

基于二维汉明码结构的闪存纠错存储系统

余辉龙¹, 张 健¹, 李 清², 覃 翠¹, 赵 静¹, 花 涛¹

(1 南京工程学院, 江苏 南京 211167, 2 中北大学, 山西 太原 030051)

摘 要: 针对闪存存储单元会出现多位错误, 提出了 NAND 型闪存存储器的二维汉明码算法, 从行和列二维结构上对任意 2^j bit 字长的存储数据, 用统一的公式化算法进行汉明编码, 通过计算纠错偶对能够纠错多位数据误码的情况, 译码时先纠错, 再校验, 保证存储系统数据的完整性, 解决了 MLC 结构闪存的纠错问题. 在此基础上分析了不能纠错的图样模式. 该算法核心结构上只有异或操作, 易于实现, 适用于 NAND 型闪存存储器. 实验表明, 闪存系统中有效数据二维结构为 $1\ 024 \times 128$ 时, 采用二维汉明码结构, 编码效率为 11%, 实际编码效率为 12.6%. 其编码速度和译码速度均达到了 102 Gbps, 适用于高速存储系统.

关键词: 二维汉明码; 纠错编码; NAND 闪存; 纠错偶对

中图分类号: TP333.5

文献标识码: A

文章编号: 1000-7180(2018)03-0001-03

Flash Error Correcting Storage System Using Two Dimensional Structure Hamming Code

YU Hui-long¹, ZHANG Jian¹, LI Qing², QIN Cui¹, ZHAO Jing¹, HUA Tao¹

(1 Nanjing Institute of technology, Nanjing 211167, China;
2 North University of China, Taiyuan 030051, China)

Abstract: NAND flash unit causes multi-bit data error, the two-dimensional Hamming code algorithm for NAND flash memory is proposed. Any two-dimensional length of 2^j bit storage data From the row and column view, the uniform Hamming encoding algorithm formula is provided. by calculating error-correcting-pairs, multiple bit data error can be corrected. The error of the MLC structure and the integrity of data storage system is solved by first decoding, and then check decoding. On this basis, the error no-corrected pattern is analyzed. XOR operation which is the core structure of the algorithm is easy to implement and suitable for NAND flash memory. Experimental results show that the effective data in two-dimensional structure $1\ 024 * 128$. Coding efficiency is 11%, actual 12.6%. Its encoding speed and decoding speed are up to 102 Gbps, suitable for high-speed storage system.

Key words: two-dimensional hamming code; error correcting code; NAND flash; error-correcting-pairs

1 引言

以 NAND 型闪存为存储介质的容量数据存储获得了越来越广泛的应用, 闪存由于在自身栅结构工艺问题, 闪存不仅有出厂无效块, 在使用过程中, 无效块会随机增长, 使得存储的数据产生软错误和硬错误^[1]. 而目前闪存存在 SLC(single level cell)和 MLC(Multi level cell). SLC 每个基本单元

存储 1 bit 数据, MLC 每个基本单元存储多位数据, 目前一般为 2 bit 和 3 bit. NAND 型闪存以页为基本存储组织单元. 然而, NAND 型闪存按页存储数据, 一页有效数据为 $2^N + K$ 字节, 需建立一个纠错码结构, 其有效信息数据量接近于 2^N 字节, 冗余编码数据小于 K 字节, 并最大化纠错能力^[2].

BCH 码和 RS 码生成多项式, 其编译码过程相对复杂, 不适用于信噪比较大, 误码概率小, 但是

实时性要求很高的场合^[3-5]. 文献[6]提出了一种汉明码为 $(2^r + 2r, 2r, 4)$ 码, 与闪存的页存储结构相匹配, 能够纠错一位数据, 检错两位数据, 适用于高速数据的存储和传输. 文献[6]仅能解决大信噪比下 SLC 闪存单元的误码问题, 无法解决 MLC 的误码问题. 因此, 需要在此基础上加以改进.

2 闪存系统存储结构

现在闪存存储系统均会采用并行总线和流水线结构, 如图 1 所示, 设系统一共有 L 级流水线, 并行总线数为 P , 即位宽为 $P \times 8$ bit, 而闪存具有页结构, 其页格式一般为 $2^N + K$ 字节. 目前 NAND 型闪存中 K 与 N 的关系如下所示:

$$K = 2^{N-5} \quad (1)$$

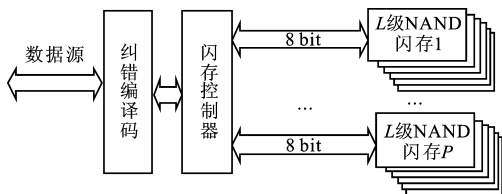
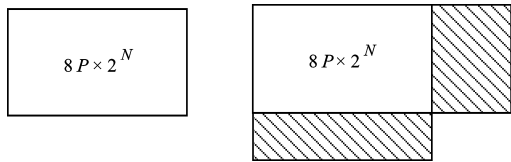


图 1 闪存阵列结构

文献[6]提出了一种汉明码为 $(2^r + 2r, 2r, 4)$, 其信息位为 2^r 比特, 冗余纠错位为 $2r$ 比特, 与闪存每页有 2^N 字节, 有 K 个冗余备份区的结构匹配. 但是该结构的汉明码仅能纠错 1 bit 的错误, 目前流行的 MLC 闪存单元, 每个单元可以存储 $2 \sim 3$ 个比特的数据, 如果出现误码, 一般会出现 $2 \sim 3$ 个比特的数据错误, 无法实现纠错, 而目前闪存阵列在单片闪存的基础上, 在位宽和流水线级上进行了延伸, 在两个维度上利用文献[6]的算法, 实现 MLC 闪存阵列的纠错与检错.



(a) 原始数据数据块 (b) 添加冗余纠错数据后数据块

图 2 闪存二维数据块

如图 2 所示, 二维原始数据块为一个 $8P \times 2^N$ 的数据, 在两个维度方向上对其添加纠错码, 按照文献[6]的方法, 需要添加三个数据块, 分别为 $(2 \log_2 8P) \times 2^N, 8P \times (2N)$.

3 算法描述

系统并行总线数为 P , 即位宽为 $P \times 8$ bit, 闪存

页结构为 2^N , 不失一般性, 设系统位宽 $8P = 2^G$. 从每一个方向上计算冗余纠错码. 设 $J = G$ 或 N , 则三维数据源的数据为 2^J .

设 2^J bit 数据源为向量 $\vec{a} = (a_0, a_1, \dots, a_{2^J-1})$, 通过异或计算得到 $2J$ 个纠错码, 记为向量 $\vec{u} = (u_0, u_1, \dots, u_{2^J-1})$. 计算过程如下所示.

$$u_0 = a_0 \oplus a_2 \oplus a_4 \oplus \dots \oplus a_{(2^J-2)}$$

$$u_1 = a_1 \oplus a_3 \oplus a_5 \oplus \dots \oplus a_{(2^J-1)}$$

$$u_2 = (a_0 \oplus a_1) \oplus (a_4 \oplus a_5) \oplus \dots \oplus (a_{(2^J-4)} \oplus a_{(2^J-3)})$$

$$u_3 = (a_2 \oplus a_3) \oplus (a_6 \oplus a_7) \oplus \dots \oplus (a_{(2^J-2)} \oplus a_{(2^J-1)})$$

...

$$u_{2^J-2} = a_0 \oplus a_1 \oplus \dots \oplus a_{(2^{J-1}-1)}$$

$$u_{2^J-1} = a_{2^{J-1}} \oplus a_{2^{J-1}+1} \oplus \dots \oplus a_{2^J-1}$$

上述符号“ $\oplus$ ”代表异或操作.

通过如上方法将 J 替换为 G 和 N 得到 $2G$ 和 $2N$ 个冗余的编码数据. 并将改数据存储到闪存的冗余区或备份闪存中.

(2) 在读取闪存中的数据时, 读取的同时按照步骤 1) 对上述数据进行编码, 得到纠错码字 $\vec{w} = (w_0, w_1, \dots, w_{2^J-1})$. 此时计算 $\vec{Q} = (S_0, S_1, \dots, S_{2^J-1}) = \vec{u} \oplus \vec{w}$.

定义纠错偶对 $\tilde{Q}_i = (S_{2i}, S_{(2i+1)})$, $\vec{Q} = (Q_0, \dots, Q_{(J-1)})$

若 $\tilde{Q}_i = (S_{2i}, S_{(2i+1)}) = (1, 0)$ 时, 令 $Q_i = 0$

若 $\tilde{Q}_i = (S_{2i}, S_{(2i+1)}) = (0, 1)$ 时, 令 $Q_i = 1$

令 $Q = Q_0 + Q_1 2 + \dots + Q_k 2^k + \dots + Q_{(J-1)} 2^{(J-1)}$

为纠错偶向量 $\vec{Q} = (Q_0, \dots, Q_{(J-1)})$ 对应的位置码.

(3) 向量 $\vec{u}$ 在 J 取 G 和 N 对应三维纠错码字分别为 $\vec{u}^G$ 和 $\vec{u}^N$, 向量 $\vec{w}$ 在 J 取 G 和 N 对应二维纠错码字分别为 $\vec{w}^G$ 和 $\vec{w}^N$. 对 2 组向量做如下计算:

$$\vec{d}^G = (d_0, d_1, \dots, d_{2G-1}) = \vec{u}^G \oplus \vec{w}^G$$

$$\vec{e}^N = (e_0, e_1, \dots, e_{2N-1}) = \vec{u}^N \oplus \vec{w}^N$$

对二维码字进行上述计算最终得到 2 个二维纠错码字块分别为 $\{g_{ij}, 0 \leq i \leq (2G-1), 0 \leq j \leq (2N-1)\}$ 和 $\{n_{ij}, 0 \leq i \leq (2G-1), 0 \leq j \leq (2N-1)\}$

此基础上进行误码率判断.

(4) 错误判断及纠错.

① 设二维码字中闪存某一页仅有不超过 1 比特元素发生错误. 则在该页出现纠错偶对现象, 计算 e^N

$= \vec{u}^N \oplus \vec{w}^N$,并计算出错误位置,然后校验即可实现纠错。

② 设二维码字中闪存的某一页有2比特及以上发生错误,则通过 $\vec{d}^G = \vec{u}^G \oplus \vec{w}^G$,计算错误位置。如图3所示,黑色代表错误信息位,白色代表出现错误,闪存一页出现多个错误(一行代表一页),但是相应的列最多只有1位出错,此时,分别计算 $e^N = \vec{u}^N \oplus \vec{w}^N$ 和 $\vec{d}^G = \vec{u}^G \oplus \vec{w}^G$,将能够纠错的部分纠正,然后再进行二次纠错,可纠正下图中间圈中的错误,然后校验即可实现纠错。

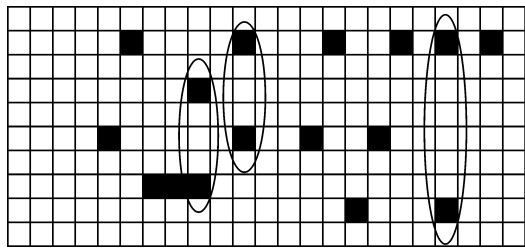
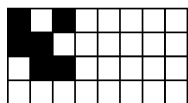
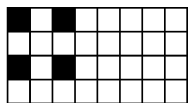


图3 多位纠错

③ 若完成初步纠错,进行二次纠错时,若出现在任何方向超过2个错误的误码,如图4所示的纠错极限,其中黑色代表错误信息位,白色代表出现错误,则不可纠错。



(a)6 比特两两错误



(b)4 比特两两错误

图4 不可纠错

4 结束语

实验系统采用三星 NAND 公司的 K9KAG08U1M 芯片,该闪存芯片每页有 4 096 字节,冗余区为 128 字节,每级流水线由 16 片闪存并列构成,即 P 为 8。有效数据并行总线数为 128。按照最大化纠错原则取 $N=10$,将每页拆分为 4 个 1 024 字节的有效数据,对应 20 字节的冗余纠错数据,此时剩余 48 字节的数据。并行总线数为 P 为 8,此时有效数据位宽为 128 比特,此时 G 为 7。因此需要扩充 2 片 K9KAG08U1M 芯片存放纠错冗余数据。

系统编码效率为:

$$\eta = \frac{4 \times 2^N \times 2G + 4 \times 2N \times 2^G}{4 \times 2^G \times 2^N + 2^N \times 2G + 4 \times 2N \times 2^G}$$

其中 G 为 7, N 为 10,此时有效数据二维结构为 $1\,024 \times 128$,计算得出编码效率为 11%,若考虑到实际使用两片闪存存储冗余数据,则实际编码效

率为 12.6%。

在实验系统中,仅施行编译码运算,由于系统运算仅有异或运算,运算简单,其编码速度和译码速度均达到了 102 Gbps,不影响系统的存储速度。

在大容量闪存存储系统架构的基础上,针对 MLC 每个基本单元存储多位数据,如果出错将会出现多比特误码,提出了 NAND 型闪存存储器的二维汉明码算法,行和列二维结构上对存储的数据进行编码,能够纠错多位数据误码的情况。并分析了不能纠错的图样,给出了较为详尽的算法描述。

参考文献:

- [1] 余辉龙,何昕,魏仲慧,等. NAND 型闪存大容量图像存储器无效块管理[J]. 微电子学与计算机, 2010, 27(2):1-4
- [2] 李锴,马航飞,张美璟,等. 基于闪存坏块表的大粒度文件隐藏[J]. 微电子学与计算机, 2015, 32(5): 161-164.
- [3] Peng X, Zhang W, Ji W, et al. Reduced-complexity multiplicity assignment algorithm and architecture for low-complexity chase decoder of reed-solomon codes [J]. IEEE Communications Letters, 2015, 19(11): 1865-1868.
- [4] 孙健,张辉,王宇飞,等. 一种基于 RS(24,20)的编译码器设计[J]. 微电子学与计算机, 2016, 33(12): 75-79.
- [5] Ding C, Fan C, Zhou Z. The dimension and minimum distance of two classes of primitive BCH codes [J]. Finite Fields and Their Applications, 2017 (45): 237-263.
- [6] 余辉龙,张健,覃翠,等. 一种指数字长的通用汉明码纠错算法[J]. 微电子学与计算机, 2015, 32(7): 59-61.

作者简介:

余辉龙 男,(1982-),博士,讲师.研究方向为光电信号处理、纠错编码. E-mail: yh@njit.edu.cn.

张健 男,(1977-),博士,副教授.研究方向为光电子技术.

李清 男,(1966-),博士,副教授.研究方向为微流体驱动控制、机电系统.

覃翠 女,(1984-),博士,讲师.研究方向为光电子技术.

赵静 女,(1986-),博士,讲师.研究方向为光电检测、微光夜视技术.

花涛 男,(1985-),博士,副教授.研究方向为超导电子学、太赫兹科学.