# NAND Flash存储的可靠性

## 导论

与其他半导体芯片相比，NAND Flash的可靠性问题更值得关注，这是因为其无论是擦除还是编程，均会施加一个>10MV/cm的高场强于阵列MOS的氧化层上，相较于此其他芯片往往不到5MV/cm。

这是由NAND型Flash的阵列架构和算法所决定的，即同一个bitline上，往往是32个单元串联，且编程均是一整页进行（通常为2KB）。不同于NOR型Flash，利用的是热电子注入的机制改变浮栅态（需要大的沟道电流，但编程电压不需要很高），NAND型Flash利用的是FN隧穿效应，这就需要很高的电压施加在阵列上（甚至高至20V）。

NAND型Flash的可靠性问题往往就是出现其本身的编程与擦除操作上。如图6-1所示，展示了当编程或者擦除时，FN隧穿对氧化层的一种退化机制，从浮栅向衬底注入的热电子，会在衬底激发电子-空穴对，而空穴在强电场下会进入氧化层，形成空穴势阱，另一方面，在FG的场压下，氧化层也会出现电子势阱，两种势阱导致氧化层退化，从而影响芯片的寿命。诸如是编程/擦除阈值窗口变窄（6.2,、6.3节会详细讨论）。

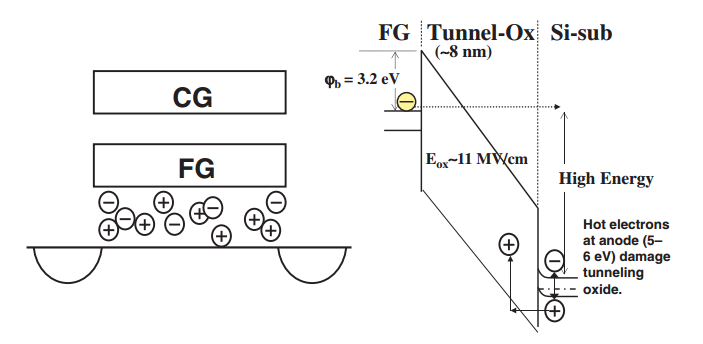


图6-1 FN隧穿应力下氧化层的退化机制

氧化层中的电子/空穴势阱同样会影响数据的保存（浮栅中电荷的维持），如图6-2所示，势阱的电荷释放是造成阈值漂移的根本原因。同时空穴势阱还会引起浮栅到衬底的漏电流，称之为SILC效应，而这一效应是造成阈值分布中拖尾效应的根本原因。

图6-3展示了读干扰的原理，芯片在经过一定的编程/擦除循环后，由于氧化层退化，SILC效应明显，当对某一页进行读循环时，会发现部分擦除态cell，发生错误，误读成编程态cell。

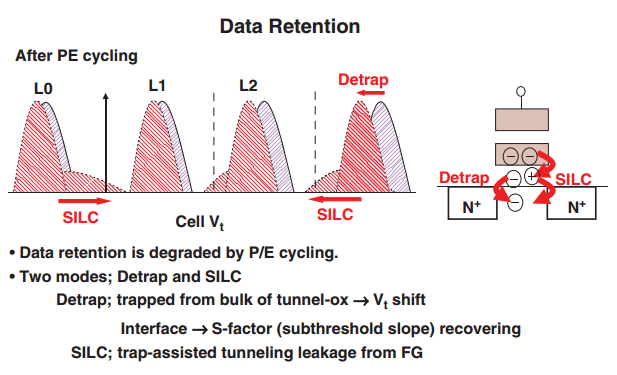


图6-2 数据维持现象

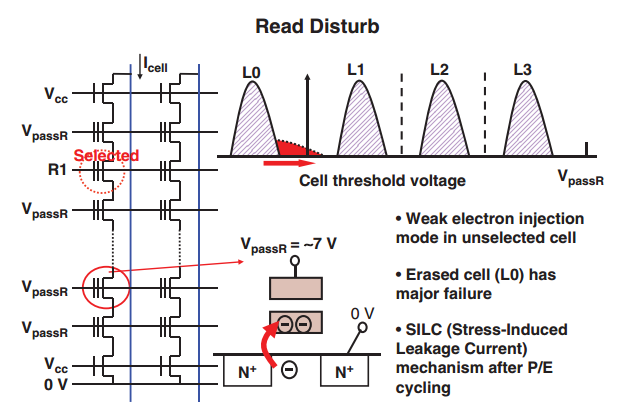


图6-3 读干扰现象

编程干扰随着NAND工艺不断推进愈发严重，主要是其算法中固有的self-boosting技术造成的，然而随着工艺节点的推进，出现了一些由于热电子注入引起的新的编程干扰，这些将在6.5章节详细描述。

6.6章节将阐述随机过编的问题，这一个问题研究认为是氧化层中空穴势阱对电子的捕获导致造成。因该问题而失效的cell数随着工艺节点的降低而增加。目前除了放慢编程操作外，没有什么很好的办法来减轻这一个问题，但通过ECC可以很好的规避这一问题。

在NAND Flash的编程与擦除操作中，会对CG和衬底施加一个高压，而这个高压并不会随着工艺尺寸的降低等比例的降低。因此会出现高场强的问题，其中一个问题便是编程过程中阈值负向偏移，这将会在6.7章节详细描述。

## 编程擦除循环数和数据保持

### 编程和擦除方案

主要有两种方案，一种是统一编程，统一擦除，另一种是统一编程，非同一擦除（选择性擦除），如图6-4、6-5所示，然而由于NAND Flash阵列架构所限，NAND Flash仅适用于前一种方案。

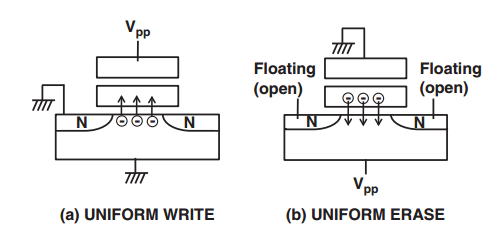


图6-4 编程擦除方案一

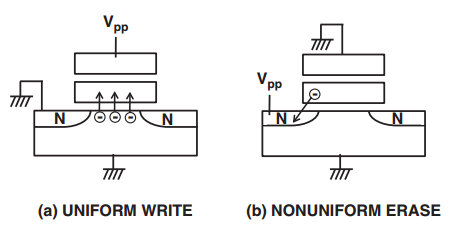


图6-5 编程擦除方案二

### 编程擦除循环数

图6-6为统一编程/擦除的循环次数与阈值的关系图，可以看到，编程阈值几乎不变，而擦除阈值在1K次之前略微下降，之后快速上升，这一现象可以通过氧化层势阱引起的隧穿电流以及平带电压的变化来解释。

对于擦除：

1K次循环之前，此时氧化层中空穴势阱占主导，且未捕获电子，因此FN隧穿电流得到加强（相同擦除电压下，浮栅被抽取的电子量更多，Vth擦得更负），平带电压增加（根据MOST Vth公式，Vth变得更负），因此图中可以看到阈值向负方向偏移。

1K次循环之后，此时氧化层空穴势阱逐渐捕获电子，同时电子势阱逐渐占主导，因此FN隧穿电流被削弱（相同擦除电压下，浮栅被抽取的电子量更少，Vth擦得更正），平带电压减少（根据MOST Vth公式，Vth变得更正），因此图中可以看到阈值在1K次循环后快速的上升。

对与编程：

1K次循环之前，此时氧化层中空穴势阱占主导，且未捕获电子，因此FN隧穿电流得到加强（相同编程电压下，浮栅抽取的电子量更多，Vth编程的更高），平带电压增加（根据MOST Vth公式，Vth变得更负），两者相互抵消，因此图中可以看到阈值基本不变。

1K次循环之后，此时氧化层空穴势阱逐渐捕获电子，同时电子势阱逐渐占主导，因此FN隧穿电流被削弱（相同编程电压下，浮栅抽取的电子量更少，Vth编程的更低），平带电压减少（根据MOST Vth公式，Vth变得更正），两者相互抵消，因此图中可以看到阈值基本不变。

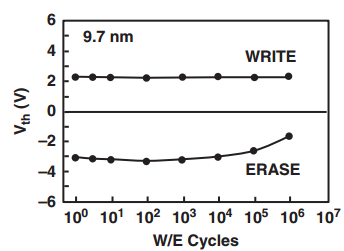


图6-6 阈值与编程/擦除循环次数的关系

图6-7为循环次数与浮栅管阈值变化ΔVth的实验结果图，该结果证实了氧化层电势阱对MOS管平带电压的影响，可以看到统一编程擦除的方案，平带电压的变化是ΔVth先负后正，这是因为氧化层在1K次之前为空穴势阱主导，之后为电子势阱主导。方案二之所以基本没有变化，是因为其不想方案一，氧化层势阱的产生以及分布是均匀随机的，方案二主要集中在漏端方向。

图6-8为图6-6基础上引起氧化层厚度的影响。从Vth变化的角度出发，可以看到氧化层厚度越薄，擦除态的Vth变化越小，阈值窗口的变化也就越低。说明，氧化层厚度和势阱的产生是正相关的。

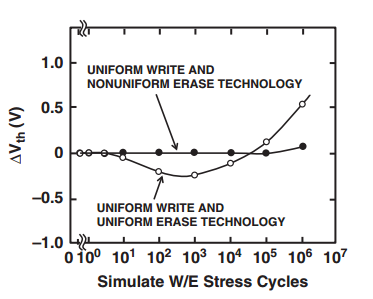


图6-7 W/E循环次数与ΔVth的关系

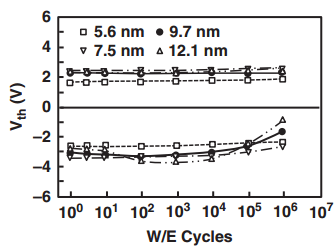


图6-8 W/E循环次数、氧化层厚度对阈值Vth的影响

### 数据保持特性

1. **编程/擦除方案的相关性**

图6-9为两种方案的高温烘培对cycling之后cell阈值的影响，cell为擦除态，可以看到统一编程和擦除，烘培时间在100分钟内，其阈值向负偏，且cycling次数越多，负偏量越大，而方案二则随着烘培时间的增加，阈值不断正偏，cycling次数越多，变化越不明显。

这说明对于方案一来说，对于百万次cycling的cell，高温烘焙有利于擦除态数据的保持。这是因为高温烘培会促使氧化层中被势阱捕获的电子脱离捕获，这等效于增加了氧化层中空穴势阱的浓度，从而使得擦除cell的Vth降低。如图6-10所示。

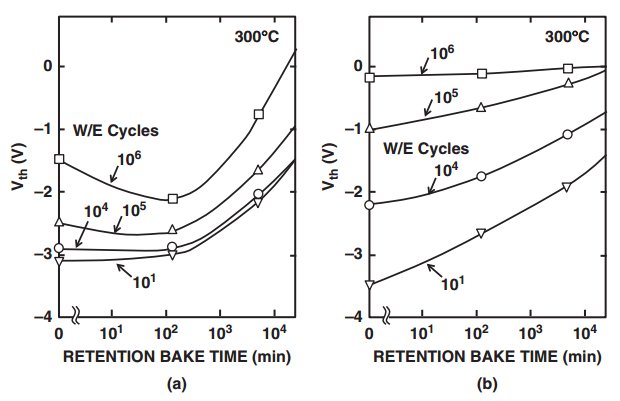


图6-9 烘焙时间与阈值的关系,（a）编擦方案一,（b）编擦方案二

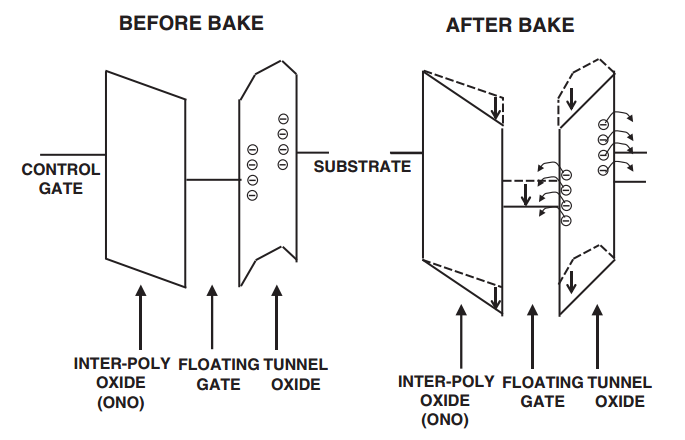


图6-10 高温烘焙前与烘培后能带图

1. **温度相关性**

图6-11是温度对阈值的影响结果，可以看到高温烘培100分钟以内，温度越高，氧化层去捕获情况越强烈，从而Vth负偏的越多。

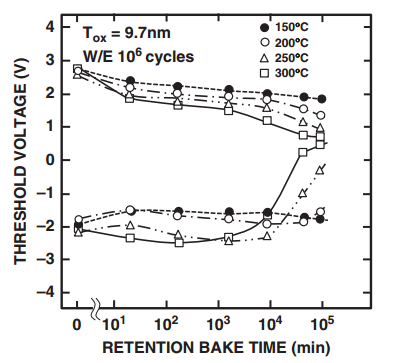


图6-11 温度对数据保持的影响

图6-12为不同温度下数据保持时间的评估，由于编程擦除的循环，数据保持时间被缩短，然而，在100℃下，100百万次的擦除循环仍能保证10年的数据保持时间。

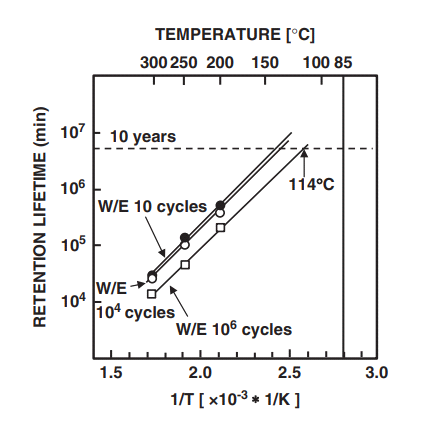


图6-12 不同温度下数据保持时间预估

数据保持时间主要由浮栅中电荷泄露机制决定，而不是氧化层中电荷去捕获。

1. **隧道氧化层厚度相关性**

为了搞清隧道氧化层厚度对数据保持的影响，实验测量了不同隧道氧化层厚度下的数据保持时间，如图6-13所示。

如6-14为20分钟烘焙后，编程态、擦除态cell的阈值变化和隧道氧化层厚度的关系，可以看到，编程态单元随着氧化层厚度的降低，Vth负向偏移量也降低，这是因为随着氧化层厚度变薄，其去捕获的电荷量也变得更少。然而对于擦除态cell，可以看到5.6n的厚度下，Vth的偏移量为正向，这是因为和去捕获影响相比，应力引起的漏电流占了主导。

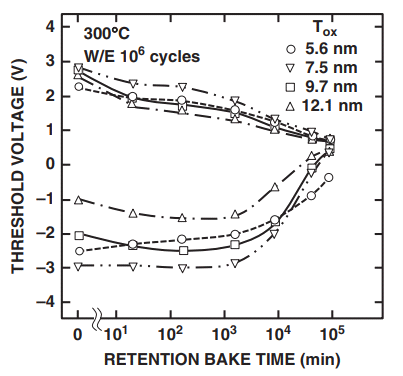


图6-13 阈值电压与氧化层厚度的关系

图6-15说明了数据保持时间与氧化层厚度以及编程擦除循环次数的相关性。对于薄的氧化层，10-10K的循环次数下，数据保持时间是不断缩短的。然而，一百万次循环后，由于电子退陷阱减少导致窗口缩小，数据保留时间延长。因此，隧道氧化层厚度的缩减只5.6nm不会受到数据保持问题的限制。

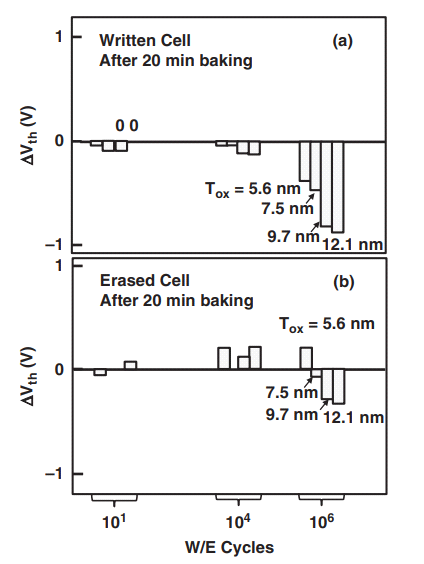


图6-14 300℃下20min烘焙，不同氧化层厚度下△Vth的偏移量。(a) 编程态cell, (b) 擦除态cell

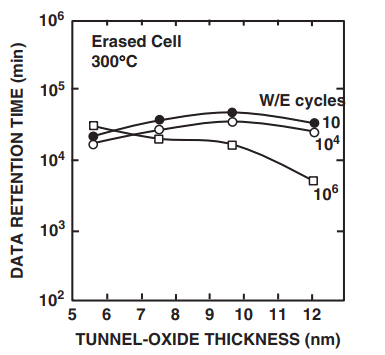


图6-15 数据保持与遂氧化层厚度以及cycling次数关系

综上所示，采用统一编程及擦除的方案，即使在一百万次循环后，仍能保证一个宽的电压阈值窗口。实验证明，由栅氧至衬底的电荷去捕获，有助于擦除态cell的数据保持。总而言之，统一擦除编程有助于实现NAND Flash的高稳定性。

## 编擦cycling与数据保持的分析

### 编擦cycling的退化

芯片在重复编程擦除后，其性能与可靠性都会退化。这主要是由FN隧穿效应导致的隧氧化层退化引起的。目前已经报道了许多在薄隧氧化层中的退化现象。

首先是关于编程擦除的脉冲对Vth的漂移的影响。NAND Flash的编程/擦除算法中，Vth的漂移通常发生在擦除态cell，正如6.2.2章节所描述，这是因为擦除态的Cell，其阈值Vth对氧化层中的电势阱更加的敏感。

为了研究编程和擦除脉冲的影响，我们对在编程和擦除脉冲形状下，擦除态Cell阈值的退化进行比较。图6-16显示了四种不同编程/擦除脉冲形状。A和B是低的擦除应力，A高编程应力，B低编程应力；C和D是低的编程应力，C高擦除应力，D低擦除应力；6-19a，b分别显示了A/B、C/D擦除态Cell阈值变化情况，如图所示，上百次的循环后，Vth出现负偏现象，这是因为隧氧化层中空穴势阱的增加，导致Vth的降低以及FN隧穿效应的增强。千次循环后，由于空穴势阱对电子的捕获，Vth逐渐正偏。由图中可以看到，A和C高应力的脉冲形状，其空穴势阱的产生是B和D的10倍，因此其Vth在千次循环内，降低的更多。

A和C相比，可以看到C条件下Vth偏移的更多，一种可信的解释是，高擦除应力会使更多的空穴势阱在更靠近Si/SiO2界面上产生，这些更近的Si/SiO2界面的空穴势阱，对read disturb，PGM disturb以及随机过编有着更大的影响。因此为了更好的可靠性，擦除算法需要更好的关注和控制。  
 接下来，讨论存储单元退化现象。如之前讨论的，编程/擦除循环会对隧氧化层与衬底界面的界面态和界面势阱的产生有着很大的影响。图6-17显示了cell的电流与迁移率随编程/擦除循环的退化关系。在编程/擦除循环100K次，并进行250℃168小时的烘焙测试中，通过测量cell的Id-Vg曲线，我们可以了解到退化的原因。氧化层势阱（Not）的产生以及电荷泄漏可以通过带隙电压的偏移来反映，而界面势阱浓度（Nit）可以通过管子的压阈值斜率来反映。

编程/擦除循环期间，氧化层中的电势阱不断产生，并捕获电子，从而使得间隙电压Vmg向正向偏移，如图6-17所示，擦除cell在经历103到105次的循环。然而这一现象在编程态cell中无法看到，这是因为电子的捕获同时会降低FN隧穿效应，从而Vth的正向偏移。图6-18是阈值在cycling和retention测试中的实验图。如图6-18b所示，阈值电压的漂移，可以由间隙电压Vmg的漂移和亚阈值斜率的变化来界定。间隙电压Vmg的漂移说明了在循环测试中，氧化层电荷阱的产生，以及在数据保持测试中电荷的泄露。亚阈值斜率的退化以及饱和电流的减少说明了在数据保持测试中界面电荷阱的产生。亚阈值斜率以及饱和电流的恢复则说明，在250℃烘培实验中由于电荷的去捕获，界面电势阱消失了。在cycling测试中，氧化层电荷阱以及界面电荷阱将会不断产生，然而在retention测试中，将产生电荷泄漏以及界面势阱的消失。界面势阱的消失对阈值Vth的影响莫种程度上是要大于电荷泄漏的。至此，我们可以得出结论，在NAND Flash中，界面势阱效应对于其数据保持以及退化至关重要。

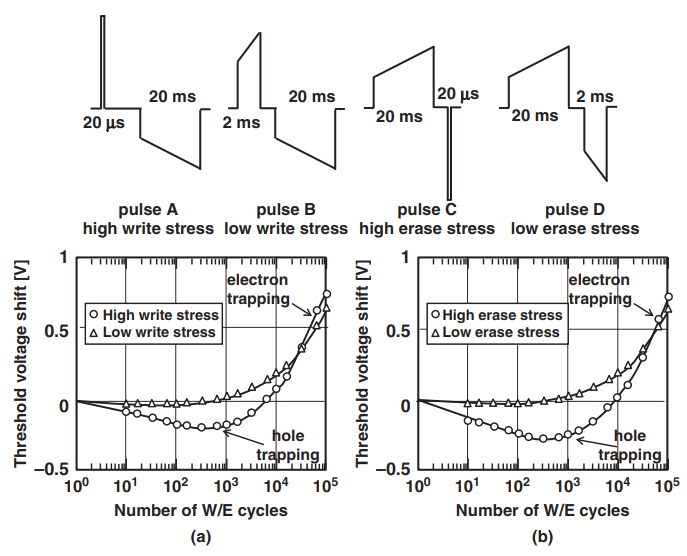


图6-16 不同的擦除编程脉冲波形以及其对擦除cell阈值的影响

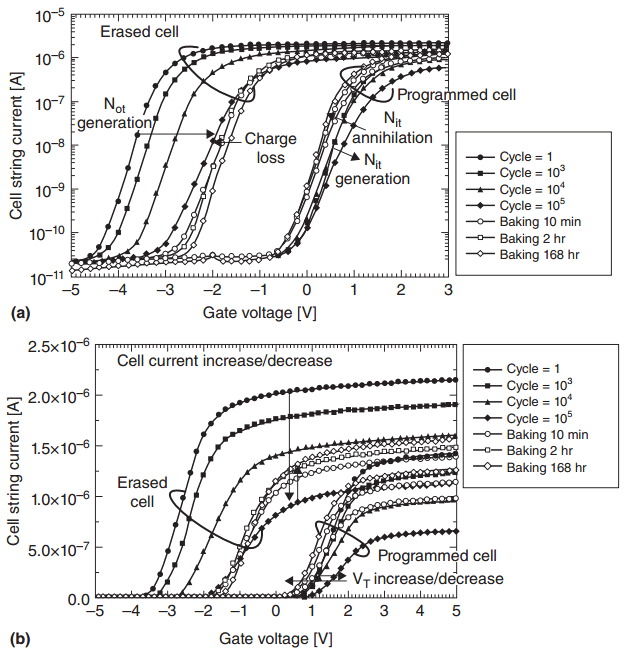


图6-17 编程/擦除态cell的Id-Vg曲线（a）为对数坐标，（b）为线性坐标

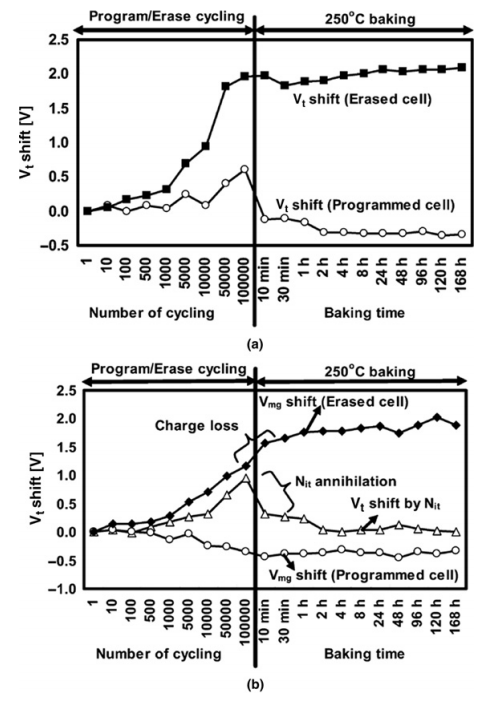


图6-18 cycling和retention实验中，阈值漂移的分析图。（a）为编程/擦除态cell的阈值变化（b）为编程擦除态cell，因Vmg以及Nit反映的Vth变化

### 应力诱生泄漏电流（SILC）

图6-19为NAND Flash单元在100K和一百万次编程/擦除循环，1000小时烘培后，室温下阈值Vt的分布。烘培之前Vt大于3.9V。有很少的一部分cell出现了电荷泄漏，Vt降低，形成了分布上的“拖尾bit”。这一数量将随着循环次数超过10万次而增加。而这一电荷泄漏效应（SILC）与Vt有着很强的相关性，越高的Vt将产生越强的SILC效应，从而导致更大的Vt偏移。这一泄漏电路由SILC效应产生的Vt偏移量计算，如图6-20所示，J/S为泄漏电流密度，J等于

其中Ccg-fg是控制栅至浮栅的电容。当场强Eox低于1.2MV/cm，即阈值Vt低于2V时，这一泄漏电流是非常小的。然而当Eox爬升至1.4MV/cm时，泄漏电流将随着Eox呈指数级增长。

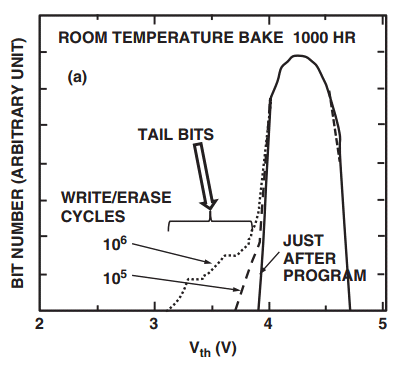


图6-19 NAND Flash单元在不同cycling次数下的阈值分布

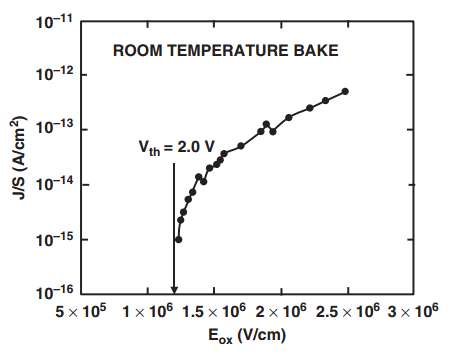


图6-20 拖尾bit泄漏电流密度与氧化层场强的关系图

图6-21为SILC重复性的研究实验结果，实验对相同的cell分别进行两次的retention实验，并实验后进行分别进行两次阈值Vt的测试，拖尾bit的行为可以大致分为两组。一组是由拖尾bit转变为正常bit，一组是仍然表现为拖尾bit。大约90%的拖尾bit再次经历编程然后retention实验后仍为拖尾bit，而10%会转变为正常bit。这说明拖尾bit可以较为容易的变回正常bit。如图6-22所示，可以看到一种意料之外的cell，定义为“stop bit”，单独追踪这些cell，我们发现其在室温retention实验中突然由拖尾bit转变为正常bit，“stop bit”的存在说明了拖尾bit与正常bit之间的变化容易性。

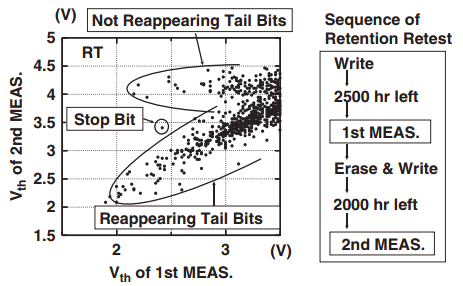


图6-21 SILC cell的复现，在经历两次编程后retention实验

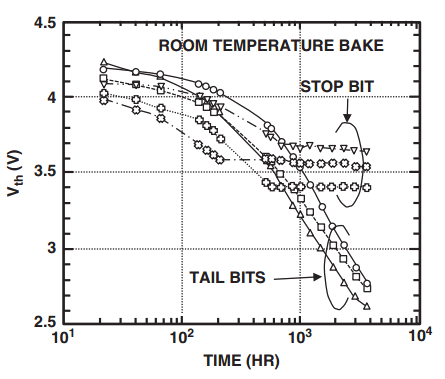


图6-22 异常cell的电荷泄漏特性，其中有三个为“stop bit”，其电荷泄漏突然终止。

实验结果说明，SILC是因为电荷阱辅助隧穿引起的电流所致。而电子的捕获和去捕获促使拖尾bit的再次出现与否。实际也证明，拖尾bit异常的泄漏电流仅通过一个或几个节点。这些泄漏通路很容易在激活与非激活状态间切换。

### NAND Flash产品的数据保持时间

这里对不同公司的NAND Flsh产品的数据保持性能做了比较，如图6-23a所示，为10K次循环后室温烘焙期间的原始误码率（RBER）。Time=0的误码率为cycling引起的错误，随着retention测试时间的增加，由数据保持失效引入的误码率逐渐增加。

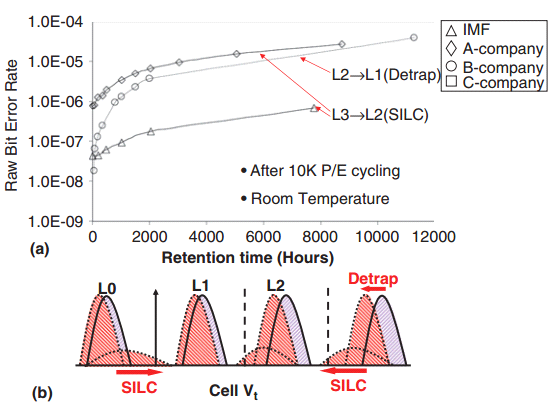


图6-23 10K P/E循环后RBER与室温保持时间的关系

数据保持失效主要是电荷损失造成的。有两种主要的机制导致数据保持错误。一个是SILC效应导致的浮栅电荷的泄漏，另一个是隧氧化层电荷阱的去捕获。图6-23b显示两种效应对阈值分布的影响。去捕获会导致分布变宽且负向偏移，而SILC会造成分布的拖尾现象。由于相对去捕获，SILC效应与场强更强相关，因此L3区域的误码率主要因为SILC效应，因为L3的阈值更高，意味着氧化层的场强更高。去捕获效应则是L1和L2区域误码率的主要因素，而去捕获效应同时还会造成“随机过编”的编程错误。

对于上述的格阈值区域（L0-L3）的误码率，各家公司有着不同的策略与侧重点，有些公司侧重于降低<1K次循环下的RBER，有些则侧重于10K次循环的RBER。

为了降低应用中RBER，定义NAND Flash的实际应用指标是很重要的，例如温度范围，主要的工作温度，编程/擦除循环次数，编程/擦除循环次数分布，读操作次数等等。根据这些应用条件，供应商可以据此优化诸如阈值设置等条件来降低RBER。由于应用的不断广泛，产品线需要进行分离，以满足每个应用的标准。

### 分布式循环测试

目前已有报道分布式循环测试结果，Compagnoni等人也对NAND Flash编程/擦除循环导致的阈值不稳定性进行了详细的实验研究，重点在编程/擦除循环时间和温度的相关性上。这项研究目的是，在SILC和去捕获效应的机理下，获得一个合理且通用的测试条件，以保证NAND Flash产品的质量。

当cell经历cycling后处于编程态时，由于在cycling中产生的损伤的逐步恢复，其阈值的不稳定性表现为，随着时间的推移，其阈值分布逐渐负向偏移。阈值的丢失不仅展现出对隧道氧化层场强的强相关，也与cycling的条件密切相关。特别是，阈值电压的瞬态特性，是被cycling中较长的时间间隔或是对数时间轴上更高的温度所延迟的。在60和41纳米技术上，将延迟因子作为循环持续时间和温度的函数进行研究，并提取NAND通用损伤恢复度量所需的参数值。

图6-24示意性地显示了在多级NAND闪存芯片（MLC）上循环后测试Vt不稳定性最常用的实验程序。（1）在tcyc=N\*twait时间内，施加一定数目N的P/E循环次数（twait为两次循环间的延迟时间）。（2）进行PV算法来确保cell某一编程态Vt。（3）在最后一次cycling后延迟一段时间，自第一次读操作，以对数间隔时间tB，对Vt进行监控。值得注意的是，Vt监控（读取）阶段对应于温度TB下的数据保留实验，温度TB可以是室温（RT），或者通常是选定的烘焙温度。后一种情况，烘培温度会周期性的中断并使芯片恢复至室温，再进行Vt的读出。

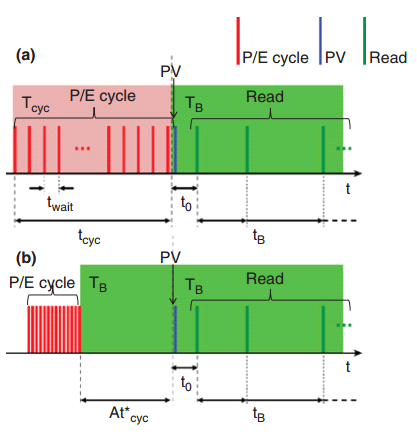


图6-24 实验程序示意图(a)循环引起的Vt不稳定性(b)分布循环的等效模型

在图6-24a的实验测试中，循环结束时出现的单元损伤量是由P/E循环产生的损伤和循环之间时间内的损伤恢复决定的。假设P/E循环产生的损伤既不取决于间隔时间twait，也不取决于实验温度Tcyc，并且循环过程中的损伤恢复可通过在损伤产生后，在温度Tcyc下的时间tcyc成比例的烘培温度TB和时间tb再现，则图6-24a的试验程序可以等效成图6-24b，也就说图6-24b可以通过更短的时间等效图6-24a的实验步骤。于6-24a中PV之前产生的单元损伤量，在后一实验中得到了等量的再现，然而却是在室温下更快的P/E cycling以及损伤恢复时间At\*cyc，其中A是由实验决定的常数。为了在单一温度下处理损伤恢复，引入了温度TB下时间t\*cyc，对应实验a中温度Tcyc下时间tcyc：

其中激活能EA的Arrhenius定律被引入计算时间的转变。假设自损伤最后出现后，由于损伤恢复，Vt有着对数性下降，那自第一次读操作后，时间tB下阈值Vt的变化量ΔVt由如下公式进行推导：

其中α表示因损伤恢复引起的Vt降低程度。根据t\*B的定义，越长的cycling时间tcyc和更高的cycling温度Tcyc会引起更低的Vt-loss瞬态变化。

## 读干扰

### 编程/擦除方案的相关性

如图6-25所示，为三种编程/擦除应力条件下的SILC效应实验。表6-2为应力条件。三种条件的应力下，SILC效应结果如图6-26所示，与b和c相比，a应力条件下的SILC效应低了大约一个数量级，这一结果表明，SILC的源头可以通过反向隧穿应力消除，同时源头可能是氧化层中具有方向性的缺陷、损伤、或者空穴势阱。通过双极应力降低SILC效应可以环节读干扰，延长数据保持时间。

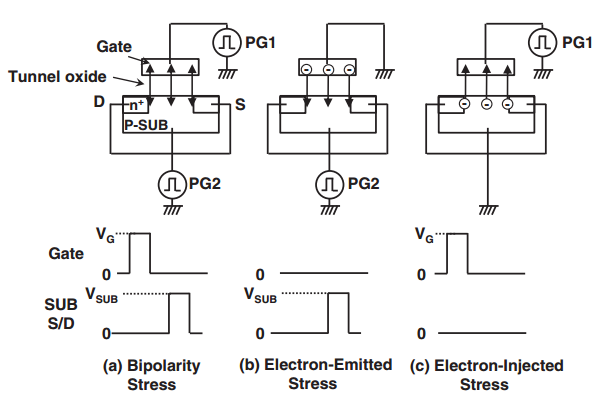


图6-25 编程/擦除应力实验方案(a)双极应力，(b)擦除应力，(c)编程应力

表6-2 应力条件

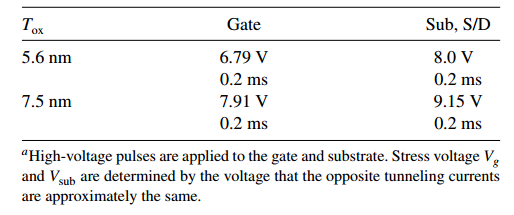


图6-26为两种编程/擦除方案，一种是双极FN-t W/E技术（即6.2.1节提到的统一编程/擦除），另一种是传统的沟道热电子编程和FN-t擦除技术（即6.2.1节提到的非统一编程，统一擦除）。

通过实验，可以得到如下结论：

如图6-27所示，不同栅压下，传统技术的阈值偏移斜率明显高于双极FN-t W/E技术，从而在数据保持时间上，双极FN-t W/E技术更占优。

如图6-28所示，降低隧氧化层厚度，有利于数据的保持，而降低隧氧化层厚度同时又能降低编程擦除电压。

如图6-29所示，300℃下，初始的数据丢失率，双极FN-t W/E技术也低于传统技术，这是因为传统技术的SILC效应高于双极FN-t W/E技术。

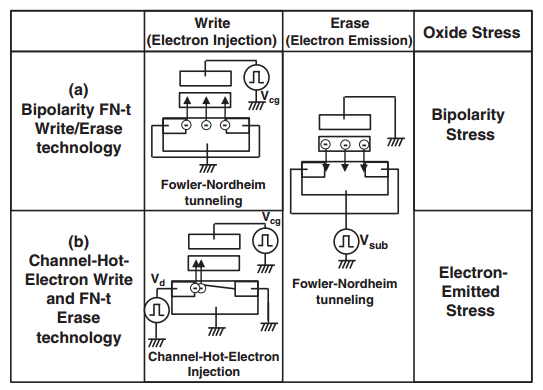


图6-26 两种不同P/E方案

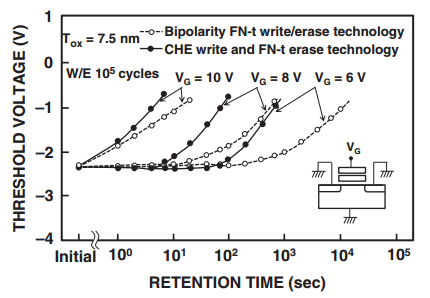


图6-27 不同栅压下，数据保持的结果

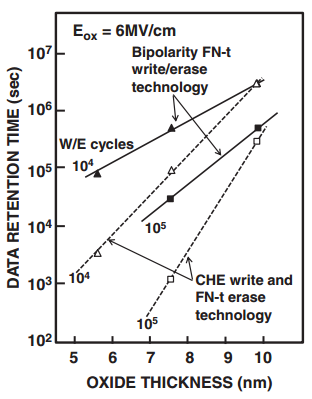


图6-28 数据保持时间与隧氧化层厚度的关系

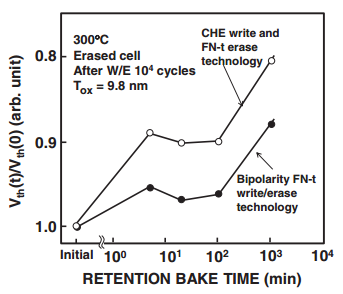


图6-29 两种技术下阈值变化与烘培时间的关系

总而言之，就读干扰以及数据保持时间而言，双极FN-t W/E技术均优于传统的沟道热电子编程和FN-t擦除技术，同时双极FN-t W/E技术更适合工艺制程的缩减（隧氧化层厚度降低）

### 去捕获和SILC

由于SILC效应，在经历P/E循环后，读干扰（read disturb）变得更加恶劣。

图6-30为不同栅压，室温一百万次P/E循环下阈值与读操作时间的关系，可以看到随着读操作时间的增加，擦除态cell的阈值不断正偏，其中SILC可以直接通过阈值偏移量以及读时间进行计算：

其中Ileak为SILC效应引起的漏电流，Cono为控制栅与浮栅之间的电容。

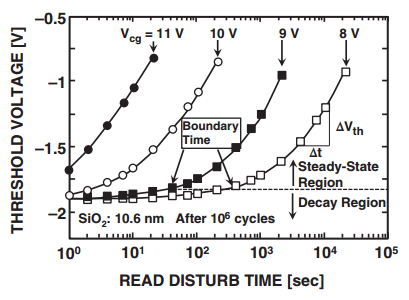


图6-30 读干扰特性与cg电压的关系

图6-31漏电流与隧氧化层场强的关系，漏电流是由阈值对读时间的差分推导出来的。从图中可以观察到泄漏电流在读取干扰应力开始时迅速衰减（衰减区域），经过衰减区后，泄漏电流达到一定的稳定值。两个区域可以通过在读时间的对数坐标系中，清楚的看出，如图6-32所示。在衰变区，dVth/dt的衰变是由SILC效应的快速衰变和隧道氧化物中陷阱载流子数量的衰变引起的。在稳定区，阈值的漂移主要由SILC效应引起，这一时期SILC电流无论栅压多少，均保持这一个定量。两个区域的边界，被定义为边界时间。

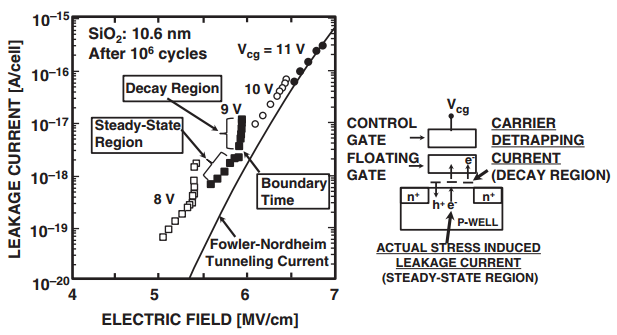


图6-31 漏电流与隧氧化层场强的关系

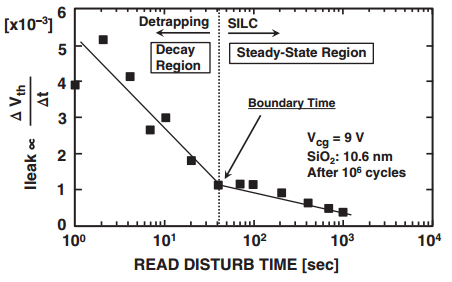


图6-32阈值对读时间的差分

图6-33为漏电流与P/E循环次数的关系，可以看到无论是在衰变去还是稳定区，漏电流均随着P/E次数的增加而增加，这些都说明，P/E次数的增加会增加隧氧化层中电荷阱数量。

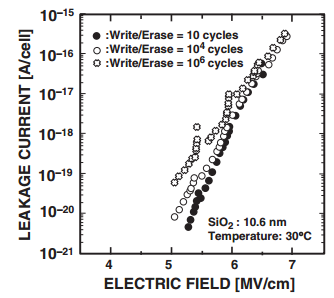


图6-33 不同P/E循环次数的漏电流

图6-34为百万次P/E后，漏电流与氧化层厚度的关系，可以看到随着氧化层的厚度的降低，漏电流是增长的。但是读干扰主要受稳定区影响，而不是衰变区影响。因此就读干扰寿命而言，降低稳定区漏电流远比衰变去更重要。

图6-35为SILC与温度的关系，可以看到，相较于室温，125℃高温下SILC稳定区电流是其三倍，因此，一方面在读干扰加速实验中，可以应用高温条件来缩短时间，另一方面边界时间也会随着温度的升高而缩短。

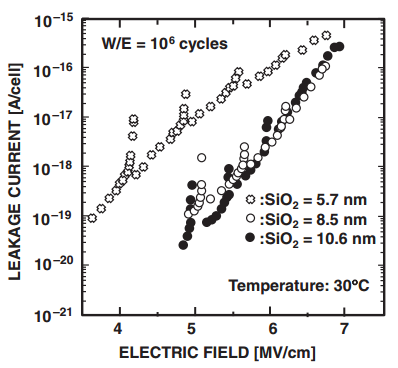


图6-34 SILC和氧化层厚度的关系

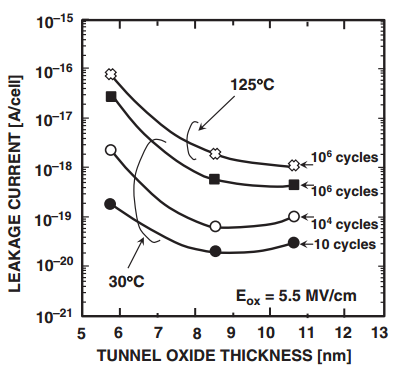


图6-35 SILC和温度的关系

### NAND Flash中的读干扰

图6-36a为芯片10K次P/E循环后，误码率与每页读次数的关系。读操作时，对于非选中的word line，偏置电压为VpassR，该值必须高于编程态cell的阈值，以保证非选中单元在读的时候处于导通状态，然后这一高压，会通过SILC以及隧氧化层中的电荷阱干扰cell的阈值。



图6-36 NAND Flash中读干扰特性。(a)RBER和读次数的关系，(b)P/E循环后的SILC机制

因为读的时候擦除态cell处于更高的场强，因此失效bit主要发生在L0区域的cell，如图6-36b所示。这符合SILC的机理，因为SILC是与场强强相关的，且在VpassR偏置下，擦除态cell有着最高的场强。通过排除编程错误，仅绘制读干扰的错误增量，来研究读干扰的特性，如图6-37所示。RBER在读取次数（图6-37a）和P/E循环次数（图6-37b）上呈幂律增加，再次与SILC一致。读干扰的失效率可以通过ECC进行补救。

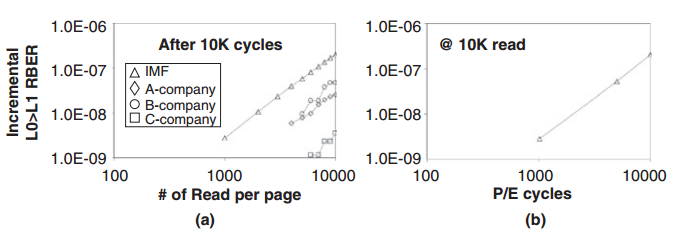


图6-37 L0区域cell的RBER与(a)读次数，(b)P/E循环次数的关系

### 读干扰中的热载流子输入机制

一种称为“增压热载流子输入效应”是读干扰的另一种机理，这一热电子注入是由于非选中单元的偏置VpassR意外引起的增压导致的。

为研究这一机理，我们对分别对四种阈值(S0-S3)施加三种读电压，如图6-38所示，进行100K次读循环。

读操作中，部分沟道电势(WL2-31区域)被VpassR耦合了上去，如图6-39所示。这一boost的电势将在选中管出产生热电子，因为此时选中管的源漏有着巨大的电势差（如6-39a，d）。由于这个巨大的电势差，一些热电子会注入至选中管相邻管子的浮栅中，如图6-40所示。

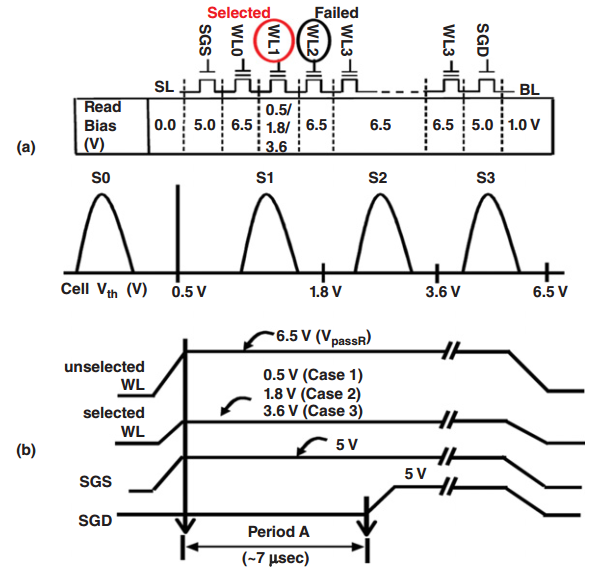


图6-38 (a)偏置条件,(b)操作波形

如图6-39所示，cell2在cell1偏置0.5V以及1.8V时，随着读循环次数的增加，产生了较大的阈值偏移，WL1=0.5V时，cell2的阈值偏移，主要发生在cell1的阈值为S2和S3区域的条件下，WL1=1.8V时，cell2的阈值偏移，主要发生在cell1阈值为S3区域的条件下。然而WL1=3.6V时，cell2没有看到任何的阈值偏移。

基于6-39的结果，选中管与非选中管沟道电势的差异与选中管阈值分布以及选中的的电压偏置Vread相关。如图6-39a所示，对于阈值为S2和S3的选中管，其沟道处于关断状态，此时cell2-cell31的沟道电势将被耦合至很高。因此，cell1和cell2之间的高电势差将导致一个高的横向电场。随着Vread的增加，选中管逐步导通，上述电势差将被减少，如图6-39c所示，cell2不再出现阈值偏移的现象。

电流密度是造成cell2阈值偏移的另一个因素，这也是为什么，图6-39a中，cell1处于S2态时，cell2的阈值偏移量比cell1处于S3态时要大的原因。

图6-40为MLC NAND Flash产品boosting HCI的原理图，沟道的电势差使得更像而场强增大，进而增加碰撞电离的可能性。结果是，电子空穴对增加，然后电子注入相邻的管子cell2，使得cell2的阈值随着读循环次数的增加而增大。

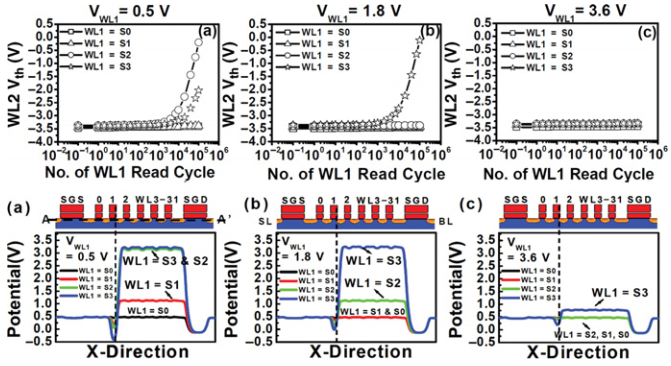


图6-39 WL2阈值随WL1读循环漂移,(a) Vwl1=0.5V,(b) Vwl2=1.8V,(c) Vwl3=3.6V

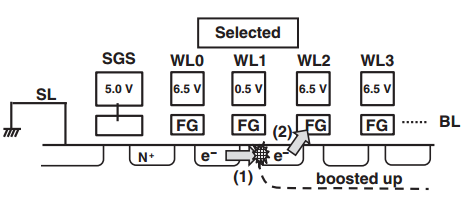


图6-40 boosting HCI原理图

## 编程干扰

### Self-Boosting模型

NAND Flash产品的编程算法中，对cell编0，通常是对沟道放电至0，栅压加Vpgm，对cell编1，则是将沟道先置位某一电压至(比如VCC)，在通过耦合效应，通过Vpgm和Vpass将沟道电压boost上去，从而抑制编程。而沟道耦合电压可以通过如下公式进行计算：

其中Cins是控制栅至沟道的电容，其值为：

图6-41为沟道boosting原理等效电路图，对于编程数据“1”的cell，沟道电压Vch必须被Vpass以及Vpgm耦合至足够高的电压来抑制FN电流。

编程抑制的能力与相邻cell的编程数据有关，如图6-41所示，电容Ccs是与相邻沟道电位相关的，当相邻沟道为0时，Ccs将为最大值，这会大大降低编程抑制的沟道耦合效率，从而造成PGM干扰，因此NOP，相对与编全1，编0和1相间的图形时是需要限制的。

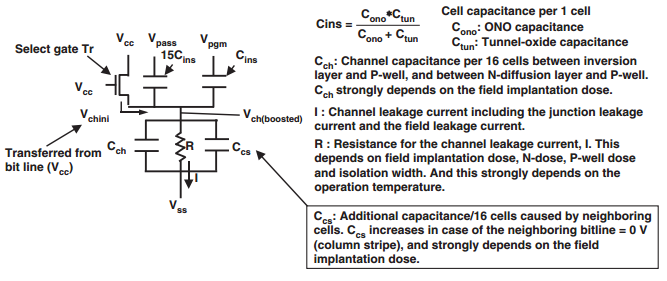


图6-41 沟道耦合等效电路图

为分析沟道漏电流，同样对编程电压算法与编程干扰的相关性进行了研究，如图6-42所示。在传统的数学模型中，认为Tpw=30us下，全1图形和柱状条纹图形之前的阈值差异是因为场漏电流所致。但是测试数据，随着脉宽增加逐渐偏离了模拟结果。然而，在引入电容Ccs的模型中，模拟结果与测试数据得到了拟合。当脉冲宽度大于1ms时，Vth增加，这是由结漏电导致的。

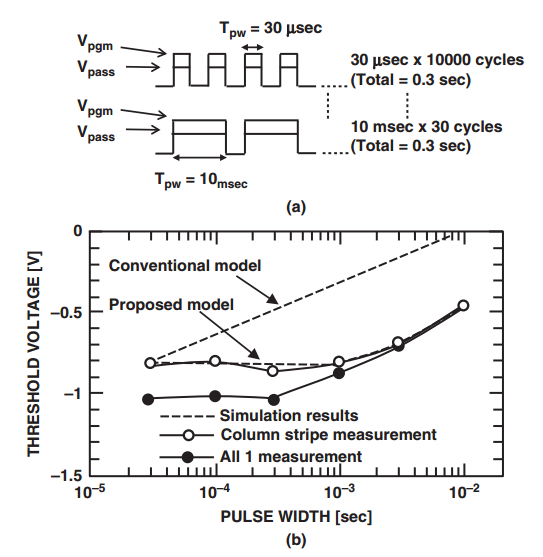


图6-42 编程脉冲宽度与PGM干扰的关系

在亚30nm NAND Flash工艺下，一种量化沟道boosting模型用来研究，沟道电容，沟道漏电流以及cell尺寸缩减对编程干扰的影响。这个模型包括沟道耦合效率（CBR），带结漏电流的瞬态沟道电势，带间隧道电流（BTBT）以及FN隧穿电流。

图6-43a为NAND Flash编程时string的简要图，6-43b为编程电压与抑制电压的波形图。

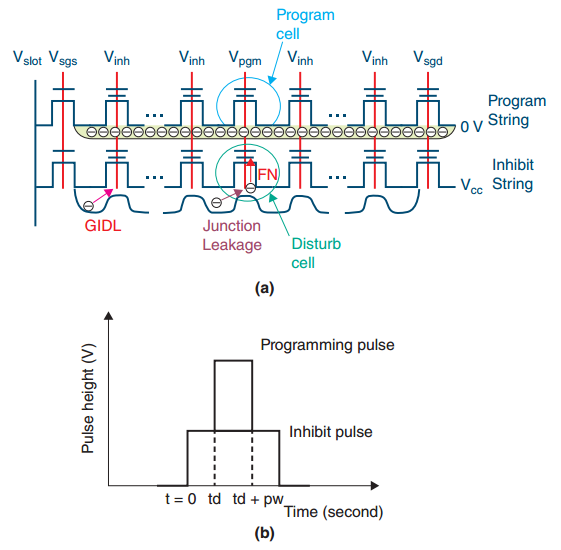


图6-43(a)NAND string编程时的简要图，(b)编程电压与抑制电压的波形

图6-44a和b为三种延迟时间(5us, 100us, 500us)下，阈值Vt与抑制电压（Vpass）的函数关系。模拟结果和实验结果在一系列抑制电压下有着很好的拟合。同时该模型低沟道浓度和高沟道浓度情况下的总体趋势。在低沟道浓度下，编程干扰随着抑制电压的升高得到改善，这说明沟道电势主要由沟道耦合率（CBR）决定。然而在高沟道浓度下，大约在7V的抑制电压下，阈值分布开始进入一种饱和状态，这说明，沟道耦合电势收到了沟道泄漏电流的限制。图6-44b所示，不带BTBT效应的模型结果与实际结果不匹配，这说明在高沟道浓度下，主要的沟道泄漏电流来自BTBT电流。

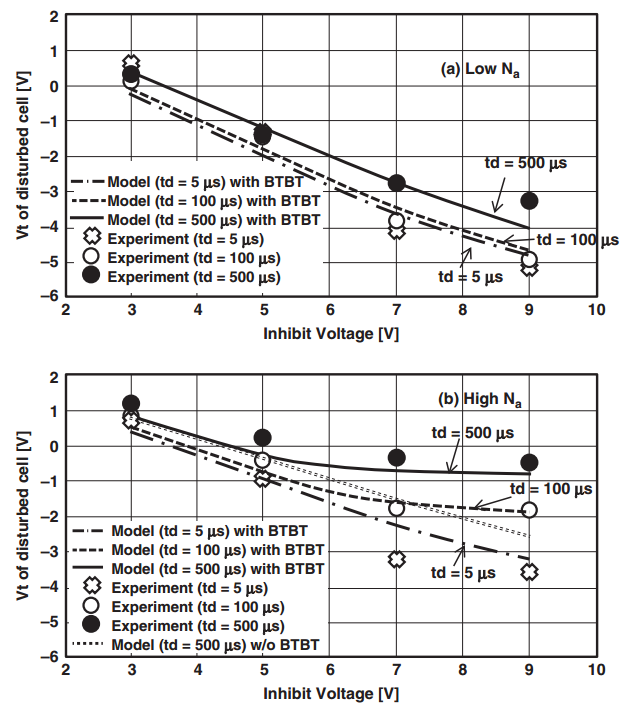


图6-44 阈值Vt分布与抑制电压的函数关系，(a)低沟道浓度下，(b)高沟道浓度下

随着NAND Flash工艺的缩进，需要更高的沟道浓度来环节短沟道效应。沟道boosting节点的漏电流增加。图6-45说明了，随着工艺迭代，其对沟道浓度的需求，以及因此导致的沟道泄漏电流的关系。沟道浓度是由维持电中性Vt所决定的。在20nm以下，BTBT电流将成为PGM干扰机制中的主导因素。

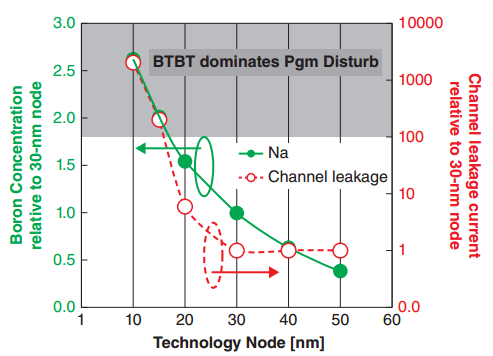


图6-45 沟道浓度以及泄漏电流关于工艺缩进趋势图

### 热载流子注入机制

Boosting模型和Vpass模型，这两种常见的编程干扰机制在2.2.4章节已经阐述。除这两种常见的失效模式外，还有一些其他的编程干扰机制。

如图6-46所示，表明了一种名为“SGS GIDL”的源漏热载流子注入的干扰。WL0单元还是WL15单元，均编程数据“1”，即抑制编程，图中可以看出，当Vpass电压低时，出现了阈值的漂移，这一现象是由于编程抑制效应降低（6.5.1章节描述），但是WL0在Vpass电压增高时同样出现了阈值的漂移，且交于前者影响更加严重，这一现象是由于MOS管的GIDL效应引起的，失效单元主要发生在靠近SGS管的边缘cell（WL0），且随着NOP的次数增加而增加。

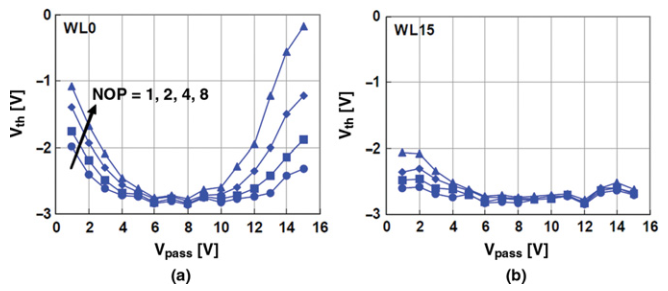


图6-46 WL0和WL15阈值随Vpass电压的变化

图6-47为GSL和WL0器件横向电势图，其阐明了GIDL效应产生的原理，从图中我们可以得到解决这一干扰的思路，即拉开GSL和WL0的距离，实际通用的做法是在GSL与WL0之间加入dummy cell。

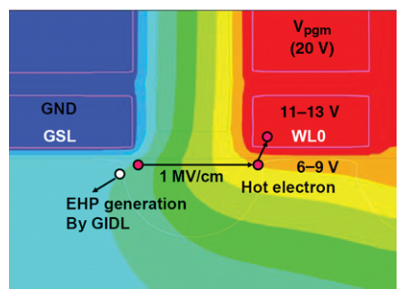


图6-47 GSL/WL0电势图

对于相同的扰动现象，也报告了另一种机制。该编程扰动同样由热电子造成，不同的是，该热电子是由SSL的氧化物-硅界面的复合中心产生，而不是GIDL效应。

### 沟道耦合

随着工艺尺寸的降低，沟道boosting电压也在下降。这一电压与相邻string的编程数据相关。Vcc-Vcc-Vcc模式，为相邻的两个string处于编程抑制状态，0-Vcc-0模式，为相邻的两个string处于编程状态，0-Vcc-Vcc，为相邻的两个string，一根处于编程抑制，一根处于编程状态。

图6-48展示了编程干扰与上述三种情况的关系。由图可以看出，相邻string沟道电势对当前沟道的boosting电压有着较大的影响，0-Vcc-0这种情况，编程抑制效果减弱，从而PGM干扰更强。

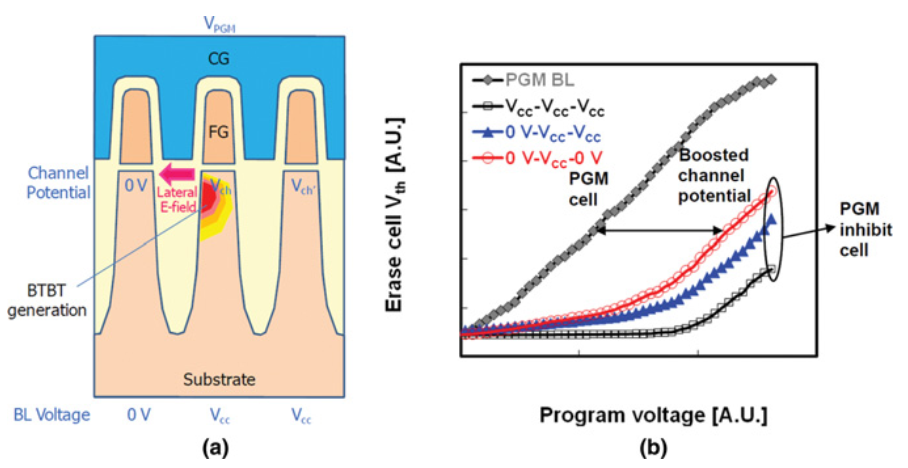


图6-48 (a)由横向场强引起的BTBT示意图，(b)三种编程图形下的编程干扰情况

为了探究上述情况的物理机制，如图6-48a所示，在0-Vcc-0的编程图形下，从器件切面图可以将相邻的三个cell等效看成一个侧壁晶体管，其中相邻沟道接地，可以等效为一个接GND的栅级，与当前string之间的介质可以等效为栅氧层，当前string靠近界面处的沟道等效为漏极，则根据编程操作时的电位关系，我们可以看到，一个大的“GIDL”电流以BTBT电流形式因沟道boosting而产生。因此我们可以看到，无论怎么增加Vpass电压，沟道电势都无法再增加，达到一个饱和值。

如图6-49a所示，为一种解决在1x-nm上述编程干扰的方案，对string之间的介质进行镂空，即引入“空气墙”，物理上改善BTBT的影响。其实验结果如图6-49b所示。

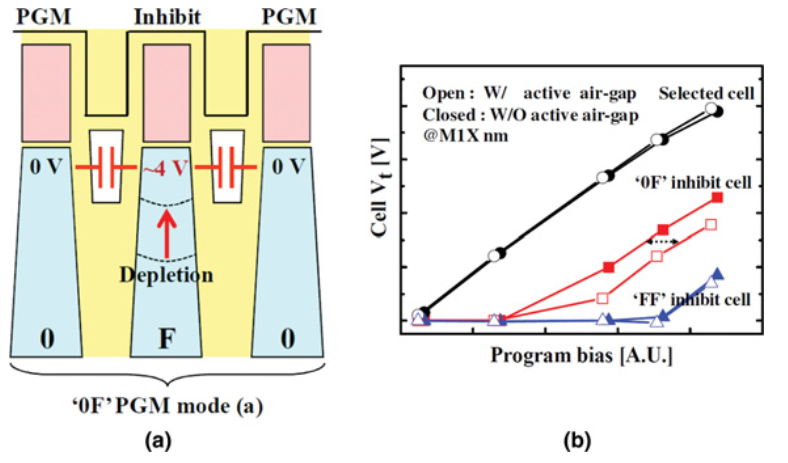


图6-49 (a)解决方案示意图，(b)解决方案与传统工艺的结果比较

## 随机过编程

随机过编程是指在编程操作中，某些cell出现异常Vth正偏，形成阈值分布上高边的拖尾现象。如图6-50所示。对于L0-L2区域的随机过编程，当这些拖尾bit阈值超过预期的Vread电压时，将产生单个bit错误，然而L3区域的随机过编程，这些拖尾bit的阈值有可能超过VpassR，从而在读操作时，直接关断整个string，导致整个block的所有页均出现读错误。

随机过编程的根本原因是隧氧化层中空穴势阱对隧穿效应的增强，如图6-51所示，隧氧化层中的空穴势阱降低了势垒高度。

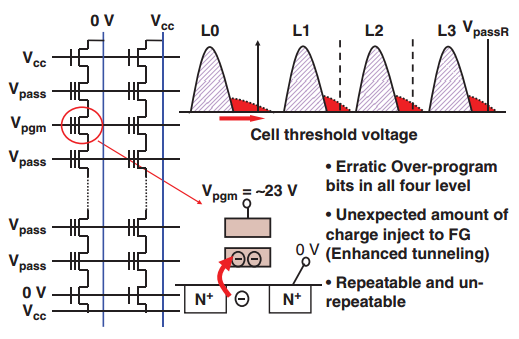


图6-50 随机过编程

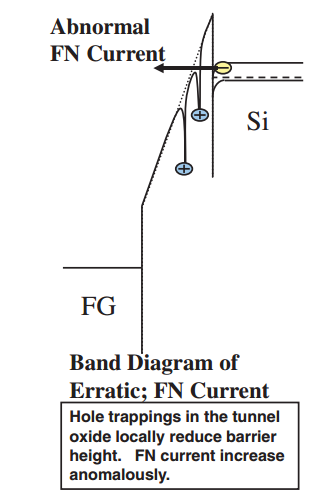


图6-51 随机过编程模型示意图

## 负向Vt漂移现象

### 背景和实验

自对准浅沟隔离单元（SA-STI单元）长期以来一直用于NAND闪存产品，从0.2μm代到目前的中1X-nm代。如图6-52所示，为SA-STI单元沿word line方向的结构图。在SA-STI单元中浮栅的侧壁用来增加FG到CG的电容，从而增加耦合效率。沟道至STI顶层的距离为场氧高度（FH），这一高度应尽可能小来提高沟道耦合效率，同时也会减小word line方向上FG-FG的耦合影响。然而，FH的减小，导致CG与沟道之间的压强增加，这对NAND Flash的性能和可靠性带来了影响。

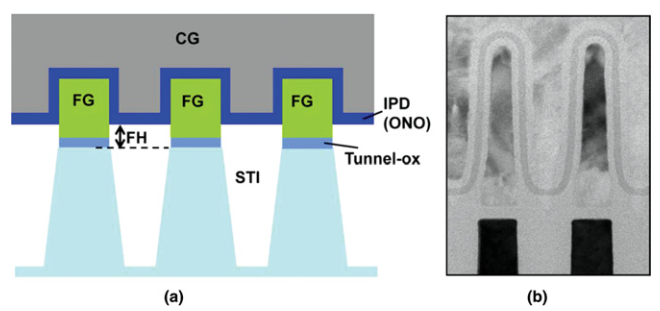


图6-52 沿word line方向的SA-STI结构图

本章节将描述，2x到3x-nm SA-STI工艺下编程抑制条件中出现的阈值负偏现象。这一现象发生在小尺寸FH情况下，即编程操作中高场强之一。由于这一负向偏移增加了阈值宽度，也就是缩小了阈值窗口，从而成为NAND Flash工艺缩减至2x-nm的一个新的障碍。

图6-53为编程抑制测试的单元排布。

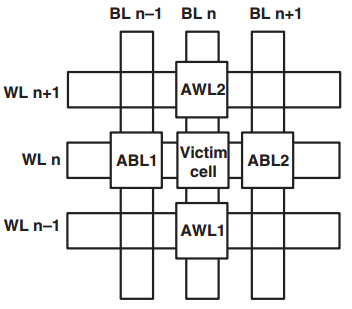


图6-53 编程抑制测试的单元排布

### 负向阈值漂移

图6-54表明了，相邻单元的编程对测试单元的阈值影响。对于AWL2上的单元，由于FG-FG耦合相应，随着其阈值的增加，测试单元的阈值也在单调增加。然而，对于ABL1和ABL2上的单元，随着其阈值的增加，测试单元的阈值先是增加，再是降低。如果尽是FG-FG耦合影响，邻近单元的阈值上升只会导致当前单元的阈值正偏，然而实际在邻近单元阈值超过7V后，出现了负偏，这一现象称之为“负向阈值漂移”。

图6-55揭示了负向阈值偏移与FH的关系。阈值负偏对FH有强相关性。当FH较小时，阈值负偏很大。在邻近cell的阈值小于6V的区间，测试cell的阈值与邻近cell的阈值比符合FG-FG耦合效应，FH越小，这一比值越小。这说明FH小，则CG对FG-FG的屏蔽效果越强。

图6-56展示了测试cell阈值与(a)邻近cell编程条件(b)邻近cell编程抑制条件的关系，b的实验条件为，对邻近cell进行编程，当其阈值Vt到达4V或者7V时，通过将其沟道电压抬高停止编程，即编程抑制。由图中可以看出，b实验条件下的负向阈值偏移情况明显小于a实验条件。

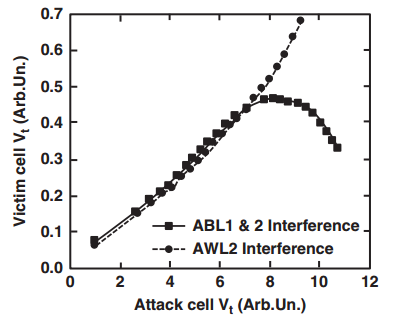


图6-54 测试单元的阈值随着邻近cell的编程的变化

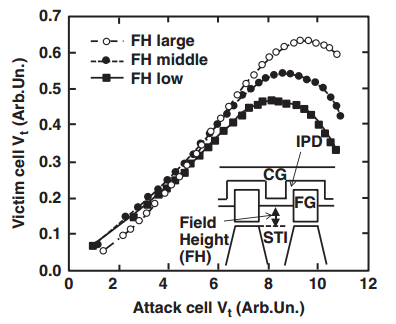


图6-55 阈值负偏与FH的关系

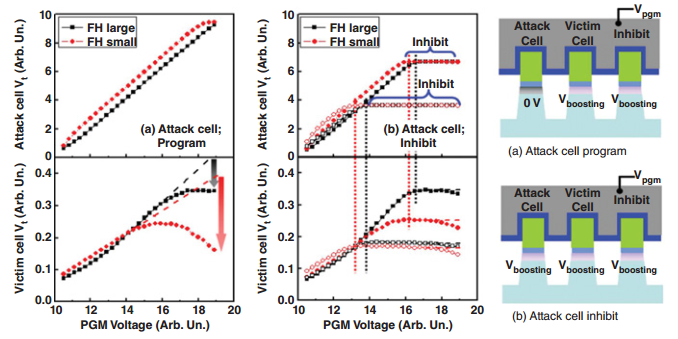


图6-56 负向阈值偏移和邻近cell操作条件的关系，(a)邻近cell持续编程，(b)邻近cell编程至一定阈值后抑制编程

### Victim Cell Vt与编程速度的关系

邻近cell的编程速度相关性是在真实的MLC NAND Flash编程序列中进行测试的。图6-57为测试cell ΔVt与邻近cell编程速度的关系。纵向坐标为经历一次编程脉冲后，整个页cell的阈值分布，图中左边cell为慢cell，编程速度较慢，右边cell为快cell，编程速度较快。如图6-57a的情况，邻近cell ABL1和ABL2均从Erase态编程至L1态，可以看到对于编程速度较慢的邻近cell，测试cell的阈值偏移更大，这是由于FG-FG耦合效应决定的，慢cell的编程，其阈值变化比快cell更大，从而造成的FG-FG耦合影响也就更大。

另一方面，如图6-57b情况，邻近cell ABL1和ABL2均从LSB态编程至L3态（更高的Vt），可以看到对于编程速度较慢的邻近cell，测试cell的阈值偏移更大，即使慢cell与快cell的阈值变化一致，即FG-FG影响一致，这说明相较于快cell，慢的邻近cell对测试cell的阈值负偏效应的影响更大。

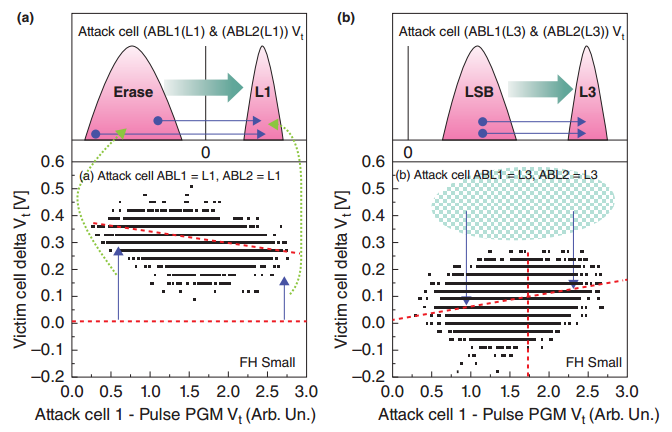


图6-57 阈值负偏效应与编程速度的关系

图6-58为阈值负偏效应与测试cell的阈值状态的关系，可以看到当测试cell阈值处于L3区域时，阈值负偏效应更强。

综上所示，我们可以总结出，在如下情况下阈值负偏效应加强：

1. 沿WL方向上的相邻cell进行编程操作时
2. 更小的FH
3. 更高的编程电压Vpgm
4. 相邻的cell阈值处于L3态（即最高的那个阈值区间，以MLC NAND Flash）
5. 相邻的cell为slow cell
6. 受阈值负偏影响的cell的阈值处于L3态

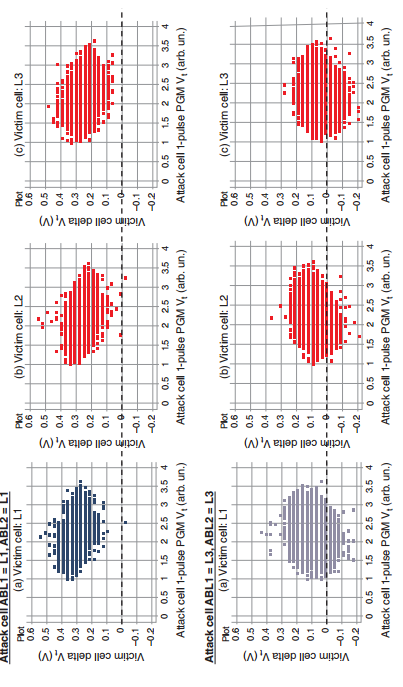


图6-58 阈值负偏效应与测试cell的阈值关系

### Carrier Separation in Programming Conditions

为了弄清阈值负偏效应的原理，我们对运用Carrier Separation技术制作的cell结构的测试电容器进行了测量，如图6-59所示。其测试条件为Vfg=8V，Vwell=Vjunction=0V下扫描Vcg。图6-60为电子流向图，图6-59中的电流Icg，Ifg，Ijunction均以电子流的形式在图6-60展示。

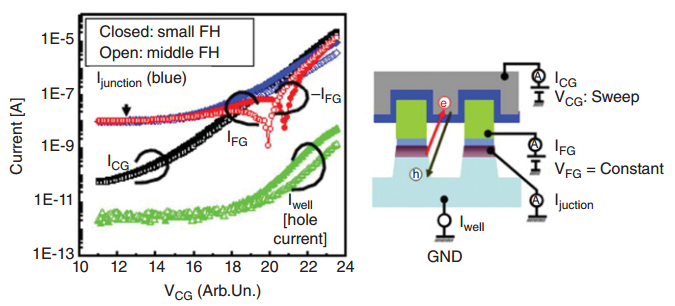


图6-59 cell结构的电容器的电流分析

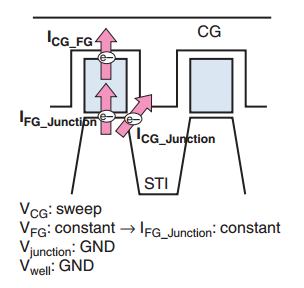


图6-60 电流流向示意图

从图6-59可以观察到当Vcg＞18V时，随着Vcg的增加，空穴电流(Iwell)增加，Icg\_junction和Icg-fg，由于Vfg为定值，所以Ifg-junction基本不变，结电流Ijunction在小FH下要大于高FH。通过图6-59和图6-60，Iwell由Icg-junction产生而不是Icg-fg，Icg-junction是从沟道到CG的FN电流。

对于Ifg，在Vcg＜20V时，由于CG-FG之间压差较小从而Icg-fg较小的缘故，其基本等于Ifg-junction = 定值。然而随着Vcg增加至＞20V，Ifg的极性将改变，这是因为Icg-fg将成为主导。

图6-61中Vcg为定值，扫描Vfg，其电流的变化。即使沟道与CG之间的电压为定值，但是电流Icg仍随着Vfg的增加而增加。这说明，随着工艺节点的推进，FG-FG耦合效应增强，因耦合效应引起的Vfg的增加，会增加电流Icg。这表明cell尺寸的缩小可能会增加阈值负偏的情况。

在图6-61中衬底空穴电流与栅电子电流的比（Iwell/Ijunction）在Vfg=3-5V范围内，大约在10-3数量级，衬底空穴电流主要由穿过隧氧化层的FN电流(Ifg\_junction)产生，这一电流值在相同的氧化层厚度与场强下与之前报道的在同一数量级上。然而在图6-60中衬底空穴电流与栅电子电流的比（Iwell/Icg）是在10-4数量级，其中Iwell主要由从沟道至CG的FN电流产生。这比之前报道的10-3-10-2的数量级低了1到2个等级。更小的衬底电流的原因目前还不清楚。然而，有别于之前报道中的平板电容器，由于是单元结构，产生的大量空穴并不会流向衬底。这说明空穴可能流向任意方向，包括浮栅方向。

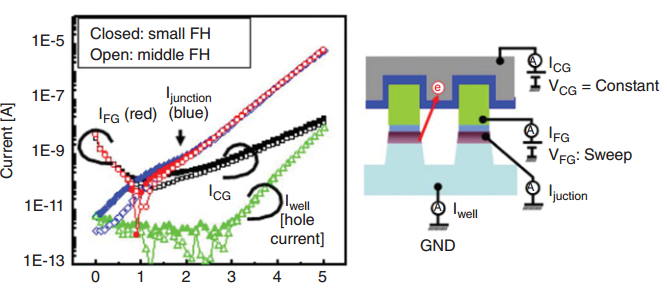


图6-61 单元结构电容的电流分析

### 模型

根据6.7.4章节的电流分析结果，阈值负偏效应的原理如图6-62所示。在同word line上的相邻cell进行编程时，电子电流直接由沟道注入至CG，而电子的注入，在CG中通过碰撞电离产生空穴，空穴注入至CG与FG之间的氧化层(IPD)。其中一部分空穴会由IPD或者是STI场氧介质注入到我们所观测的cell浮栅中，从而导致其阈值发生负偏。

这一现象将随着阵列单元的缩减而严重，因为FG-FG的距离变短导致更高的耦合效应，进而引起更大的Icg电流。也就说阈值负偏效应将会在未来尺寸缩进中更加恶劣，成为推进更小尺寸MCL、TLC等NAND Flash产品的新瓶颈。

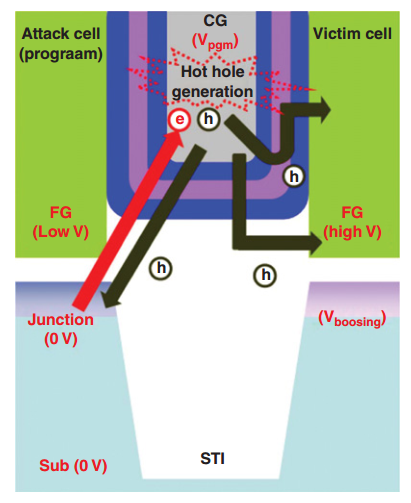


图6-62 阈值负偏效应原理图

## 总结

在整个第六章，对NAND Flash存储芯片的可靠性问题进行了阐述。

在6.2章节，描述了P/E循环中芯片的退化以及数据保持问题。一种统一编程/擦除方案，即编程和擦除时对整个沟道区域采用FN隧穿效应，可以保证在1百万次P/E循环后仍由较宽的阈值窗口。同时也能够保证较长时间的数据保持，因此统一编程/擦除方案是NAND Flash产品的业界标准。

在6.3章节，我们讨论了关于P/E循环次数和数据保持的一些可靠性问题。因P/E循环应力导致的芯片退化，主要是由于电子/空穴电荷阱，SILC以及界面态引起的，并且随着工艺的缩进，这些退化现象将变得更加严重。

在6.4章节，阐述了读干扰的特征。实验证明，相较于编程采用HCI效应擦除采用FN隧穿效应，编程和擦除均采用FN隧穿效应，有着10倍的数据保持时间。这主要是因为后者在薄氧化层下，有着更小的SILC效应，也因此，后者可以保证工艺的缩进。

6.4章节也阐述了读干扰的几种分析，起先是电子的去捕获导致阈值漂移，之后是SILC主导。6.4.4节同样介绍了热电子注入效应引起的读干扰，诸如GIDL效应等。

在6.5章节，讨论了编程干扰。由于难以意料的热载流子注入和沟道耦合的退化机制，随着阵列单元尺寸的缩进，编程干扰将变得难以控制。如6.5.2节描述的因相邻沟道电压差导致的热载流子注入效应。6.5.3节描述的相邻沟道电位差，引起耦合效率降低以及BTBT电流的影响，进而导致的抑制电压效率的降低。

在6.6章节，讨论了随机过编的问题。随机过编是因为编程过程中隧穿电荷量过量导致的，这一现象可以通过强大的ECC进行管理控制。

在6.7章节，阐述了一种阈值负偏移现象。这一现象出现这样的情况下：一个单元处于抑制编程操作，而其word line方向上的邻近单元处于编程操作下。其阈值负偏量会随着如下情况增加：更高的编程电压，更低的场强高度（FH），邻近cell更慢的编程速度（相邻的cell为slow cell），本身阈值处于高阈值区间。实验结果表明这一现象的原理是由于邻近cell的FN电流碰撞电离产生了空穴，而空穴进入了当前cell的浮栅中。

图6-63对NAND Flash的可靠性和物理现象进行了总结。可以看出，氧化层中的载流子陷阱，去捕获效应以及SILC是芯片稳定性退化的最主要的根本原因。

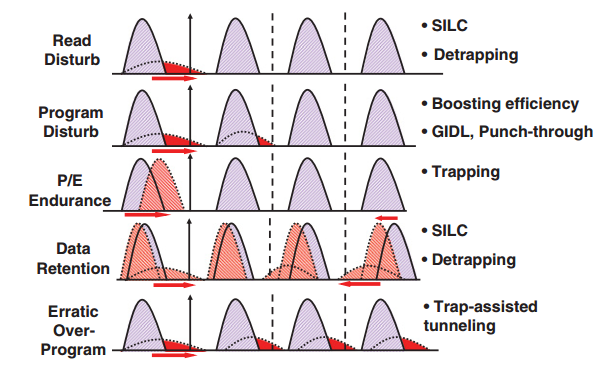


图6-63 可靠性总结

P/E循环次数和数据保持时间，两者是一种折中关系，如图6-64所示。即使使用系统解决方案，P/E循环和数据保留的未来目标也将受到影响，即P/E循环<1K，数据保留<1年。

如图6-65所示，性能与可靠性也是一种折中关系。如果需要高速编程，可靠性（如P/E循环）将降低，因为在编程过程中，较高的电场会作用于存储单元中的隧道氧化物。另一方面，如果某些应用程序需要高可靠性，例如大于10K的P/E循环，大于3年的数据保持时间，则诸如高速编程的性能会受到影响。因此，NAND闪存可靠性的目标规范将大大细分到每个应用程序，如存储卡、消费类应用程序（智能手机、平板电脑等）、高端应用程序（企业服务器SSD）等。

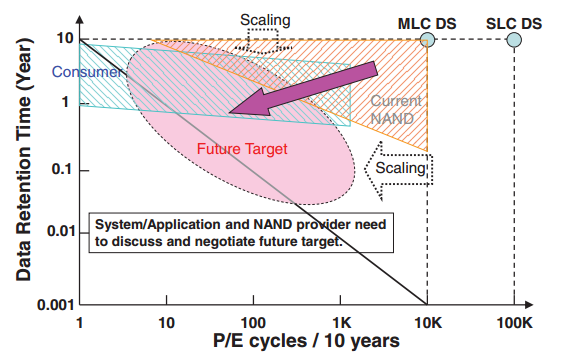


图6-64 数据保留与程序/擦除循环的展望

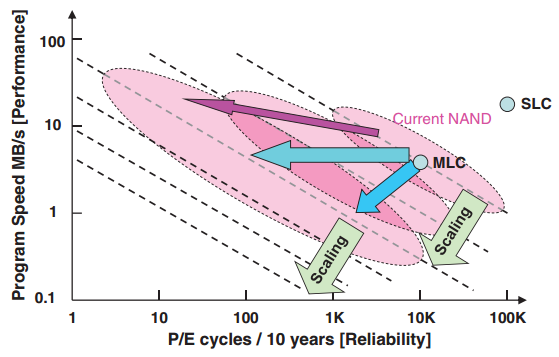


图6-65性能与可靠性的展望