种原因发生改变,例如,编程时相邻存储元的相互影响、保存时间以及擦除次数等,都会对存储元中的电荷产生影响,导致存储元的电压状态发生偏移,数据发生错误。

2.2 LDPC 码基础简介

1962 年, Gallager 第一次提出 LDPC 码的概念^[11]。但由于当时的技术条件有限,无法实现可行的译码算法,LDPC 码

一直没有得到应用和发展。直到 1995 年前后,学者对 LDPC 码重新进行了研究,提出了解码算法,并发现 LDPC 具有良好的性能,迅速引起了研究者的极大关注。在接下来的 15 年中,研究人员通过进一步的研究和发展,在各方面取得了突破性的进展,LDPC 码的相关技术也日趋成熟。由于其具有极强的纠错能力,目前闪存系统中也开始采用 LDPC 码替代传统的 BCH 码作为纠错码。

LDPC 码是一种稀疏校验矩阵线性分组码,其校验矩阵只含有很少量的非零元素。正是校验矩阵的这种稀疏性,保证了译码复杂度和最小码距都只随码长呈线性增加。除了校验矩阵是稀疏矩阵外,码本身与任何其他分组码并无一致。如果现有的分组码可以被稀疏矩阵所表达,那么用于码的迭代解码算法也可以成功地移植到它身上。不同的是,LDPC 码的设计是先构造一个校验矩阵,如式(1)所示,校验矩阵 H 的大小为 $M \times N$,其中 N 为发送的码字长度。设 K 为信息位长度,校验信息长度为 $M = N - K$ 。然后通过它来构造一个生成矩阵进行后续编码,将收到的信息位和矩阵的每一行的每一个元素对应相乘再相加,满足 $H \cdot C^T = 0$,即编码成功。

它的稀疏奇偶校验矩阵通常用 Tanner 图来表示,如图 3 所示,Tanner 图也是 LDPC 码校验矩阵的一种表示。Tanner 图包含两类顶点:1) n 个码字比特顶点(称为比特节点),分别与校验矩阵的各列相对应;2) m 个校验方程顶点(称为校验节点),分别与校验矩阵的各行对应。校验矩阵的每行代表一个校验方程,每列代表一个码字比特。矩阵中的“1”表示对应的比特节点包含在相应的校验方程中,每个比特节点都有一个可靠性信息,该信息通过比特节点和校验节点之间的信息传输进行迭代更新。解码器基于更新的比特节点可靠性信息做出最佳决定。如果比特节点和校验节点之间传播的可靠性信息更准确,解码器可以更快做出决定。

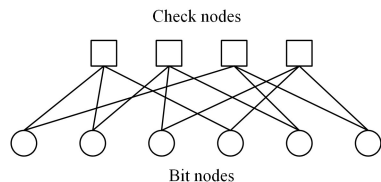


图 3 Tanner 图的一个例子

Fig. 3 An example of tanner graph

LDPC 的纠错原理是基于概率信息进行解码,解码分为硬解码和软解码。LDPC 硬解码与 BCH 类似,即在闪存的两个相邻电压状态之间设置一个读取电压,如果存储元电压比该电压小则表示处于左边的电压状态;反之,处于读取电压右边的电压状态。在数据错误率较小或者不存在错误的情况下,使用硬解码就可以正确读出数据。然而,当闪存数据错误率升高,存储元电压发生偏移时,如图 4 所示,相邻电压状态之间可能发生交叉。这时需要在相邻电压状态之间设置多个读取电压,来确定处于中间错误区域的存储元属于哪个电压状态。

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \quad (1)$$

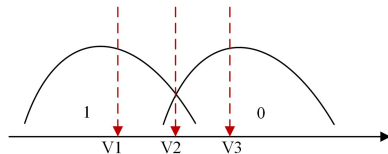


图 4 电压偏移状态

Fig. 4 Voltage offset status

2.3 研究动机

当执行读取请求时,闪存控制器总是在出错数据返回主机之前对数据进行纠错,此时就需要用到 LDPC 解码。LDPC 的解码流程总是从硬判决解码开始,硬判决解码只需要一次读取和一次解码,其优点是过程简单,计算量少,实现简单,纠错速度快,在相同误码率的情况下,可以用较小的纠错成本完成。在闪存的早期寿命中,错误较少,LDPC 硬解码可以完成纠错工作。但硬判决译码的缺点是纠错能力差,在闪存中进行大量 P/E 循环后,硬件会磨损,错误会大大增加。此外,闪存编程错误、读取错误和保留错误对闪存数据的影响也在增加。当硬解码失败时,LDPC 执行软解码,如图 5 所示。软解码具有很强的纠错能力,但计算量很大,因为软决策解码需要经过 3 个步骤^[12]:

- (1) 重读操作:重读操作根据软数据的构造方式在闪存中读取一组或多组数据,可以获得多组对解码有用的信息。
- (2) 数据重组:重读操作将读出由“0”和“1”组成的多组不同的闪存数据信息。LDPC 将这些信息映射成软数据,通过 LDPC 解码进行解码。
- (3) 软判决解码:软数据构造完成后,码字中的每个对数似然比(Log likelihood Ratio, LLR)都被传递给 LDPC 解码器。LLR 中的每个实数表示每位数据为“0”和“1”的概率。一个正数代表为“0”的概率更大,一个负数代表为“1”的概率更大,这样就可以纠正闪存的错误。

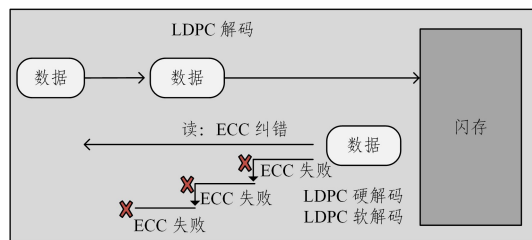


图 5 LDPC 解码流程

Fig. 5 LDPC decoding process

LDPC 使用迭代方式进行解码,解码迭代的目的是积累比特决策信息。迭代次数越多,位决策的可靠性越高,就会导致更多的延迟。事实上,LDPC 软解码共有 7 个读取级别,它会一直读取数据,直到解码成功,但随着量化级别的提升,所造成的延迟会越来越大。鉴于此,本文提出了一种基于 LDPC 读延迟的刷新和副本结合的策略优化方案来提高闪存存储器的读取性能。

3 相关工作

传统纠错码 BCH 以及 RS 等在过去的闪存应用中扮演了重要的角色,但随着随机错误的增大,传统纠错码逐渐不能

满足要求,于是通信领域研究中非常火的一种纠错码——低密度奇偶校验码被运用在 NAND Flash 中。与通信领域应用特点不同,存储领域要求高编码率、解码速度更快以及硬件空间可扩展等,这给 LDPC 码在存储领域的实现带来很大的挑战。低密度奇偶校验码是近 20 年重新引起研究者极大关注的一种纠错技术。LDPC 码不仅性能逼近香农极限,纠错能力极强,而且解码复杂度低,解码速度极快,码率自适应。本文围绕在闪存纠错应用中的特点,介绍基于 LDPC 读延迟的刷新技术和副本技术在闪存存储器中的应用。

3.1 通过刷新技术优化闪存读取性能

通过闪存的刷新操作可以在一定程度上解决延迟问题,刷新可以将出现错误数据的页面重新写入无错的数据,将高级感测延迟的页面重置回延迟最低的硬判决解码,当 LDPC 达到高读取延迟时,会激活刷新操作以纠正错误并将无错误的数据重新编程为新块,这有利于后续读取的读取性能,如图 6 所示。

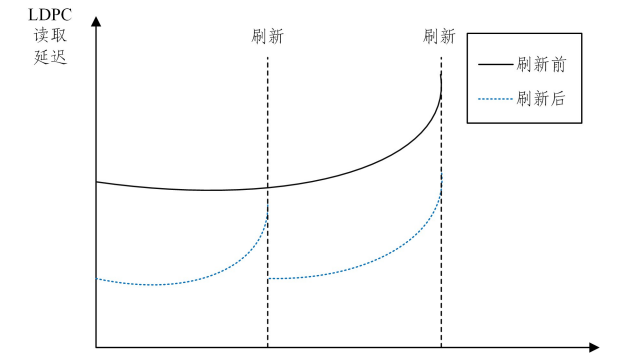


图 6 刷新前后对比图
Fig. 6 Comparison before and after refreshing

Cai 等^[13]提出用闪存校正和刷新的策略来减少闪存的擦除次数和响应时间。Pan 等^[14]提出了一种准非易失闪存存储技术,该技术可以加强非易失性并使用刷新来提高 P/E 循环耐久性。Li 等^[15]为了缓解保留错误导致的闪存访问速度下降的问题,应用刷新的概念,针对数据读取的访问特性,执行页面级刷新,大大提升了闪存存储器的读取性能。Lv 等^[16]根据 LDPC 的纠错时间来判断是否进行刷新操作,该方案的基本思想是在数据需要较长的纠错延迟之前刷新数据,从而降低纠错延迟。Di 等^[17]研究了闪存的工艺差异,发现闪存块最长保留时间因闪存块而异。该方案为不同的设备设定不同的刷新周期,利用闪存块的保留时间的预测值来最大限度地降低刷新成本,使闪存的寿命和性能都得到了显著提高。Luo 等^[18]将热写数据与冷写数据的判断加入刷新的判断中,缩短了数据块的 P/E 周期,延长了闪存的寿命。

不断对刷新操作进行合理的优化^[19-20],可以极大延长闪存存储器的使用寿命并提升其读取性能,进一步满足企业和个人的数据存储需求。

3.2 通过副本技术优化闪存读取性能

现代数据中心内的分布式存储系统主要通过两种方式来保证数据的可靠性和可用性:副本和纠删码。副本和纠删码各有优缺点,对于副本技术而言,每个副本都能提供独立的服务,当一个副本出现故障时,另一个副本可以快速接管,并且

在进行数据恢复时,只需要一个同样大小的数据块,所以恢复速度较快且方便,但是副本需要的空间较大。纠删码在相同环境下更能提高系统的可用性和可靠性,并且在拥有相同容错能力的情况下,纠删码技术的存储开销更低,在大数据量下有明显优势。根据纠删码的特性,对于 (N,M) 的保护级别,最多任意 M 个数据块丢失后,丢失的数据需要从其他 N 个数据块恢复,对网络要求较高,而且实现难度较大。考虑到纠删码在重建数据时的缺点以及实现的难度,现在大多数局域和广域的分布式存储系统都采用副本冗余技术,例如存储系统 Google 的 GFS^[21]、Amazon 的 Dynamo^[22]、Coda^[23]、Ceph^[24]、Skute^[25]、LBFS^[26]、Pangaea^[27]都单纯采用了副本冗余技术。而随着数据量的增多,纠删码的优势也逐渐体现出来,因此很多公司也在研究纠删码技术,Microsoft 的 Windows Azure Storage^[28]和 OceanStore^[29]就同时采用了纠删码和副本技术。

基于对刷新技术和副本技术的研究,本文将刷新技术和副本技术结合应用到基于 LDPC 读延迟的闪存存储器中,以缩短平均响应时间和闪存存储器的读取响应时间。

4 提出的方案

本节主要介绍基于 LDPC 读延迟的刷新和副本结合策略的方法,首先根据数据的冷热程度来选择恢复数据的方法,当数据的读取频率为 1 时定义为冷数据,当读取频率大于 1 时定义为热数据。本文方法主要是针对热数据而言,当热数据采用 LDPC 软解码的量化级别大于或等于 3 时,采用刷新的方法;当量化级别大于或等于 5 时,采用副本来恢复出错的数据。这种方法利用了刷新和副本的优点,在一定程度上减少了闪存存储器的读取延迟,提升了闪存存储器的读取性能。该方案主要分为 3 部分:量化级别的确定、选择刷新的条件以及选择副本的条件。根据对量化级别的分析确定了适合刷新的量化级别为 3,适合副本的量化级别为 5,两者结合缩短了闪存存储器的平均响应时间,提升了读取性能,延长了寿命。接下来介绍这两部分内容以及过程中重要的步骤和算法。

4.1 使用刷新和副本的量化级别的确定方案

由于 LDPC 的解码方式分为硬解码和软解码,软解码的量化级别分为 7 级,根据量化级别的特点,每提升一个级别,读取延迟就会增加一部分,延迟如表 1 所列。本文根据 LDPC 软解码的级别特性,进行大量的实验研究量化级别之间的关系,并在不同的负载下进行实验,实验结果表明,在量化级别大于或等于 3 时采用刷新技术,大于或等于 5 时采用副本技术,可以提升闪存存储器的读取性能,延长其寿命。

表 1 软解码每一步延迟	
Table 1 Soft decode each step delay	
LDPC 软解码步骤	读取延迟/us
第一次读取	50
额外读取	14
第一次读取的数据传输	20
一次额外读取的数据传输	10
执行一次软解码	15

4.2 基于 LDPC 读延迟的刷新和副本结合策略的优化方案

为了提升闪存存储性能,闪存控制器具有多种功能,包括

地址映射、磨损均衡、垃圾收集、ECC 纠错等,图 7 给出了此工作在控制器中构建的框架。地址映射模块为系统保存逻辑块地址和物理块地址之间的对应关系。刷新框架可以从地址映射模块获取所有块的地址、生存期、有效页面数等信息;副本框架可以获取到读取的数据,方便之后对出错的数据进行恢复。目前,大多数闪存存储器都选择 LDPC 码作为 ECC 模块的纠错码。LDPC 码的解码流程为先用硬解码对出错的数据进行纠错,若纠错未成功则使用软解码,随着软解码的量化级别越来越高,纠错能力也越来越强。在闪存存储器的寿命早期,数据误码率较低,使用硬解码就可以纠错成功,但随着后期数据越来越多,误码率越来越高,出错的概率也越来越大,必须使用软解码才能纠错成功。但是软解码的解码复杂度较高,会造成大量的延迟,在一定程度上影响了闪存存储器的读性能。鉴于此,需要对解码流程进行优化,来优化闪存存储器的读性能,便于用户更好地使用。本文通过分析 LDPC 软解码量化级别的特点,将刷新技术和副本技术以最优的方式结合起来优化闪存存储器的读取性能。量化级别 3 和 5 是 LDPC 软解码延迟变化幅度大的两个级别,因此当数据到达量化级别 3 之后纠错成功时,ECC 模块将 LDPC 读级信息发送到刷新帧,提供 LDPC 延迟信息;当数据到达 5 之后纠错成功时,采用副本对其出错的数据进行恢复。考虑到副本需要对出错概率高的数据进行备份,会占用一定的空间,因此选择量化级别到达 5 之后使用副本,便于在减少空间的同时提升读取性能。

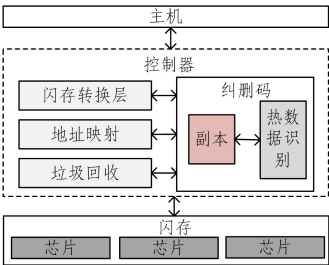


图 7 刷新和副本结合的方案框架

Fig. 7 Refresh and duplications combined scheme framework

本文提出的基于 LDPC 读延迟的刷新和副本结合的策略优化方案是根据一个块中 LDPC 读级别的数量来决定是否刷新和是否采用副本。如果一个块的 LDPC 读级别超过了设置的阈值,则需要对该块进行刷新和使用副本。基于 LDPC 读延迟的刷新和副本结合策略的优化方案利用了刷新和副本的优点,本文将此方案应用到 LDPC 软解码中来减少平均响应时间,从而提升闪存存储器的读性能。图 8 给出了该方案的流程。首先判断读入数据是热数据还是冷数据,根据数据的冷热程度来选择合适的方法。当数据为热数据时,首先需要记录第一读取时的方法,若为 LDPC 硬解码,则直接采用硬解码即可。若为软解码则需要判断量化级别,若量化级别大于或等于 3,则采用刷新来纠正出错的数据;若量化级别大于或等于 5,则采用副本来恢复数据。基于 LDPC 读延迟的刷新和副本结合策略的优化方案的详细流程如下,流程图如 8 所示。

步骤 1 统计数据的读取次数和访问频率,根据其访问

频率将其划分成热数据和冷数据。热数据为访问频率大于 1 的数据,反之即为冷数据。

步骤 2 统计第一次读取数据的解码方式,判断其是硬解码还是软解码。

步骤 3 统计数据进行软解码的量化级别,并进行记录。

步骤 4 根据数据的量化级别选择合适的方式,统计量化级别大于 3 的数据,采用刷新的方式进行纠错。

步骤 5 量化级别大于 5 之后,采用副本进行恢复。

步骤 6 测试数据是否纠错完成,如果完成则结束,否则继续进行纠错。

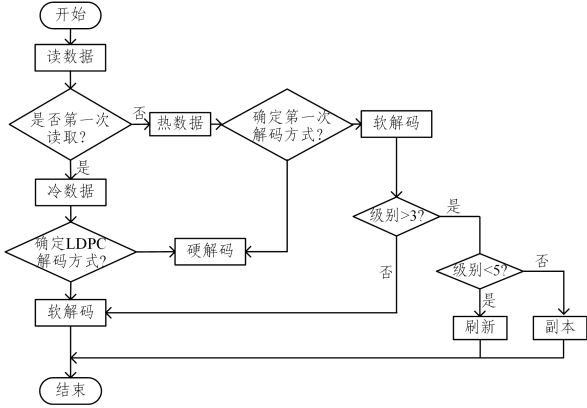


图 8 刷新和副本结合的方案流程图

Fig. 8 Flow chart of refresh and duplications combination scheme

上述内容从如何选择适当的量化级别开始,在确定量化级别后,将其与刷新和副本相结合,从固态驱动器的平均响应时间进行优化。本文提出的基于 LDPC 读延迟的副本和刷新策略结合的关键算法如算法 1 和算法 2 所示。

(1)如何将刷新和副本结合,如算法 1 所示。

算法 1 刷新和副本结合

输入:LDPC 软解码的量化级别

输出:合适的解码方式

- 1. 如果数据为热数据
- 2. 软解码量化级别大于或等于 3
- 3. 使用刷新
- 4. 软解码量化级别大于或等于 5
- 5. 使用副本
- 6. 否则
- 7. 执行软解码
- 8. 结束
- 9. 否则
- 10. 执行硬解码
- 11. 结束

(2)区分热数据和冷数据,如算法 2 所示。

算法 2 区分热数据和冷数据

输入:读取的数据

输出:数据的冷热程度

- 1. 如果数据频率大于 1
- 2. 此数据为热数据
- 3. 否则
- 4. 数据为冷数据
- 5. 结束

5 实验

本节将介绍闪存存储设备的实验设置和基本配置。通过在不同的负载环境下进行实验,我们得到了使用刷新和副本实验的量化级别,给出了基于 LDPC 读延迟的刷新和副本结合策略的优化方案的效果,分析了该方案对闪存设备平均响应时间和寿命的影响。

5.1 实验环境配置

为了评价该方案的有效性,本文使用带有闪存存储器扩展名的 Disksim 模拟了闪存系统。表 2 列出了实验所需要的参数,延迟来自于参数文件中的设置,LDPC 解码中的 7 个读取延迟根据表 1 的延迟分析计算。收集了 Financial1(F1), Financial2(F2), Websearch1(W1), Websearch2(W2), Websearch3(W3), Mds0, Hm0, Prxy0 和 Rsrch2 这 9 个实际工作负载^{1),2)}的统计数据和绩效结果。表 3 列出了工作负载的规格,包括读/写比率和持续时间。

表 2 闪存存储器参数配置
Table 2 Flash memory parameter configurations

参数配置	值
每个包中平面数	8
每个平面中块数	2048
每个块中页数	64
每个页面大小	8
芯片传输延迟	0.000025
页面读取延迟	0.025
页面写入延迟	0.200
块擦除延迟	1.5

表 3 负载读写数目统计 Table 3 Workloads read and write statistics			
负载	读	写	读比率/%
Financial1	1 235 633	4 099 354	23
Financial2	3 046 112	653 082	82
Websearch1	7 992 543	1 422 974	85
Websearch2	4 578 819	990	99
Websearch3	16 410 267	26 772	98
Mds0	143 973	1 067 061	12
Hm0	1 417 748	2 575 568	37
Prxy0	383 524	12 135 444	3
Rsrch2	136 364	71 223	65

5.2 适合刷新和副本的量化级别的确定

我们使用表中的数据做了一个实验来展示平均访问延迟。实验中闪存和工作负载的实验设置如表 2 和表 3 所列。我们在级别 1 到级别 7 的范围内测试了不同 LDPC 读取级别和不同读/写比率下闪存存储设备的平均访问延迟。随着 LDPC 读取电压的增加,器件的平均响应延迟逐渐增加。不同的负载表现出不同的特点,如图 9 所示。例如,从级别 1 到级别 3,Financial2 的平均响应延迟增加了 45%,Hm0 增加了 31%。从 3 级到 5 级,Financial2 的平均响应延迟增加了 46%,Hm0 增加了 33%。从 5 级到 7 级,Financial2 的平均响应延迟增加了 39%,Hm0 增加了 32%。虽然每个工作负载的平均响应时间增加不同,但结果表明,当 LDPC 的读取级别

为 3 级和 5 级时,设备的平均响应时间明显增加,表明产生了更多的 LDPC 解码延迟。降低 LDPC 解码延迟可以提升设备读写性能并减少平均响应时间。因此,我们选择了量化级别 3 和量化级别 5 为分界线,在此基础上,提出了一种基于 LDPC 读延迟的刷新和副本结合的策略方案。

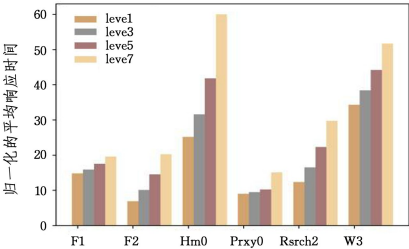


图 9 不同量化级别下各负载的平均响应时间
Fig. 9 Average response time of each load at different quantization levels

5.3 基于 LDPC 读延迟的刷新和副本方案实验结果分析

闪存存储器在寿命前期,数据出错的概率较小,使用简单的硬解码就可以纠错成功。随着后期读取数据量增多,出错的概率也变大,因此需要 LDPC 码软解码才能纠错成功。但软解码造成的延迟较多,为此我们分析了闪存存在早期和后期对平均响应时间的影响,如图 10 所示,决定在后期使用副本和刷新结合策略方案。在实验分析阶段,我们分析了基于 LDPC 读取延迟的刷新和副本的优化策略方案造成的平均响应时间的优化效率。

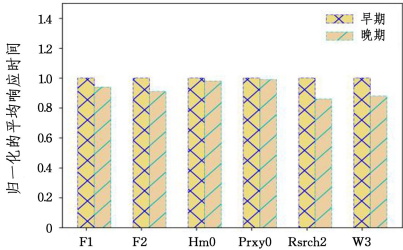


图 10 闪存存储器早期和晚期造成的平均响应时间对比
Fig. 10 Comparison of average response times caused by flash memory in early and late periods

本文实验评估了 3 种策略:

- (1)原始策略:在原始策略中,闪存存储器只依赖设备的垃圾收集模块、ECC 模块和磨损均衡模块来保证数据的可靠性。
- (2)传统刷新策略:在传统刷新策略中,只有达到 LDPC 的量化级别才会刷新数据。
- (3)刷新与副本相结合的策略优化方案:本文提出的方案根据热数据软解码的量化级别,选择使用刷新还是副本来纠正错误数据,以保证数据的稳定性,提升闪存存储器的读取性能。

与原始策略相比,刷新和副本结合的策略平均响应时间显著缩短,如图 11 所示,这进一步优化了闪存存储器的读取

¹⁾ STORAGE PERFORMANCE COUNCIL. SPC Trace File Format Specification. <http://traces.cs.umass.edu/index.php/Storage/Storage>
²⁾ “MSR cambridge traces”, <http://iotta.snia.org/tracetypes/3>, accessed:2016-11

性能。从图 11 可以看出,负载 Prxy0 和 Mds_0 下的平均响应时间减少最多,结合表 3 中负载的特点可知,读比率越小,提高闪存存储器读取性能的可靠性就越大。其中,Prxy0 负载下读取性能提高了 28%,Mds_0 负载下提高了 12%。

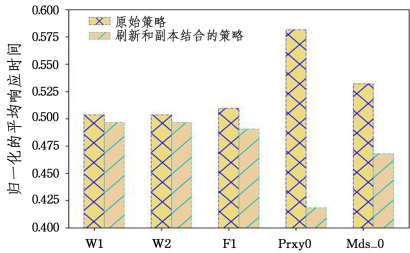


图 11 原始策略与刷新和副本结合的策略在不同负载下的平均响应时间对比

Fig. 11 Comparison of average response time of original strategy and the combined refresh and duplications strategy under different workloads

与传统刷新策略相比,刷新和副本结合的策略延长了闪存存储器的寿命,如图 12 所示,这更便于闪存存储器长久地使用。从图 12 中可以看出,负载 W1 和 W2 下的寿命延长情况较好,结合表 3 负载的读写比率可知,读比率越大,寿命的延长情况越好。

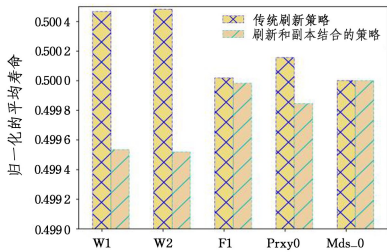


图 12 传统刷新策略与刷新和副本结合的策略在不同负载下的寿命对比

Fig. 12 Lifetime comparison of traditional refresh strategy and refresh and duplication combination strategy under different workloads

结束语 本文提出了一种基于 LDPC 读取延迟的刷新和副本结合策略的优化方案,旨在在闪存存储器的后期降低其读取延迟。根据数据的冷热程度,对于热数据,该方法使用刷新来纠正软解码量化级别大于或等于 3 的错误数据,并使用副本来恢复大于或等于 5 的错误数据。通过不同负载下的测试,证明了方案的可行性,但是使用副本会浪费空间,在未来的进一步研究中,我们将继续考虑是否有其他混合方案可以从 LDPC 软解码方面提升闪存存储器的读取性能。

参考文献

[1] CERNEA R A, PHAM L, MOOGAT F, et al. A 34 MB/s MLC write throughput 16 Gb NAND with all bit line architecture on 56 nm technology[J]. IEEE Journal of Solid-State Circuits, 2008, 44(1): 186-194.

[2] TOKUTOMI T, TANAKAMARU S, IWASAKI T O, et al. Advanced error prediction LDPC for high-speed reliable TLC nand-based SSDs[C]// 2014 IEEE 6th International Memory Work-

shop(IMW). IEEE, 2014: 1-4.

[3] SUKKWANG P, JAEKYUN M. Characterization of Inter-Cell Interference in 3D {NAND} Flash Memory[J]. IEEE Trans. Circuits Syst. {I} Regul. Pap., 2021, 68(3): 1183-1192.

[4] RIZVI S S, CHUNG T S. Flash SSD vs HDD: High performance oriented modern embedded and multimedia storage systems [C]// 2010 2nd International Conference on Computer Engineering and Technology. IEEE, 2010: 297-299.

[5] LEE S, LEE J, PARK I, et al. 7.5 A 128Gb 2b/cell NAND flash memory in 14nm technology with tPROG=640μs and 800MB/s I/O rate[C]// 2016 IEEE International Solid-State Circuits Conference (ISSCC). IEEE, 2016: 138-139.

[6] JEONG W, IM J, KIM D H, et al. A 128 Gb 3b/cell V-NAND flash memory with 1 Gb/s I/O rate[J]. IEEE Journal of Solid-State Circuits, 2015, 51(1): 204-212.

[7] LI S, ZHANG T. Improving Multi-Level NAND Flash Memory Storage Reliability Using Concatenated BCH-TCM Coding[J]. IEEE Transactions on Very Large Scale Integration Systems, 2010, 18(10): 1412-1420.

[8] REN J, DING X, XIN X N, et al. An NB-LDPC decoder Algorithm combined using channel information for Storage Application[C]// 2020 IEEE 5th International Conference on Integrated Circuits and Microsystems(ICICM). IEEE, 2020: 306-309.

[9] KANG D, JEONG W, KIM C, et al. 256 Gb 3 b/cell V-NAND flash memory with 48 stacked WL layers[J]. IEEE Journal of Solid-State Circuits, 2016, 52(1): 210-217.

[10] LI Q, SHI L, XUE C J, et al. Improving LDPC performance via asymmetric sensing level placement on flash memory[C]// 2017 22nd Asia and South Pacific Design Automation Conference (ASP-DAC). IEEE, 2017: 560-565.

[11] GALLAGER R. Low-density parity-check codes[J]. IRE Transactions on Information Theory, 1962, 8(1): 21-28.

[12] SHOKROLLAHI A. An Introduction to Low - Density Parity-Check Codes[C]// TACSCI 2000. Berlin: Springer. 2002: 175-197.

[13] CAI Y, YALCIN G, MUTLU O, et al. Flash correct-and-refresh: Retention-aware error management for increased flash memory lifetime[C]// 2012 IEEE 30th International Conference on Computer Design (ICCD). IEEE, 2012: 94-101.

[14] PAN Y, DONG G, WU Q, et al. Quasi-nonvolatile SSD: Trading flash memory nonvolatility to improve storage system performance for enterprise applications[C]// IEEE International Symposium on High-Performance Comp Architecture. IEEE, 2012: 1-10.

[15] DU Y, LI Q, SHI L, et al. Reducing LDPC soft sensing latency by lightweight data refresh for flash read performance improvement[C]// 2017 54th ACM/EDAC/IEEE Design Automation Conference(DAC). IEEE, 2017: 1-6.

[16] LV Y, SHI L, LI Q, et al. Optimizing Tail Latency of LDPC based Flash Memory Storage Systems Via Smart Refresh[C]// 2019 IEEE International Conference on Networking, Architecture and Storage(NAS). IEEE, 2019: 1-8.

[17] DI Y, SHI L, GAO C, et al. Minimizing retention induced refresh through exploiting process variation of flash memory[J]. IEEE

Transactions on Computers,2018,68(1):83-98.

[18] CAI Y,LUO Y,GHOSE S,et al. Read disturb errors in MLC NAND flash memory;Characterization,mitigation,and recovery [C]//2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. IEEE,2015:438-449.

[19] LI P,ZHANG Y,YIN D,et al. An Efficient Refresh Strategy of Flash Memory via High Delay Blocks in LDPC[C]//2021 6th International Conference on Integrated Circuits and Microsystems(ICICM). IEEE,2021:299-304.

[20] LI P,ZHANG Y,YIN D,et al. A High Precision Refresh Method to Improve The Performance of Flash Storage Devices [C]//2021 20th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES). IEEE,2021:214-217.

[21] GHEMAWAT S,GOBIOFF H,LEUNG S T. The Google file system[C]//Proceedings of the nineteenth ACM Symposium on Operating Systems Principles. 2003:29-43.

[22] DECANDIA G,HASTORUN D,JAMPANI M,et al. Dynamo: Amazon's highly available key-value store[J]. ACM SIGOPS Operating Systems Review,2007,41(6):205-220.

[23] KISTLER J J,SATYANARAYANAN M. Disconnected operation in the Coda file system[J]. ACM Transactions on Computer Systems(TOCS),1992,10(1):3-25.

[24] WEIL S A,BRANDT S A,MILLER E L,et al. Ceph: A scalable,high-performance distributed file system[C]// Proceedings of the 7th Symposium on Operating Systems Design and Implementation. 2006:307-320.

[25] BONVIN N,PAPAIOANNOU T G,ABERER K. Dynamic cost-efficient replication in data clouds[C]// Proceedings of the 1st Workshop on Automated Control for Datacenters and Clouds. 2009:49-56.

[26] MUTHITACHAROEN A,CHEN B,MAZIERES D. A low-bandwidth network file system[C]// Proceedings of the eighteenth ACM Symposium on Operating Systems Principles. 2001:174-187.

[27] SAITO Y,KARAMANOLIS C,KARLSSON M,et al. Taming aggressive replication in the Pangaea wide-area file system[C]// Proceedings of the 5th Symposium on Operating Systems Design and Implementation. USENIX Association,2002:15 -30.

[28] CALDER B,WANG J,OGUS A,et al. Windows azure storage:a highly available cloud storage service with strong consistency [C]// Proceedings of the Twenty-Third ACM Symposium on Operating Systems Principles. 2011:143-157.

[29] BINDEL D,CHEN Y,EATON P,et al. Oceanstore:An extremely wide-area storage system[C]// Proceedings of the 9th International Conference on Architectural Support for Programming Languages and Operating Systems. 2000:190-201.



ZHANG Yaofang, born in 1995, post-graduate candidate,is a member of China Computer Federation. Her main research interest is network storage.



XIE Ping, born in 1979, Ph.D, professor, is a member of China Computer Federation. His main research interests include computer architecture,and mass storage system.

(责任编辑:何杨)