

Chapter6 层次存储结构_1

第一讲 存储器概述

程序及指令的执行过程

回顾：程序及指令的执行过程

- ° 在内存存放的指令实际上是机器代码（0/1序列）

08048394 <add>:

1	8048394:	55	push	%ebp
2	8048395:	89 e5	mov	%esp, %ebp
3	8048397:	8b 45 0c	mov	0xc(%ebp), %eax
4	804839a:	03 45 08	add	0x8(%ebp), %eax
5	804839d:	5d	pop	%ebp
6	804839e:	c3	ret	

栈是主存中的一个区域！

- ° 对于add函数的执行，以下情况下都需要访存：

- ✓ 每条指令都需从主存单元取到CPU执行 **取指**
- ✓ PUSH指令需把寄存器内容压入栈中 **存数** POP指令则相反 **取数**
- ✓ 第3条mov指令需从主存中取数后送到寄存器 **取数**
- ✓ 第4条add指令需从主存取操作数到ALU中进行运算 **取数**
- ✓ ret指令需从栈中取出返回地址，以能正确回到调用程序执行 **取数**

访存是指令执行过程中一个非常重要的环节！取指、取数、存数

基本术语

记忆单元（存储基元 / 存储元 / 位元）（Cell）

具有两种稳态的能够表示二进制数码0和1的物理器件

存储单元 / 编址单位（Addressing Unit）

具有相同地址的位构成一个存储单元，也称为一个编址单位

存储体 / 存储矩阵 / 存储阵列（Bank）**

所有存储单元构成一个存储阵列

编址方式（Addressing Mode）

字节编址、按字编址

存储器地址寄存器（Memory Address Register - MAR）

用于存放主存单元地址的寄存器

存储器数据寄存器（Memory Data Register-MDR (或MBR)）

用于存放主存单元中的数据的寄存器

存储器分类

（1）按工作性质/存取方式分类

- 随机存取存储器 Random Access Memory (RAM)

每个单元读写时间一样，且与各单元所在位置无关。如：**内存**。

(注：原意主要强调地址译码时间相同。现在的DRAM芯片采用行缓冲，因而可能因为位置不同而使访问时间有所差别。)

- **顺序存取存储器 Sequential Access Memory (SAM)**

数据按顺序从存储载体的始端读出或写入，因而存取时间的长短与信息所在位置有关。例如：**磁带**。

- **直接存取存储器 Direct Access Memory(DAM)**

直接定位到读写数据块，在读写数据块时按顺序进行。如**磁盘**。

- 相联存储器 Associate Memory (AM)

Content Addressed Memory (CAM)

按内容检索到存储位置进行读写。例如：快表。

(2) 按存储介质分类

半导体存储器：双极型，静态MOS型，动态MOS型

磁表面存储器：磁盘 (Disk)、磁带 (Tape)

光存储器：CD, CD-ROM, DVD

(3) 按信息的可更改性分类

读写存储器 (Read / Write Memory): 可读可写

只读存储器 (Read Only Memory): 只能读不能写

(4) 按断电后信息的可保存性分类

非易失 (不挥发) 性存储器 (Nonvolatile Memory)

信息可一直保留，不需电源维持。

(如：ROM、磁表面存储器、光存储器等)

易失 (挥发) 性存储器 (Volatile Memory)

电源关闭时信息自动丢失。(如：RAM、Cache等)

(5) 按功能/容量/速度/所在位置分类

- **寄存器(Register)**

封装在CPU内，用于存放当前正在执行的指令和使用的数据

用触发器实现，速度快，容量小 (几~几十个)

- **高速缓存(Cache)**

位于CPU内部或附近，用来存放当前要执行的局部程序段和数据

用SRAM实现，速度可与CPU匹配，容量小 (几MB)

- **内存存储器MM (主存储器Main (Primary) Memory)**

位于CPU之外，用来存放已被启动的程序及所用的数据

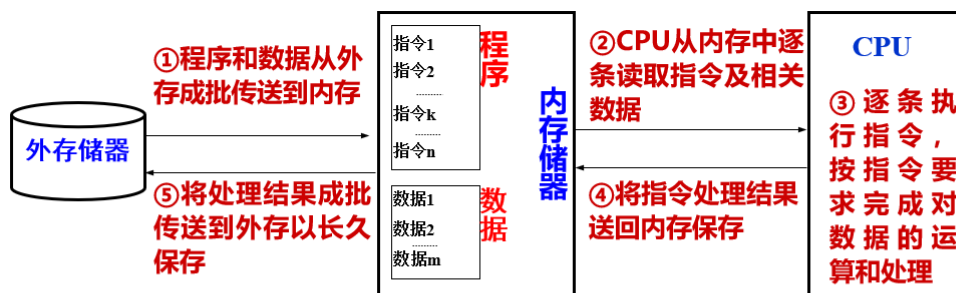
用DRAM实现，速度较快，容量较大 (几GB)

- **外存储器AM (辅助存储器Auxiliary / Secondary Storage)**

位于主机之外，用来存放暂不运行的程序、数据或存档文件

用磁表面或光存储器实现，容量大而速度慢

内存与外存的关系及比较



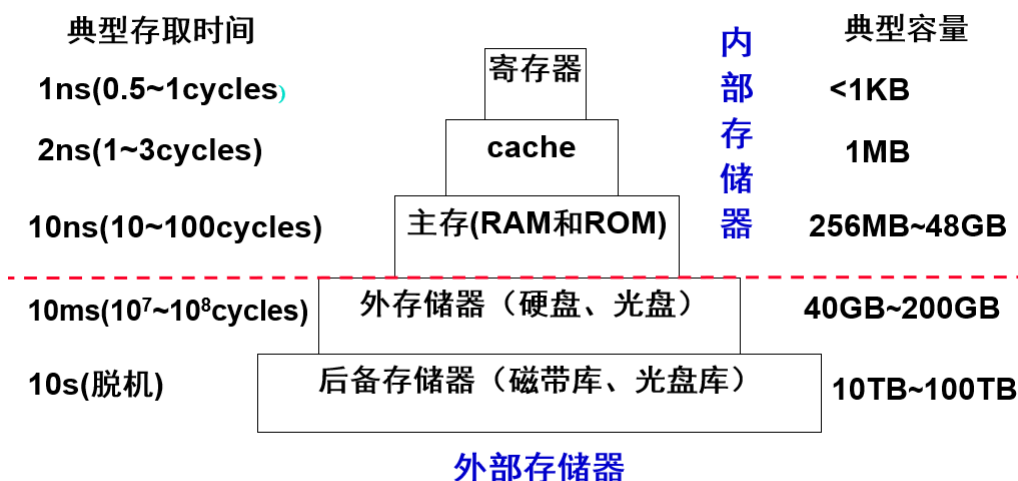
✓ 外存储器（简称外存或辅存）

- 存取速度慢
- 成本低、容量很大
- 不与CPU直接连接，先传送到内存，然后才能被CPU使用。
- 属于**非易失性**存储器，用于长久存放系统中几乎所有的信息

✓ 内存储器（简称内存或主存）

- 存取速度快
- 成本高、容量相对较小
- 直接与CPU连接，CPU对内存中可直接进行读、写操作
- 属于**易失性**存储器(volatile)，用于临时存放正在运行的程序和数据

存储器的层次结构与主存的结构



列出的时间和容量会随时间变化，但数量级相对关系不变。

PC机主存储器的物理结构

- 由若干内存条组成
- 内存条的组成：

把若干片**DRAM**芯片焊装在一小条印制电路板上制成

- 内存条必须插在主板上的内存条插槽中才能使用



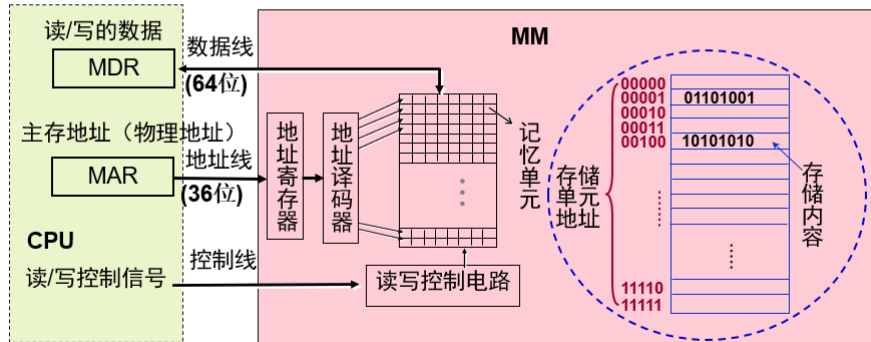
主存的结构

问题：主存中存放的是什么信息？CPU何时会访问主存？

指令及其数据！CPU执行指令时需要取指令、取数据、存数据！

问题：地址译码器的输入是什么？输出是什么？可寻址范围多少？

输入是地址，输出是地址驱动信号（只有一根地址驱动线被选中）。
可寻址范围为 $0 \sim 2^{36}-1$ ，即主存地址空间为64GB（按字节编址时）。



主存地址空间大小不等于主存容量（实际安装的主存大小）！

若是字节编址，则每次最多可读/写8个单元，给出的是首(最小)地址。

主存地址空间：可寻址的主存范围

主存容量：实际安装的主存大小

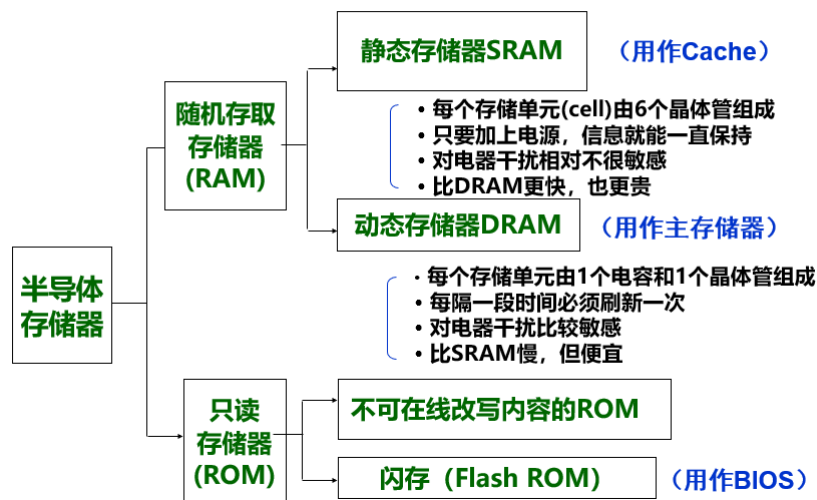
主存的主要性能指标

性能指标：

- 按字节连续编址，每个存储单元为1个字节（8个二进制位）
- 存储容量：所包含的存储单元的总数（单位：MB或GB）
- 存取时间 T_A ：从CPU送出内存单元的地址码开始，到主存读出数据并送到CPU（或者是把CPU数据写入主存）所需要的时间（单位：ns, $1 \text{ ns} = 10^{-9} \text{ s}$ ），分读取时间和写入时间
- 存储周期 T_{MC} ：连续两次访问存储器所需的最小时间间隔，它应等于存取时间加上下一次存取开始前所要求的附加时间，因此， T_{MC} 比 T_A 大（因为存储器由于读出放大器、驱动电路等都有一段稳定恢复时间，所以读出后不能立即进行下一次访问。）
(就像一趟火车运行时间和发车周期是两个不同概念一样。)

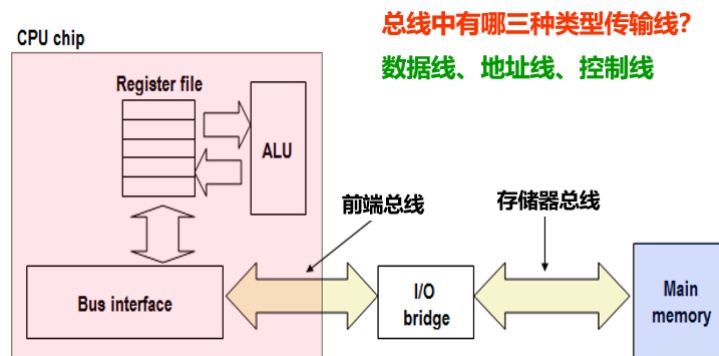
内存储器的分类及应用

内存由半导体存储器芯片组成，芯片有多种类型：



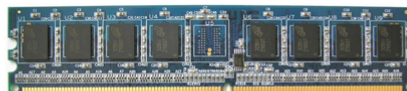
第二讲：主存与CPU的连接及其读写操作

主存与CPU的连接

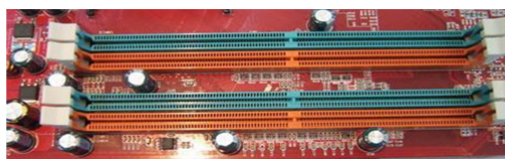


PC机主存储器的物理结构

由若干内存条组成
内存条的组成:



把若干片DRAM芯片焊装在一小条印制电路板上制成
内存条必须插在主板上的内存条插槽中才能使用

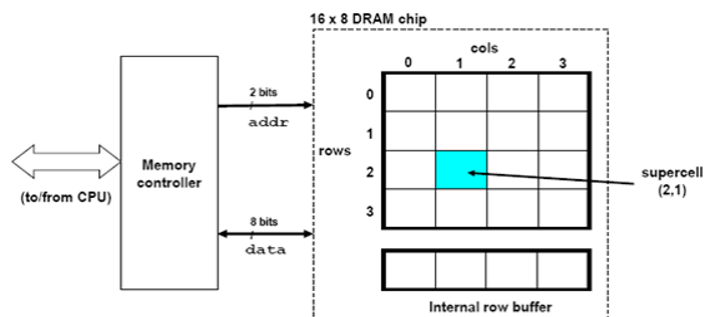


目前流行的是DDR2、DDR3内存条:

- 采用双列直插式, 其触点分布在内存条的两面
- DDR条有184个引脚, DDR2有240个引脚
- PC机主板中一般都配备有2个或4个DIMM插槽

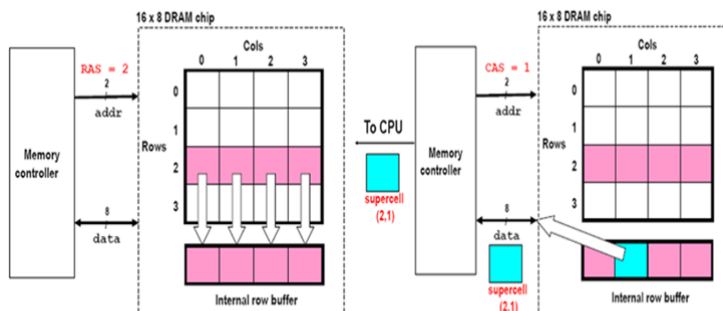
主存模块的连接和读写操作

° DRAM芯片内部结构示意图



图中芯片容量为16×8位, 存储阵列4行×4列, 地址引脚采用复用方式, 因而只需2根地址引脚, 每个超元 (supercell) 有8位, 需8根数据引脚, 有一个内部的行缓冲 (row buffer), 通常用SRAM元件实现。

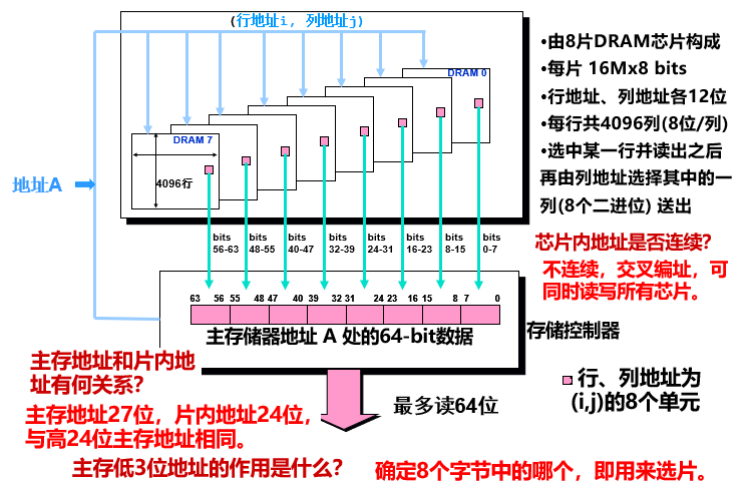
° DRAM芯片读写原理示意图



首先, 存储控制器将行地址“2”送行译码器, 选中第“2”行, 此时, 整行数据被送入行缓冲。然后, 存储控制器将列地址“1”送列译码器, 选中第“1”列, 此时, 将行缓冲第“1”列的8位数据supercell(2,1)读到数据线, 并继续送往CPU。

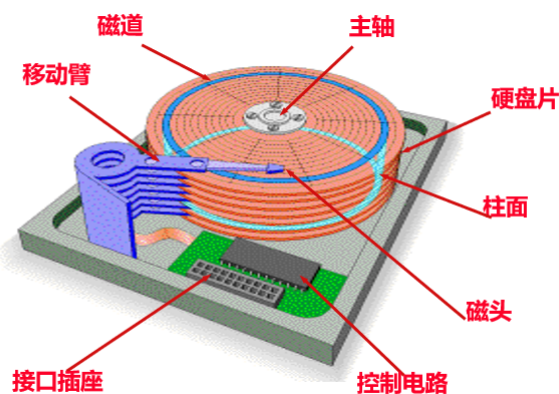
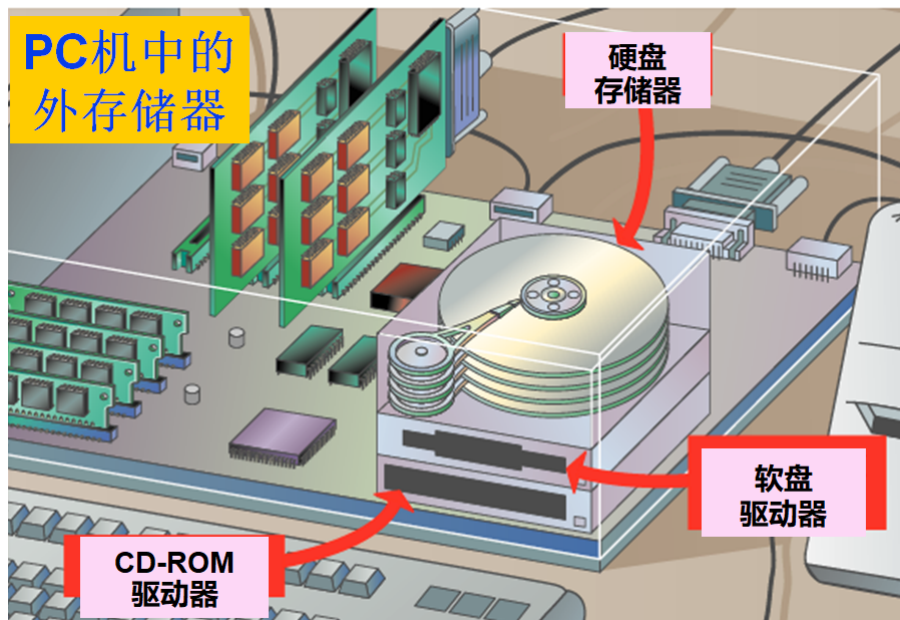
DRAM存储器实例

举例：128MB的DRAM存储器

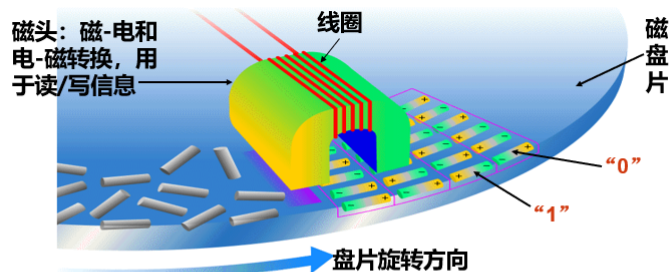


第三讲：磁盘存储器

磁盘图示



磁盘存储器的信息存储原理

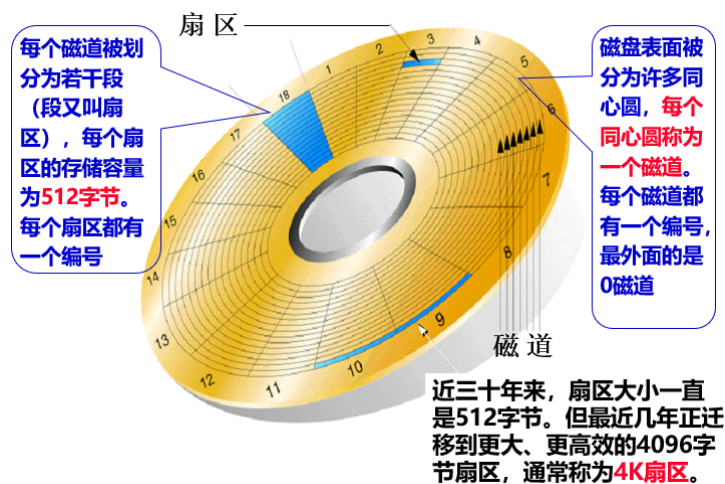


写1：线圈通以正向电流，使呈N-S状态
 写0：线圈通以反向电流，使呈S-N状态

不同的磁化状态被记录在磁盘表面

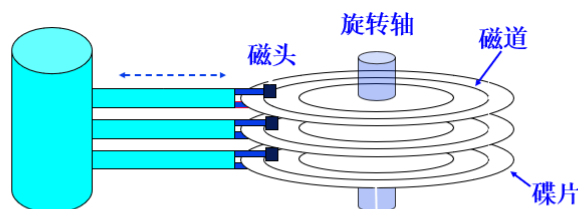
读时：磁头固定不动，载体运动。因为载体上小的磁化单元外部的磁力线通过磁头铁芯形成闭合回路，在铁芯线圈两端得到感应电压。根据感应电压的不同的极性，可确定读出为0或1。

磁盘的磁道和扇区



磁盘格式化操作指在盘面上划分磁道和扇区，并在扇区中填写ID域信息的过程

磁盘操作流程



硬盘的操作流程如下：

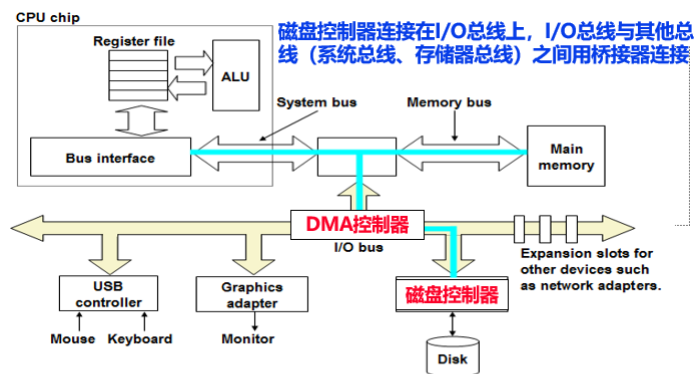
所有磁头同步寻道（由柱面号控制）→ 选择磁头（由磁头号控制）→ 被选中磁头等待扇区到达磁头下方（由扇区号控制）→ 读写该扇区中数据

° 磁盘上的信息以扇区为单位进行读写，平均存取时间为：

$T = \text{平均寻道时间} + \text{平均旋转等待时间} + \text{数据传输时间（忽略不计）}$

- 平均寻道时间——磁头寻找指定磁道所需平均时间（大约5ms）
- 平均旋转等待时间——指定扇区旋转到磁头下方所需平均时间（大约4~6ms）（转速：4200 / 5400 / 7200 / 10000rpm）
- 数据传输时间——（大约0.01ms / 扇区）

磁盘存储器的连接



磁盘的最小读写单位是扇区，因此，磁盘按**成批数据交换**方式进行读写，采用**直接存储器存取 (DMA, Direct Memory Access)** 方式进行数据输入输出，需用专门的DMA接口来控制外设与主存间直接数据交换，数据不通过CPU。通常把专门用来控制总线进行DMA传送的接口硬件称为**DMA控制器**

固态硬盘 (SSD)

固态硬盘 (Solid State Disk, 简称SSD) 也被称为电子硬盘。

它并不是一种磁表面存储器，而是一种**使用NAND闪存组成的外部存储系统**，与U盘并没有本质差别，只是容量更大，存取性能更好。

它用**闪存颗粒代替了磁盘作为存储介质**，利用闪存的特点，以**块写入和抹除**的方式进行数据的读取和写入。

电信号的控制使得固态硬盘的内部**传输速率远远高于常规硬盘**。

其接口规范和定义、功能及使用方法与传统硬盘完全相同，在产品外形和尺寸上也与普通硬盘一致。目前接口标准上使用USB、SATA和IDE，因此**SSD是通过标准磁盘接口与I/O总线互连的**。

在SSD中有一个闪存翻译层，它将来自CPU的逻辑磁盘块读写请求翻译成对底层SSD物理设备的读写控制信号。因此，**这个闪存翻译层相当于磁盘控制器**。

闪存的**擦写次数有限**，所以频繁擦写会降低其写入使用寿命。