

一种用于减少闪存磨损的数据塑型方法

张文辉¹ 曹强¹

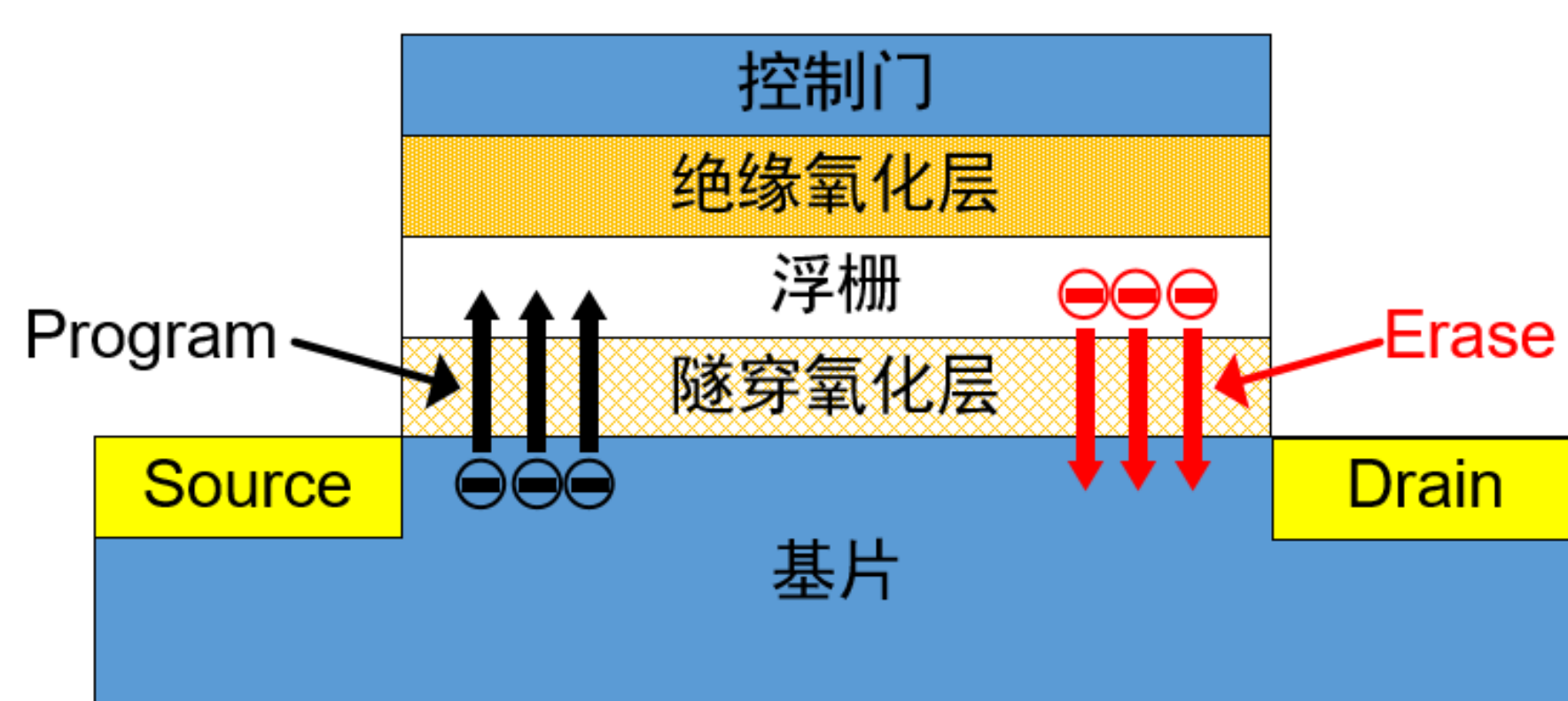
¹武汉光电国家实验室， 华中科技大学

摘要

在现今的计算机系统中，应用对存储系统的性能要求越来越高。因此，存储系统愈来愈趋向于使用读写性能远超磁盘的闪存作为存储介质。但是闪存单元每次擦写都会产生磨损，从而有使用寿命的限制。除此以外，随着闪存单元擦写次数的增加，比特错误率也会随之增高。所以，减慢闪存磨损是闪存存储系统的一个重要研究方向。已有研究表明，在闪存中，写入“0”对闪存单元造成的磨损要远高于写入“1”对闪存单元造成的磨损。我们利用闪存单元的这个特性，提出一种基于比特翻转的数据塑型方法，称为BFbDS (Bit-Flip based Data Shaping)。BFbDS通过改变数据的“0”、“1”比例，减少数据写入闪存时对闪存单元造成的磨损，从而达到延长闪存使用寿命的目的。我们在实验中研究了包括文字、图片、音频等6类数据的“0”、“1”比例，并且使用BFbDS对这些数据进行塑型。实验表明，使用BFbDS最多可以减少数据中31%的“0”，从而大幅度减少写入这些数据时对闪存单元造成的磨损。

1. 闪存特性

对闪存单元的擦除操作和写入操作都会导致电子穿过单元内的隧穿氧化层，而这个过程里电子会对隧穿氧化层造成伤害，从而对闪存单元造成磨损。有研究表明，对于一个SLC闪存单元，写入“0”对闪存单元造成的磨损要远高于写入“1”对闪存单元造成的磨损。这是因为“1”对应于闪存单元的擦除态，也就是一个闪存单元被擦除后表现的状态，写入“1”不需要电子穿过隧穿氧化层；相对地，“0”对应于闪存单元的编程态，写入“0”是需要向闪存单元的浮栅注入电子的。



闪存单元示意图

2. SLC磨损模型

对于SLC的一个浮栅单元来说，以 W_E 表示写入擦除态“1”时对浮栅单元造成的磨损，以 W_P 表示写入编程态“0”时对浮栅单元造成的磨损，此外，使用 W_T 表示这个浮栅单元总共可承受的磨损，即当对该单元的累积磨损超过 W_T 时，该浮栅单元损坏。设该浮栅单元损坏前共被擦写 N 次，且写入“0”的概率为 p ，相对地，写入“1”的概率为 $1-p$ 。已有研究表明对SLC单元写入“1”造成的磨损要远小于写入“0”，即 $W_P \gg W_E$ 。SLC磨损模型表示如下：

$$N = \frac{W_T}{p * W_P}$$

即SLC单元的可擦除次数与写入“0”的概率成反比。

3. 比特翻转塑型方法——BFbDS

BFbDS的基本思想是，通过计算一个数据段中“0”和“1”的占比，如果数据段中“0”的占比高于“1”的占比，则对该段数据实施比特翻转后再存储，也就是数据段中的“0”翻转为“1”，而“1”翻转为“0”；相反，如果数据段中“0”的占比低于“1”的占比，则该段数据无需进行翻转直接存储。从BFbDS的基本思想可以知道，BFbDS的塑型效果具有很高的稳定性，不存在数据塑型后“1”的比例比塑型前“1”的比例更低的情况。但是，BFbDS的塑型效果很大程度上依赖于原数据段中“0”、“1”比例的差异，当原数据中“0”、“1”比例差异大时塑型效果好，相反，则塑型效果较差。为提高BFbDS的塑型效果，可以缩小实施比特翻转的单元。举例来说，当对一段长度为2KB的数据进行塑型时，可以把数据划分为多个子数据段，例如每32字节为一个子段，这样，BFbDS的塑型效果不再完全取决于数据段整体的“0”、“1”比例，而是与每一个子段中“0”、“1”的比例相关，这样即使数据整体的“0”、“1”比例差异较大，BFbDS也可能获得不错的塑型效果。实际上，使用较小的比特翻转单元一定可以获得不差于使用较大翻转单元的塑型效果。但是每个比特翻转单元在存储时需要记录一个比特以表明它在存储前是否被翻转，因此，使用的比特翻转单元越小，需要额外存储的比特翻转信息也就越多，因此不能无限制地缩小使用的比特翻转单元。

4. 实验效果展示

在实验中，我们使用了包括文本、图文混合、数据表格和图片等7中类型、13种不同的文件格式共112个数据样本，并且对这些样本实施BFbDS塑型，记录塑型前后“0”的占比的变化。我们的实验结果表明，BFbDS对数据表格、图文混合以及音频等类型数据的塑型效果较好，以8字节为比特翻转单元时最多可以在塑型后减少数据中30%的“0”。即使对于数据中“0”、“1”比例较为均衡的压缩包类型数据，使用BFbDS也可以减少数据中10%的“0”。

