

# 衬底负偏置对 FLASH 器件耐久性退化的影响\*

石 凯 许铭真† 谭长华

(北京大学微电子学系, 北京 100871)

摘要: 研究了 ETOX™ 结构 FLASH memory 单元器件在  $V_{FG} \approx V_D/2$  的热载流子写入应力条件下, 衬底负偏置对单元器件耐久性退化的影响. 结果表明: 在既定的栅、漏偏置条件下, 随着衬底负偏置的增加, 器件耐久性退化会出现极小值. 综合考虑了器件耐久性退化以及写入效率两方面的要求以后, 确定了在  $V_{FG} \approx V_D/2$  热载流子写入应力模式下, FLASH memory 单元器件具有增强写入效率以及最小耐久性退化的最佳衬底负偏置条件.

关键词: FLASH memory; 热载流子; 耐久性; 碰撞电离

PACC: 7325; 7340 EEACC: 2550; 2560

中图分类号: TN386 文献标识码: A 文章编号: 0253-4177(2006)06-1115-05

## 1 引言

1988 年 Intel 公司提出了经典的 ETOX™ 结构 FLASH 存储器, 现今大部分 FLASH 新结构都是它的基础上发展而来的. 在 FLASH 的生产设计中, 沟道热电子注入(channel hot electron injection, CHEI) 是一种被广泛使用的电荷注入(写入)方式. 早期, 通过在控制栅与漏端施加高正偏压, 使得沟道电子被加热, 获得足以越过 Si-SiO<sub>2</sub> 界面势垒(3.15eV) 的能量, 注入浮栅实现写入<sup>[1~3]</sup>. 然而, 这种电荷注入(写入) 方式有三个主要缺点: (1) 应力电压偏高, 以致于需要外围电路提供额外的高电压; (2) 注入(写入) 电流密度较高, 使得隧道氧化层在应力过程中的退化严重; (3) 写入电荷量不易控制. 针对上述问题, 设计研究人员从结构(如: HIMOS<sup>[4]</sup>、non-planar cell<sup>[5]</sup> 等)、写操作应力模式(如: CHISEL、BBHE<sup>[6]</sup> 等) 方面进行了努力. Bude 等人<sup>[7~11]</sup> 通过 Monte Carlo 模拟和实验验证提出, ETOX™ 结构 FLASH memory 单元器件在衬底负偏置热载流子写入应力条件下, 由于“次级电子(channel initiated secondary electron, CHISEL)” 的参与使得写入效率大幅度提高, 因此缩短了写入时间  $t_p$ , 并提高了器件的耐久性. 文献[12] 的研究工作更为系统地从事件操作性能和可靠性方面将 CHISEL 与 CHE 进行了比较, 得出随衬底负偏置的增加, 写入效率提高, 并且热电子注入范围和应力诱生的  $N_{it}$  分布从沟道漏端向沟道中心扩展, 同时  $N_{it}$  量有所下降. 说明 CHISEL 引起的退化程度要轻于 CHE.

然而上述研究所采用的衬底负偏置电压大都在 1.5V 以上, 并且分析侧重于 CHISEL 与 CHE 之间的比较. 本文将在此基础上, 对 CHISEL 写应力条件下衬底负偏置对 FLASH 器件耐久性退化的影响做进一步分析研究.

## 2 实验

### 2.1 测量样品

ETOX™ 结构 n 沟 FLASH 单元器件, 控制栅与浮栅之间等效氧化层厚度约 30nm, 隧道氧化层厚度 9nm, 沟道长为 1 $\mu$ m, 宽长比为 8:1. 相应的耦合系数为 0.4~0.5.

### 2.2 实验设备与测量设置

实验设备为计算机程控 HP4156B 精密半导体参数分析仪 (precision semiconductor parameter analyzer), HP41501B (SMU and Pulse Generator Expander) SMU/PGU 扩展器和 HP16440A (SMU/Pulse Generator Selector) 电子通道 (SMU / PGU) 转换开关, 在室温 (26℃) 无光照条件下进行测量.

ETOX™ 结构 n 沟 FLASH 单元器件耐久性实验:

耐久性实验采用周期性的擦写交替应力, 每个擦/写操作后进行常规特性(包括输出特性、转移特性) 的测量. 从转移特性实验数据中提取阈值电压  $V_{TH}$  和跨导  $g_m$ , 作为监测器件退化情况的主要参量. 器件采用漏端沟道热电子写入应力模式(如图 1(a) 所示); 控制栅  $V_{CG} = 16.0V$  (相应  $V_{FG} = 3.5V$ ),

\* 国家重点基础研究专项基金资助项目(批准号: TG2000-036503)

† 通信作者. Email: mxu@pku.edu.cn

2005-09-06 收到, 2005-10-28 定稿

源端接地,漏端  $V_D = 7.0\text{V}$ ,衬底偏压分别采用  $0, -0.5, -1.0, -1.5$  和  $-2.0\text{V}$ ,以获得衬底负偏置对器件耐久性退化的影响,写应力脉冲宽度  $T_P =$

$100\text{ms}$ ;源端 FN 擦除应力模式(如图 1(b)所示):控制栅  $V_{CG} = -6.0\text{V}$ ,源端  $V_S = 8.0\text{V}$ ,漏端浮起,衬底接地,擦应力脉冲宽度  $T_E = 1\text{ms}$ .

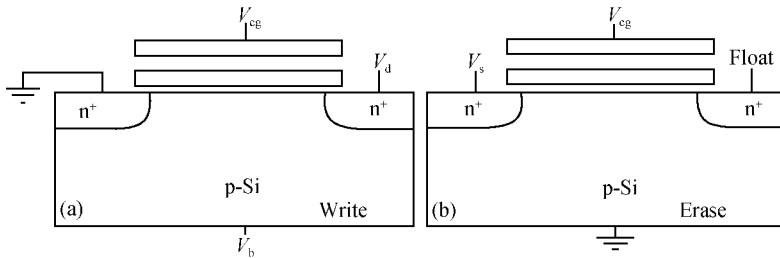


图 1 FLASH 单元器件擦写模式 (a)写入模式,  $V_{GG} = 16.0\text{V}$ ,  $V_D = 7.0\text{V}$ ,  $V_S = \text{GND}$ ,  $V_B = 0 \sim -2.0\text{V}$ ; (b)擦除模式,  $V_{CG} = -6.0\text{V}$ ,  $V_D = \text{FLOAT}$ ,  $V_S = 8.0\text{V}$ ,  $V_B = \text{GND}$

Fig.1 Erasing and writing stress mode of flash memory cell (a) Writing stress mode,  $V_{CG} = 16.0\text{V}$ ,  $V_D = 7.0\text{V}$ ,  $V_S = \text{GND}$ ,  $V_B = 0 \sim -2.0\text{V}$ ; (b) Erasing stress mode,  $V_{CG} = -6.0\text{V}$ ,  $V_D = \text{FLOAT}$ ,  $V_S = 8.0\text{V}$ ,  $V_B = \text{GND}$

### 3 结果与讨论

普遍认为隧道氧化层缺陷的产生,以及载流子的陷落与解陷是 FLASH EEPROM 器件耐久性(阈值窗口)退化的主要原因.而且隧道氧化层损伤集中在漏端附近,主要是在热载流子写入应力过程中,由碰撞电离产生的热载流子注入导致的.

当沟道电子在源漏横向电场  $E_{LAT}$  的作用下从源向漏移动,其中一部分在漏端附近的高电场中得到足够的动能,越过 Si-SiO<sub>2</sub> 界面势垒 ( $3.15\text{eV}$ ) 进入氧化层;另一部分在漏端附近引起碰撞电离,产生初级电子-空穴对.这些初级电子的一部分注入氧化层,成为栅电流的一部分;其余部分被漏端收集,形成漏极电流.碰撞电离产生的初级空穴向衬底漂移,一部分引起二次碰撞电离并再次产生电子-空穴对.这样栅电流成分中将包含沟道热电子以及初次和二次碰撞电离产生的初级和次级电子,栅电流的大小关系到热载流子对栅氧化层的损伤程度,以及 FLASH 器件的写入效率;衬底电流的成分主要是碰撞电离产生的空穴流.因而衬底电流和栅电流,是两个最基本的监控热载流子效应对 FLASH 器件特性影响的参数.图 2 给出 FLASH 热载流子写入应力模式下的载流子行为.当  $V_B < 0$ ,  $|V_B|$  不太大时,一次热电子从源到漏的漂移过程中,会发生碰撞电离(一次碰撞电离)并产生电子-空穴对,其中的空穴被衬底收集形成了衬底电流  $I_B$ .因而可以用  $I_B/I_D$  来近似反映每个沟道电子发生碰撞电离的机会.由于碰撞电离的发生与热载流子注入栅氧化层密切相关,因而碰撞电离与栅氧化层损伤必然有着密切的关系,阈值窗口退化率( $\Delta V_{TW}/V_{TW0}$ )是栅氧化层损

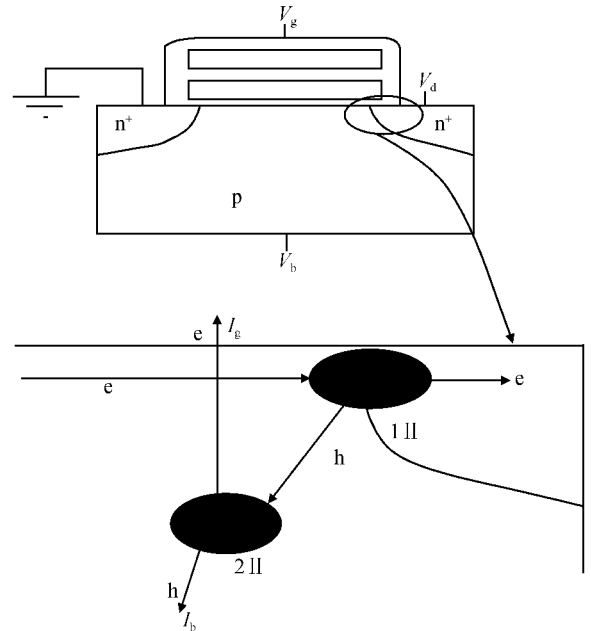


图 2 FLASH 器件热载流子写入操作中沟道电子的行为  
Fig.2 Behavior of CHE in FLASH memory cell writing operation

伤的重要表征量.众所周知,长沟器件的衬底电流可由以下公式描述<sup>[13]</sup>:

$$I_B = I_D \alpha \Delta L \quad (1)$$

$$\alpha \propto \exp\left(-\frac{l B_i}{V_D - V_{Dsat} + |V_{BS}|}\right) \quad (2)$$

$$\Delta L \propto \sqrt{V_D + \phi_{bi} + |V_{BS}|} \quad (3)$$

其中  $\alpha$  为碰撞电离率,它与碰撞电离化长度  $l$ ,碰撞电离常数  $B_i$ ,饱和漏电压  $V_{Dsat}$  和衬底偏置电压  $|V_{BS}|$  相关; $\Delta L$  为漏端耗尽区宽度,与内建势  $V_{bi}$  和衬底偏置电压相关.由(3)式当  $V_{BS} < 0$ ,  $V_{FG}$ ,  $V_D$  偏

置电压固定时,  $\Delta L$  随  $|V_{BS}|$  的增大而增大, 使得碰撞电离发生的空间范围扩展, 相应地增加了碰撞电离发生的机会. 又由于阈值电压  $V_{TH}$  随  $|V_{BS}|$  增大而增大, 而且  $V_{Dsat}$  随  $|V_{BS}|$  增大而减小. 如下式表述:

$$V_{Dsat} = V_{FG} - V_{TH}, \quad V_{FG} > V_{TH}^{[14]} \quad (4)$$

$$V_{TH} = V_{fb} + \phi_{Si} + \gamma \sqrt{\phi_{Si} + V_{SB}}, \quad \phi_{Si} = 2\phi_B^{[15]} \quad (5)$$

所以, 随  $|V_{BS}|$  的增加, 碰撞电离率  $\alpha$  指数增加.

对于长沟器件, 在饱和区  $I_D$  可以表述为<sup>[15]</sup>:

$$I_D = \frac{\beta}{2} (V_{FG} - V_{TH})^2 \quad (6)$$

所以, 当阈值电压  $V_{TH}$  随  $|V_{BS}|$  的增加而增加时,  $I_D$  将减小.

将(2)~(6)式代入(1)式, 得到:

$$I_B = C (V_{FG} - V_{fb} - \phi_{Si} - \gamma \sqrt{\phi_{Si} - V_{BS}})^2 \times \sqrt{V_D + \phi_{bi} - V_{BS}} \times \exp\left(-\frac{IB_i}{V_D - V_{FG} - V_{BS} + V_{fb} + \phi_{Si} + \gamma \sqrt{\phi_{Si} - V_{BS}}}\right) \quad (7)$$

在(7)式中,  $C$  为与工艺相关的常数. 综上所述, 当  $V_{BS} < 0$  时, 随着  $|V_{BS}|$  的增加  $I_B - V_B$  曲线必定有极值. 这与图 3 给出的实验结果是一致的, 图 3 中的曲线是用(7)式计算得到的模拟结果.

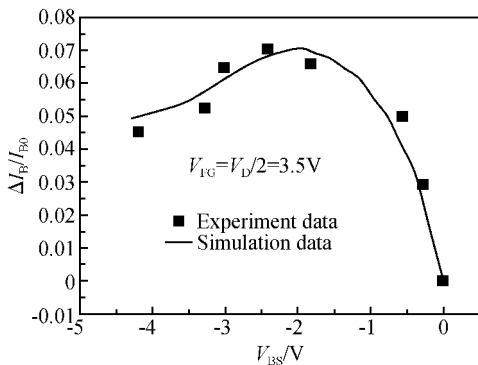


图 3 FLASH memory 在  $V_{FG} = V_D/2$  应力模式下的归一化  $I_B - V_B$  曲线

Fig.3 Unitary  $I_B - V_B$  curve of FLASH memory under  $V_{FG} = V_D/2$  stress mode

此外, 随着写入应力中衬底负偏置电压的加大, 一次碰撞电离发生的范围( $\Delta L$ )展宽, 相应的热电子注入范围也随之扩展. 与此同时, 由于漏-衬底 pn 结反向偏置的增加导致纵向电场的增强, 使得初次碰撞电离产生的热空穴在纵向电场的作用下, 向衬底漂移成为衬底电流的主要成分, 而其中的一部分会激发二级碰撞电离, 并产生新的电子-空穴对. 二级碰撞电离产生的电子在纵向电场的作用下向沟道区注入, 它的数目虽少, 但在电场作用下都获得了较高的能量. 因而随着衬底负偏置电压的加大, 热电子注

入范围不再局限于漏端附近, 而是逐渐向沟道扩展. 依据文献[7~11]的模拟结果, 当  $V_B = 0$  时(或  $|V_B|$  比较小时), 沟道中热电子主要分布在靠近漏端附近, 这将会使热电子在漏端附近的狭窄区域集中注入栅氧化层, 导致局部较高的注入电流密度, 使栅氧化层损伤集中在一个较小的范围, 这样会造成相对更为严重的局部退化; 而在  $V_B < 0$  时, 随  $|V_B|$  的增大, 热电子分布从漏端向沟道扩展, 局部热电子密度相对降低, 进而降低了注入电流密度, 使局部退化相对减轻. 如果  $|V_B|$  继续增大, 将会有更多电子注入栅氧化层, 加剧器件退化. 因而  $V_B < 0$  时, 随  $|V_B|$  增大, 阈值窗口的退化不会是单调变化, 而应是先减小后增大.

图 4 给出了  $V_{FG} = V_D/2$  应力模式下, FLASH 器件在擦写操作 500 次后, 阈值窗口退化率( $\Delta V_{TW}/V_{TW0}$ )与写应力衬底偏置条件之间的关系. 从图 4 的实验结果中不难发现, 在  $V_B < 0$  时, 随  $|V_B|$  的增加, 退化率先下降后增加, 曲线整体呈现 U 形, 而且  $V_B = -1.5V$  时, 阈值窗口退化率( $\Delta V_{TW}/V_{TW0}$ )最小. 此实验结果证实了上述分析.

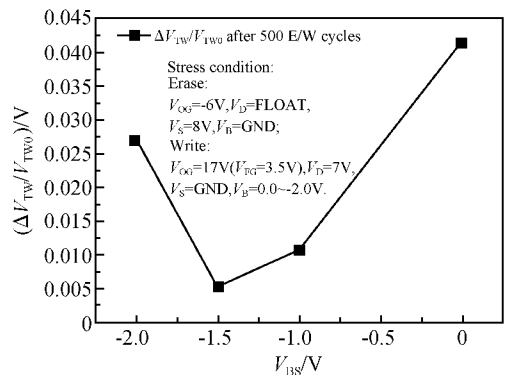


图 4 FLASH memory 器件阈值窗口退化率与写入应力衬底偏置条件之间的关系

Fig.4 Relationship between  $\Delta V_{TW}/V_{TW0}$  and  $V_{BS}$

图 5 插图给出跨导相对退化量  $\Delta g_m(t)/g_{m0}$  与应力时间  $t$  ( $t = \text{cycle} \times \text{period}$ ) 在双对数坐标下呈现的近似线性关系, 说明二者间存在幂指数关系:

$$\Delta g_m(t)/g_{m0} = (g_{m0} - g_m(t))/g_{m0} \propto t^\alpha \quad (9)$$

其中  $\alpha$  为跨导退化时间因子. 图 5 给出了跨导退化时间因子  $\alpha$  与衬底负偏置的关系. 它与图 4 中阈值窗口退化率与衬底负偏置的关系类似, 也为“U”型关系.

器件跨导  $g_m$  正比于沟道中载流子迁移率  $\mu^{[16]}$ :

$$g_m = \frac{\partial I_d}{\partial V_g} \Big|_{V_d = \text{const}} = C_i V_d \times \frac{W}{L} \mu \quad (10)$$

当界面陷阱密度远大于氧化层体陷阱密度时(本文所采用的测量样品, 其界面陷阱密度高达  $10^{12}$

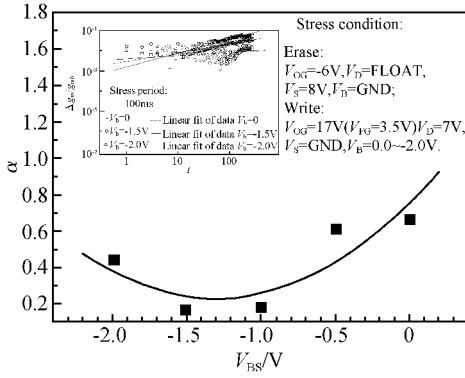


图 5 FLASH memory 器件在写入模式下的跨导退化时间因子  $\alpha$  与衬底负偏置电压  $V_{BS}$  的关系。插图为跨导相对退化量  $\Delta g_m(t)/g_{m0}$  与应力时间  $t(t = \text{cycle} \times \text{period})$  的幂指数图, 其中直线为线性拟合结果。

Fig.5 Relationship between the time factor  $\alpha$  of  $g_m(t)$  and  $V_{BS}$ . The inset shows the relationship between  $\Delta g_m(t)/g_{m0}$  and stress time  $t(t = \text{cycle} \times \text{period})$  in dual-log plot, the line is the simulation result.

$\text{cm}^{-2}$ ), 迁移率  $\mu$  可以近似地表示为<sup>[17]</sup>:

$$\mu(t) = \frac{\mu_0}{1 + KN_{it}(t)} \quad (11)$$

因而跨导  $g_m$  的退化与界面陷阱密度  $N_{it}$  的关系可以表示为:

$$g_m(t) = C_i V_d \times \frac{W}{L} \times \left( \frac{\mu_0}{1 + KN_{it}(t)} \right) \quad (12)$$

将(9)式改写为  $(g_{m0} - g_m(t))/g_{m0} = At^\alpha$  (其中  $A$  为拟合参数)可以得到  $g_m(t) = g_{m0}(1 - At^\alpha)$ , 代入(12)式得到:

$$N_{it}(t) = \frac{At^\alpha}{K(1 - At^\alpha)} \quad (13)$$

(13)式表明, 跨导退化时间因子  $\alpha$  的大小与该应力条件下界面陷阱的产生速率, 以及器件的退化快慢相关。

根据以上分析, 当衬底负偏置不太大时(如:  $|V_B| < 1.5V$ ), 由于热载流子注入范围的扩展, 随衬底负偏置电压的逐渐增大, 局部注入电流密度降低. 这将使得栅氧化层的损伤范围增大, 使界面陷阱分布于更大的范围, 因此界面陷阱密度相对降低. 这会使得局部阈值退化减小, 图 5 给出  $\alpha$  随衬底负偏置的增加而逐渐减小, 使得 FLASH memory 器件的耐久性退化减缓. 然而, 随着衬底负偏置电压进一步增加, 由于热载流子注入的栅电流密度增加, 加速了器件退化, 使得  $\alpha$  增加. 图 6(a)和(b)分别给出了 FLASH memory 器件阈值窗口退化率  $\Delta V_{TW}/V_{TW0}$  与跨导退化时间因子  $\alpha$  的关系, 以及器件阈值窗口退化率  $\Delta V_{TW}/V_{TW0}$  与界面态密度  $N_{it}$  随应力时间的增长率  $dN_{it}/dt$  的关系, 均显示出良好的线性关系.

这说明在衬底负偏置条件下, FLASH 器件的

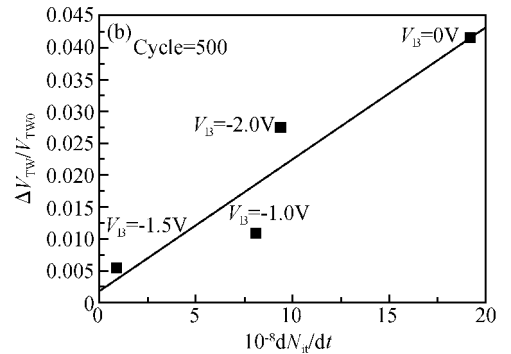
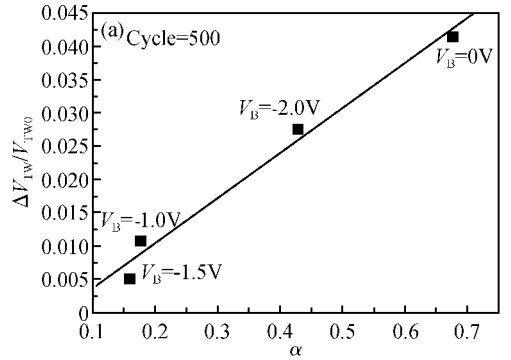


图 6 (a) FLASH memory 器件阈值窗口退化率与跨导退化时间因子  $\alpha$  的关系; (b) FLASH memory 器件阈值窗口退化率  $\Delta V_{TW}/V_{TW0}$  与界面态密度随应力时间的增长率  $dN_{it}/dt$  的关系

Fig.6 (a) Relationship between  $\Delta V_{TW}/V_{TW0}$  and  $\alpha$ ; (b) Relationship between  $\Delta V_{TW}/V_{TW0}$  and  $dN_{it}/dt$

耐久性退化不仅与氧化层体陷阱相关<sup>[18,19]</sup>, 其早期失效主要与隧道氧化层/硅界面陷阱相关。

依据上述实验及分析结果, 可以知道: 对于采用衬底负偏置的热载流子写入模式的 ETOX<sup>TM</sup> FLASH memory 器件, 存在着某种优化的写入应力条件, 它在既定的栅、漏偏置条件下, 通过调整衬底偏压, 在保持较高编程效率的同时, 尽可能将器件的耐久性退化降至最低. 对于本文所测量的样品在  $V_{FG} = V_D/2 = 3.5V$  的栅、漏偏置条件下,  $V_B = -1.5V$  为其最佳衬底负偏置条件。

### 4 结论

本文研究了 FLASH 器件耐久性退化率与写入应力中衬底负偏置之间的关系, 从理论和实验上分析了衬底负偏置对 FLASH 器件耐久性退化的影响. 研究表明, 在 FLASH 器件热载流子写入应力模式下施加衬底负偏置使得热载流子注入范围扩大和注入栅电流密度增加, 进而造成隧道氧化层损伤范围以及损伤程度的变化, 引起器件耐久性退化(包括阈值窗口退化率、跨导变化率等)随衬底负偏

置的非单调变化. 因此, 通过对衬底负偏置的选择, 可获得在热载流子写入模式下最佳的应力条件. 它在既定的栅、漏偏置条件下, 通过调整衬底偏压达到, 在保持较高编程效率的同时, 尽可能将器件的耐久性退化降至最低. 对于本文所测量的样品在  $V_{FG} = V_D/2 = 3.5V$  栅、漏偏置条件下,  $V_B = -1.5V$  为其最佳衬底负偏置条件.

致谢 感谢 Motorola 公司为本工作提供了测试设备, 感谢无锡华晶公司为本工作提供了测试样片.

## 参考文献

- [1] Bez R, Camerlenghi E, Modelli A, et al. Introduction to flash memory. Proceedings of the IEEE, 2003, 91(4): 489
- [2] Aritome S, Shirota R, Hemink G, et al. Reliability issues of flash memory cells. Proceedings of the IEEE, 1993, 81(5): 776
- [3] Pavan P, Bez R, Olivo P, et al. Flash memory cells - an overview. Proceedings of the IEEE, 1997, 85(8): 1248
- [4] Van Houdt J. HIMOS—a high efficiency flash EEPROM cell for embedded memory applications. IEEE Trans Electron Devices, 1993, 40(12): 2255
- [5] Ou Wen, Li Ming, Qian He. A novel non-planar cell structure for flash memory. Chinese Journal of Semiconductors, 2002, 23(11): 1158
- [6] Pan Liyang, Zhu Jun, Liu Zhihong, et al. A novel flash memory using band-to-band tunneling induced hot electron injection to program. Chinese Journal of Semiconductors, 2002, 23(7): 690
- [7] Bude J. Gate current by impact ionization feedback in sub-micron MOSFET technologies. Symposium on VLSI Technology Digest of Technical, 1995, 9A-2: 101
- [8] Bude J D. Monte Carlo simulations of impact ionization feedback in MOSFET structures. VLSI Design, 1998, 8: 13
- [9] Bude J, Pinto M R, Smith R K. Monte carlo simulation of the CHISEL flash memory cell. IEEE Trans Electron Devices, 2000, 47(10): 1873
- [10] Mohapatra N R, Mahapatra S, Rao V R, et al. Effect of programming biases on the reliability of CHE and CHISEL flash EEPROMs. IEEE 41st Annual International Reliability Physics Symposium, Dallas, Texas, 2003: 518
- [11] Mahapatra S, Shukuri S, Bude J. CHISEL flash EEPROM-part I, II. IEEE Trans Electron Devices, 2002, 49(7): 1296
- [12] Driussi F, Esseni D, Selmi L. Performance, degradation monitors, and reliability of the CHISEL injection regime. IEEE Trans Device and Materials Reliability, 2004, 4(3): 327
- [13] Ong T C, Fazio A, Mielke N, et al. Erratic erase in EROX™ FLASH memory array. VLSI Technology Symposium, 1993: 83
- [14] Arora N. Translator: Zhang Xing, Li Yingxue, et al. MOS-FET models for VLSI circuit simulation theory and practice. Beijing: Science Press, 1999 (in Chinese) [艾罗拉 N 著, 张兴, 李映雪, 等译. 用于 VLSI 模拟的小尺寸 MOS 器件模型——理论与实践. 北京: 科学出版社, 1999]
- [15] Sze S M. Physics of semiconductor devices. New York: John Wiley, 1981
- [16] Marchand B, Cretu B, Ghibaudo G, et al. Secondary impact ionization and device aging in deep submicron MOS devices with various transistor architectures. Solid-State Electron, 2002, 46: 337
- [17] Chung J E, Ko Ping-Keung, Hu C. A model for hot-electron-induced MOSFET linear-current degradation based on mobility reduction due to interface-state generation. IEEE Trans Electron Devices, 1991, 38(6): 1370
- [18] Ielmini D, Spinelli A S, Lacaita A L, et al. Correlated defect generation in thin oxides and its impact on flash reliability. IEDM, 2002: 143
- [19] Peng J Z, Haddad S, Fang H, et al. Flash EPROM endurance simulation using physics-based models. IEDM, 1994: 295

## Effects of Reverse Substrate Bias on the Endurance Degradation of FLASH Memory Devices\*

Shi Kai, Xu Mingzhen<sup>†</sup>, and Tan Changhua

(Department of Microelectronics, Peking University, Beijing 100871, China)

**Abstract:** The effects of reverse substrate bias on the endurance degradation of ETOX™ FLASH memory devices under the stress mode  $V_{FG} \approx V_D/2$  are investigated. The results indicate that as the reverse substrate bias increases the injection efficiency, the endurance degradation of the device is minimized under certain reverse substrate bias. Taking both the device endurance degradation and the injection efficiency into account, a FLASH memory device with optimal reverse substrate bias under the stress mode  $V_{FG} \approx V_D/2$  is obtained with the minimum endurance degradation and the greatest injection efficiency.

**Key words:** FLASH memory; hot-carrier; endurance; impact ionization

**PACC:** 7325; 7340 **EEACC:** 2550; 2560

**Article ID:** 0253-4177(2006)06-1115-05

\* Project supported by the Special Funds of the National Key Basic Research Plan of China (No. TG2000-036503)

<sup>†</sup> Corresponding author. Email: mxu@pku.edu.cn

Received 6 September 2005, revised manuscript received 28 October 2005

©2006 Chinese Institute of Electronics