# NAND Flash 微缩挑战

## 读出窗口裕度（Read Window Margin）

在本节中讨论了超过2X到0X-nm代的NAND闪存的自对准STI单元(SA-STI单元)的读窗口裕度。通过FG-FG耦合干扰（浮栅电容耦合干扰）、电子注入扩散（EIS)、反向模式依赖（BPD)和随机电报噪声(RTN)等物理现象，研究了读窗口Margin。读窗口Margin的退化不仅是体现在编程后的cell Vt分布宽度变宽，有时也表现为擦除后cell状态的Vt变高，主要是由于较大的阵列间FG-FG耦合。然而，读窗口Margin在1Z-nm(10nm)生成中仍然是正的，Air gap制程使FG-FG耦合干扰减少了60%。因此，SA-STI cell有望能够缩小到1Z-nm(10nm)的一代，减少60%的FG-FG耦合干扰。

### 读窗口Margin推理

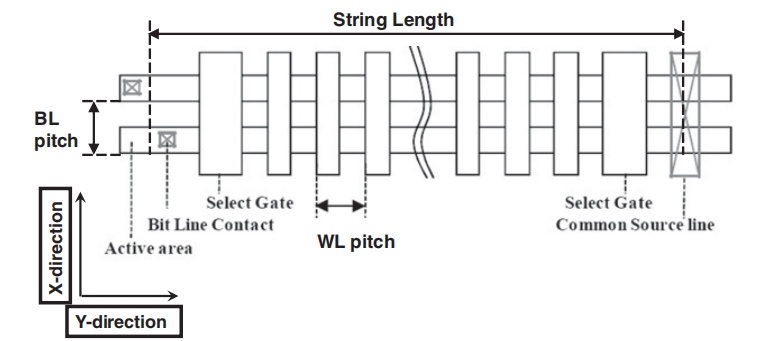


图1‑1 NAND单元格字符串的俯视图

图1‑1显示了传统的NAND单元格字符串的俯视图。为了研究NAND闪存单元的比例，我们假设单元尺寸超过2xnm(26nm)一代，如表5.1所示。给出了2Xnm的尺寸，位线(BL)半间距（half-pitch）为27nm，字线(WL)半间距 为26nm。假设超过2Xnm的尺寸，BL半间距的固定比例因子为×0.85，WL间距的固定比例因子为×0.8。同样的，假设沟道长度、宽度和poly间介电(inter-poly dielectric IPD)厚度分别降低×0.9和×0.95。

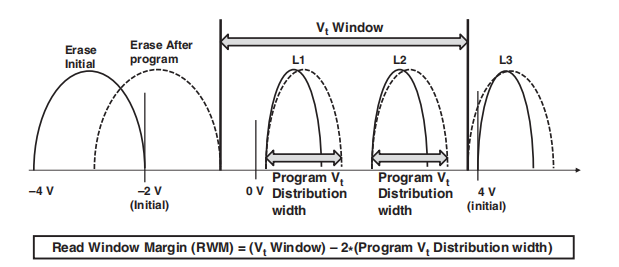


图1‑2 MLC NAND cell的Vt读出窗口

图1‑2显示了MLC（2位/单元）NAND单元中读取Vt窗口示意。“Vt窗口”可以定义为：擦除态cell阈值分布的最右边边界和L3（最高编程状态）的阈值分布最左边之间的窗口。 L1/L2的两个编程Vt分布必须在Vt窗口内，才能完成一个可靠的读取操作。读取窗口裕度(RWM)由RWM=(Vt窗口)−2∗(编程的Vt分布宽度)定义，因此RWM表示每个状态的Vt分布。

随着工艺微缩，一些物理现象越来越糟， 导致RWM已经被严重恶化。因此，为了进一步缩放NAND cell，分析和预测未来按比例缩小的NAND cell中的RWM是非常重要的。为了研究RWM，我们假设电子注入扩散(electron injection spread)、FG-FG耦合干扰、Random Telegraph Noise 和背景图形依赖(back pattern dependence)的物理现象的尺度趋势如下。以及对页面编程顺序、参数设置等的其他假设，也如下所示。

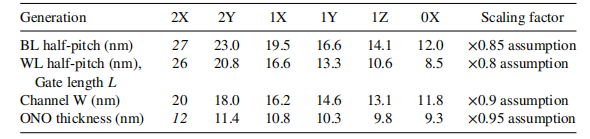


表 1-1 世代2X-0X纳米, cell尺寸和ONO(IPD)厚度

**RWM计算的假设**

1. 由于电子注入扩散(EIS)、FG-FG耦合干扰、RTN和背景图形依赖(BPD) 因素的叠加，假设Vt分布宽度（@±3σ）变得更宽。每个值给出2Xnm代，并用以下公式外推2Ynm到0Xnm代。
2. 电子注入扩散(EIS)是在编程操作过程中由于编程脉冲期间少量注入电子的统计扩散而引起的。

### 编程态Vt分布宽度

在常规程序操作中，通过使用ISPP和逐位验证操作，可以压缩编程的Vt分布宽度。初始编程的Vt分布宽度由ISPP\_step+EIS确定。然后，在所有页面都完成编程后，由于RTN、FG-FG耦合干扰和BPD使它变得更宽。

根据第5.2.1节的假设，计算了所有页面编程后的Vt分布宽度，如图1‑3所示。当存储单元从2Xnm缩小到0Xnm时，编程的Vt分布宽度从1320mV增加到2183mV。很明显，增加Vt分布宽度的主要原因是FG-FG耦合和RTN。

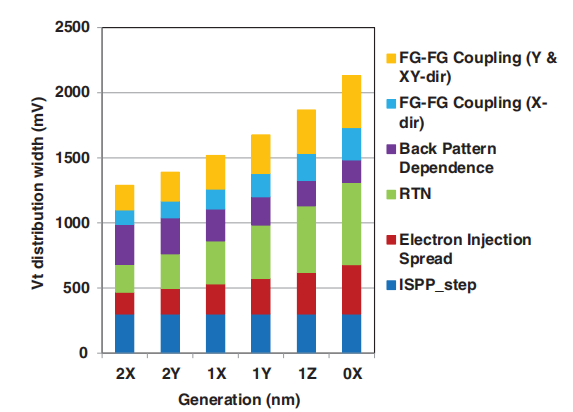


图1‑3 编程Vt分布宽度预测

为了获得适当的FG-FG耦合干扰的Vt位移值，我们推导出了相邻攻击单元（受目标单元影响）的deltaVt，如图1‑4所示。页面编程步骤也在图1‑4中描述了Vt分布。对于编程状态的FG-FG耦合，攻击单元的deltaVt被描述为dVt\_E\_L1或dVt\_LSB\_L2+dVt\_LSB\_L3，如图1‑4中MSB程序后的（3）所示。dVt\_E\_L1表示deltaVt从MSB程序之前的擦除状态（@2）转移到L1状态。采用较大的dVt\_E\_L1或dVt\_LSB\_L2+dVt\_LSB\_L3值计算FG-FG耦合Vt位移。

编程Vt分布(3v宽度)的cell通过ISPP编程算法上移。特定单元(单元A)停止编程，通过验证Vt\_cellA的阈值电压，而相邻单元(单元B)尚未通过验证Vt\_cellB的阈值电压。相邻单元(单元B)通过以下ISPP步骤编程，然后导致FG-FG耦合单元A的Vt差异(Vt\_cellA−Vt\_cellB)。在此假设下，假设Vt差分的分布(Vt\_cellA−Vt\_cellB)由Vt分布(3-V宽度)组成，则Vt位移的𝜎为SQRT（𝜎2+𝜎2）=𝜎∗SQRT（2）。同时，我们假设相同的FG-FG耦合Vt转变发生在L1和L2之间，通过假设使用更好的程序操作，如ABL并行程序方法[44]，BC状态第一程序算法[45]，和p3模式预脉冲方案[46]，减少FG-FG耦合L1和L2。

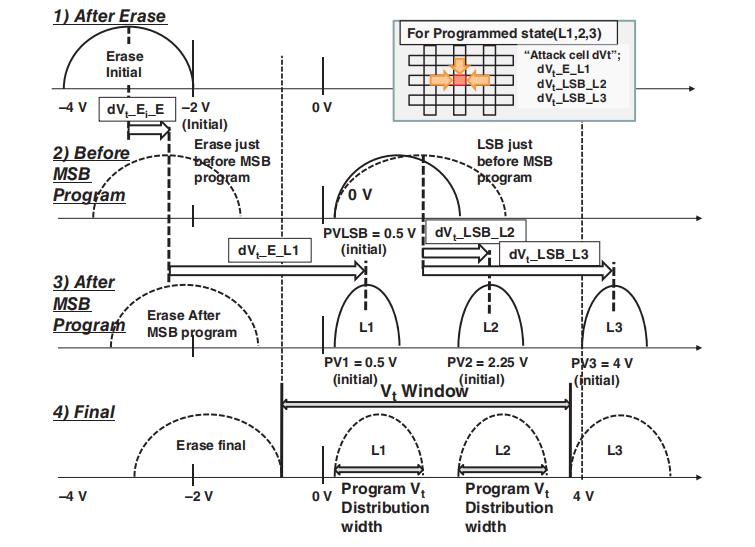


图1‑4 在页面编程步骤中的Vt分布

### Vt窗口

如图1‑2显示，“Vt窗口”可以定义为：擦除态cell阈值分布的最右边边界和L3（最高编程状态）的阈值分布最左边之间的窗口。

图1‑5显示了在Airgap或低k介质使FG-FG耦合减少0%、30%和60%。

在x方向加入STI Airgap 和y方向加入WL Airgap均可减少FG-FG耦合，最理想状态下， 在−60%FG-FG耦合减少的情况下，即使在1Z-nm一代中，Erase cell分布的右侧边缘也能保持在小于0V以内。这样的话，在1Z-nm一代中， Vt窗口可以保持在4000mV以上。

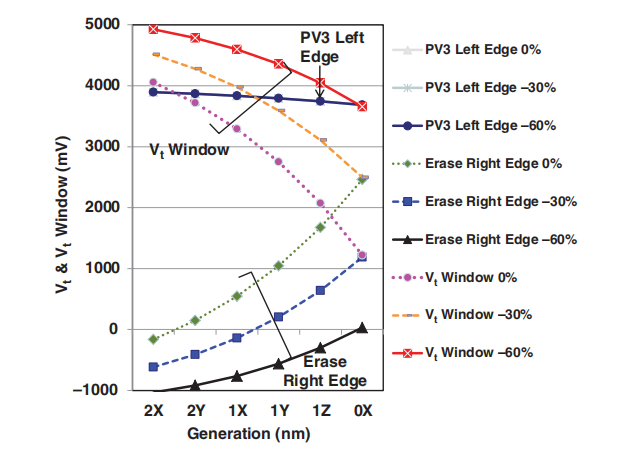


图1‑5 工艺微缩下的Vt窗口

### 读取窗口裕度(RWM)

图1‑6为RWM的比例趋势。RWM随着cell尺寸微缩而退化。在“Airgap”的情况下，1Xnm有临界的RWM，1y-nm世代有负的(−719mV)RWM。在“FG-FG耦合−30% Airgap”的情况下，1Ynm有临界RWM，1Z-nm一代的RWM为负。更进一步，在“FG-FG耦合−60% Airgap”的情况下，1Z-nm一代仍然具有正的RWM。这意味着实现1Y-nm 工艺下cell需要减少30%的FG-FG耦合，而1Z-nm工艺下cell需要减少50-60%。

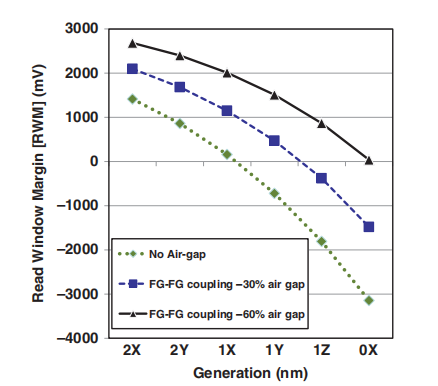


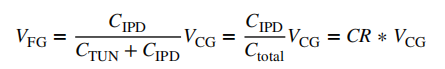
图1‑6 读取窗口裕度（在无airgap，FG-FG耦合减少了30%以及60%的情况下）

## 浮栅电容式耦合干扰

浮栅电容耦合干扰（FG-FG耦合）是缩小浮栅NAND闪存单元的主要限制问题，因为读窗边缘（RWM)主要由浮栅电容耦合干扰降低，如第5.2节所述。随着特征大小(F)的缩放，浮栅到浮栅的空间变得更小，从而通过8个相邻单元的Vt变化导致Vt位移。这个微缩问题导致了Vt分布宽度的扩大。

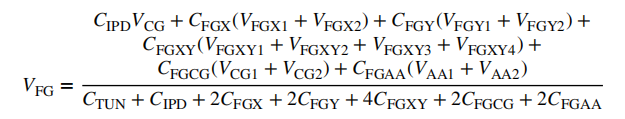
### 浮栅电容耦合干扰模型

在大cell尺寸的旧概念中，浮栅电压仅由控制栅电压以及CR=CIPD/Ctotal 决定，其中CIPD为控制栅与浮栅电容，Ctotal为浮动栅的总电容，可以表示为：



随着NAND闪存设计规则的缩小，浮栅周围的寄生电容器(CFGX、CFGY、CFGXY、CFGCG、CFGAA)相对变大，如图1‑7所示。他们不能被忽视。

此时，浮栅电压不仅由相应的控制栅电压决定，还由周围浮栅、控制栅和活动区域的电压决定，如下式所示



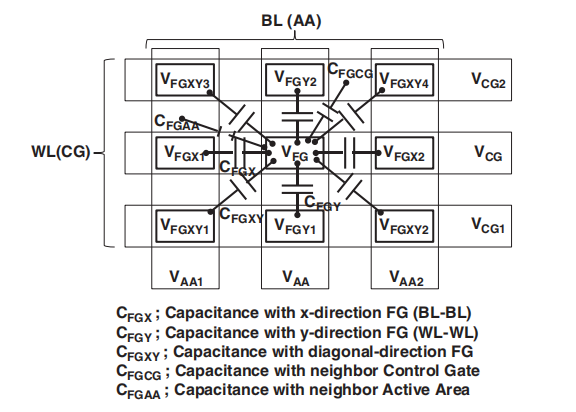


图1‑7 基于寄生电容耦合的浮栅电容耦合干扰模型

图1‑8显示了由单元技术节点缩小后，浮栅电容耦合干扰引起的vth位移的模拟结果。在三维TCAD模拟中，假设63nm的存储单元具有8nm的隧道氧化物厚度，15nm ONO厚度和85nm的浮栅高度。存储单元晶体管目前已经从63纳米缩小到20纳米， 随着cell尺寸的减小，场氧化物凹槽保持在+5nm，并调整掺杂浓度以防止cell晶体管punch through。随着技术节点的微缩，浮栅电容耦合干扰的vth位移显著增加。

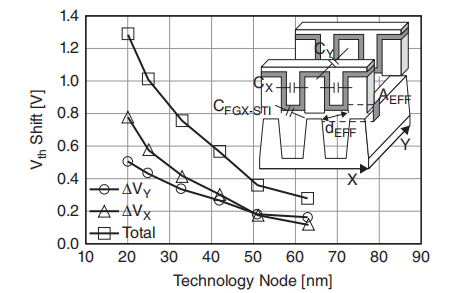


图1‑8 cell间干扰与cell节点微缩引起的Vth位移的模拟结果

图1‑9显示了浮栅电容性耦合干扰对技术节点微缩造成的影响。 浮栅电容耦合干扰占总Vt位移的百分比显著提高。 对于30nm节点，FG电容耦合干扰超过cell总Vt位移的30%， 20nm节点上干扰接近于cell总Vt位移的50%。

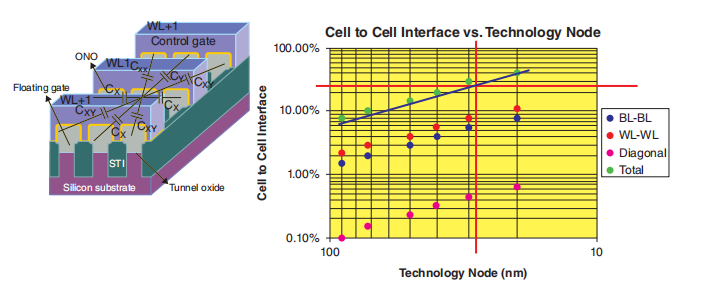


图1‑9 FG-FG耦合干扰

### 通道直接耦合

根据传统的浮栅电容耦合干扰理论，63nm技术节点的y方向(ΔVY)的Vt位移比x方向(ΔVX)的Vt位移更严重，如图1‑8所示，浮栅在y方向上直接面对面，在x方向浮栅被控制栅屏蔽。在工艺节点大小为50nm时，ΔVX超过了ΔVY，并且随着技术节点大小减小到20nm时，ΔVX急剧增加。

在50nm以下的技术节点中，单元晶体管的沟道边缘与相邻单元的浮栅之间的距离非常接近，因此相邻单元的浮栅电压直接影响沟道边缘，改变了沟道边缘上的电场分布。然后，Vth移位是由相邻单元的浮栅电势的直接场效应引起的。由于大约70%的cell电流在沟道边缘流动，因此cell晶体管的Vth主要取决于电场拥挤的条件和沟道边缘的掺杂浓度。因此，存储单元遭受了强烈的Vth位移，特别是在x方向上，其中浮栅极面对沟道边缘的整个表面。这意味着观察到的浮栅电容耦合干扰包括与相邻单元的沟道边缘耦合。

在45nm的cell中， cell间干扰的实验数据如图1‑10所示。随着场氧化物recess的减小，相邻单元对channel边缘的直接场效应增加，从而使Vth位移变大。图1‑10中有三条线，即常规浮栅电容耦合干扰的Vth位移（ΔVX-Indirect）、直接场效应的第V位移（ΔVX-Direct）和实验测量的Vth位移（ΔVX）。

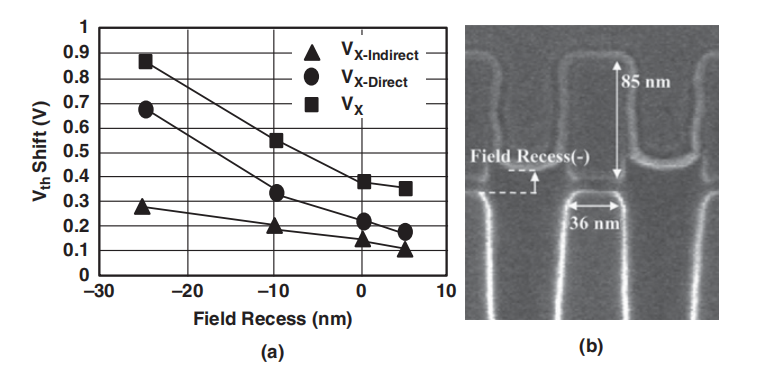


图1‑10 (a) Vth位移依赖于filed recess (b) 45nm节点尺寸的NAND flash横断面扫描电镜照片

### Source/Drain耦合

一种新的cell间浮栅诱导势垒增强(floating-gate induced barrier enhancement FIBE)干扰现象已在40nm节点以下的cell中被报道。与传统的浮栅之间的电容耦合不同，当干扰单元(interfering)的Vth超过一定的门限后， 被干扰单元(interfered)的阈值电压(Vth)位移会明显变大。这是由于通过源极/漏极和干扰单元的浮栅之间的电容耦合对源极和漏极区域的导带进行了调制。通过实验和仿真，验证了该模型的有效性。为了降低FIBE效应，在相邻的WL中，较高的掺杂量和较高的Vread方案可能是有效的。

图1‑11显示了WLs(y方向)之间的cell间干扰。受干扰cell的Vth在B区异常增加(较高的干扰cell Vth)，而受干扰cell的Vth与a区干扰cell的较低Vth呈线性依赖，显示了传统的浮栅电容耦合。在NAND闪存cell的工艺节点小于40nm后，干扰cell的Vt越高，这种B区Vt的位移越大。

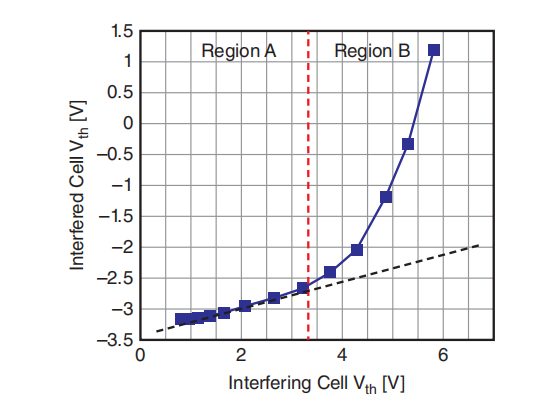


图1‑11 被干扰(interfered) cell Vth依赖于干扰(interfering) cell的Vth

图1‑12显示了在具有寄生电容和寄生电容+FIBE效应的情况下，模拟受干扰单元的导带分布。传统浮栅电容耦合的受干扰单元的FG电位变化似乎只增加了通道中心的导带，如图1‑12b所示。而干扰单元的FG电位变化增强了漏极区的导带，影响了被干扰单元的沟道导带，如图1‑12b所示。

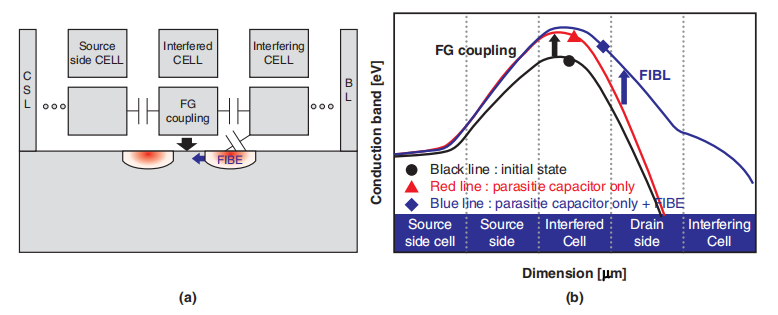


图1‑12 (a)NAND cell在WL方向上的寄生耦合电容的简化模型 (b)隧道氧化物下的导带分布

### Air Gap与Low-k 材料

介绍了利用低low k介质的gate spacer，诸如low-k氧化物和air gap填充可以改善浮栅电容耦合干扰。

图1‑13显示了在栅极之间形成Air gap工艺流程的一个例子。栅极成型后，淀积缓冲氧化物/氮化物（图1‑13 A，b部分），然后沉积氧化物以填充栅极空间（图1‑13 A，c部分）。在那之后，通过干蚀去除栅极Poly上的氧化物（图1‑13 A，d部分）。然后，为了在栅极之间形成wing，淀积并蚀刻SiN(图1‑13 A，f部分)。通过湿蚀刻去除spacer wing内的氧化物后（图1‑13 A，g部分），氧化物淀积形成栅间spacer内的air gap（图1‑13 A，h部分）。在图1‑13 B中，从制作器件的扫描电镜图像可以清楚地观察到air gap。

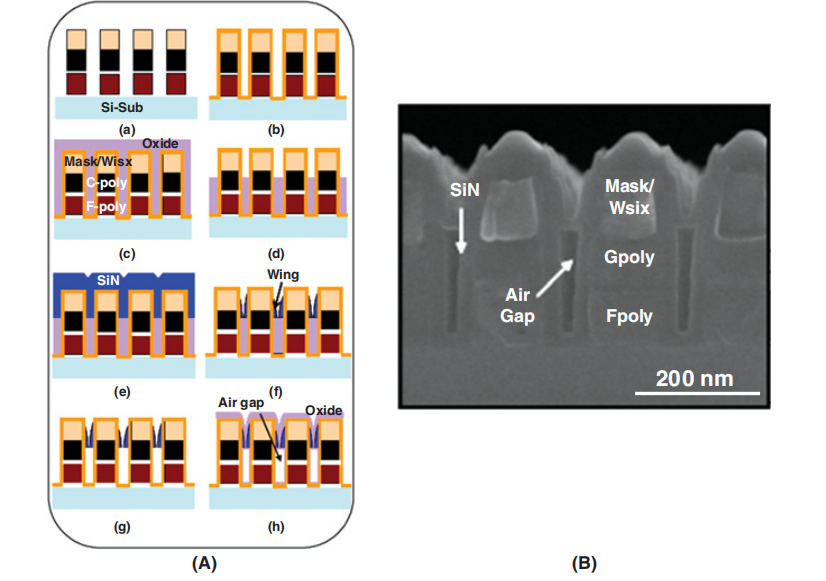


图1‑13 (A)Air Gap的工艺流程 (B)在一个90nm的cell中，Air gap的扫描电镜图像

图1‑14显示了在90nm cell中具有SiN、氧化物和Air gap作为栅极space材料， WL30和WL31的Vt分布。通过对同一字线(WL30)/奇数位线和WL31/偶数位线cell进行编程，统计WL30/偶数位线上的cell的Vt移动。当cell从Vth=−3V编程到Vth=1.5V时，三种栅极space材料分别为SiN、氧化物和Air gap的阈值电压变化分别为0.16V、0.07V和0.02V。Vth位移几乎对应于介电常数(SiN：氧化物：air=8：4：1)。Air gap减少的Vth位移是由于其浮动栅之间的寄生电容较低。图1‑14(d)比较了单脉冲程序下1Gbit单元的单元Vth分布。由于改进了浮栅电容耦合干扰，利用Air gap改善了cell 的Vth分布。

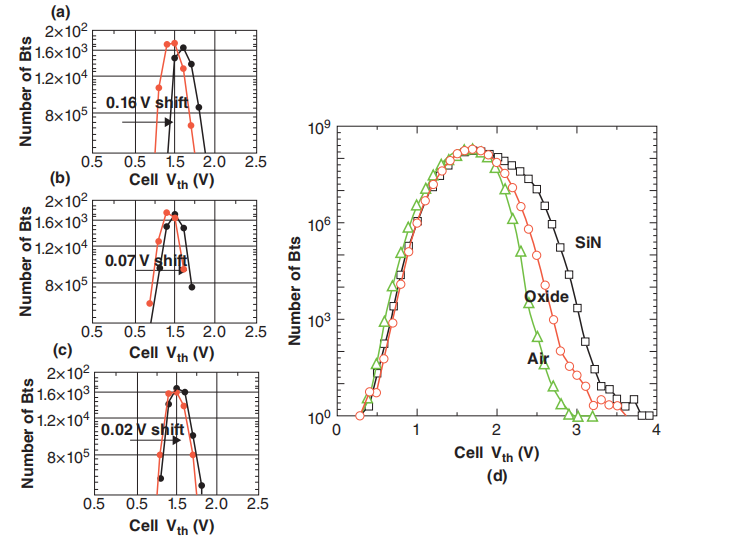


图1‑14在WL30上测量了浮栅干扰的阈值电压位移

1. SiN间隔物，(b)氧化物，(c)Air gap，(d)浮栅干扰造成的cell vth分布位移对比

图1‑15为1xnm cell的字线(WL) Air gap和STI Air gap结构的横截面透射电镜显微图。 该工艺成功地制备了WL和STI Air gap。从25nm一代产品开始使用WL Air gap，可以减少浮栅电容性耦合干扰的同时，WL Air gap也可以改善WL高场问题。STI Air gap不仅能改善浮栅电容耦合干扰，而且是非常有效的改善编程干扰。

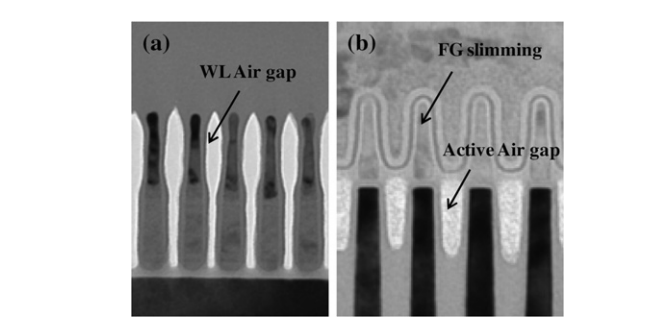


图1‑15 1xnm SA-STI NAND Flash cell中(a)字线Air gap和(b)STI Air gap的横断面TEM显微图

## 编程电子注入扩散

### 编程电子注入扩散理论

当微缩存储单元的大小时，会减少浮栅中存储的电子数量，如图1‑16所示。在1xnm存储单元中，存储的电子数接近100，对应于3V Vt位移。这意味着在一个编程脉冲中，只有10个电子被注入到浮动门中，每个脉冲之间有300毫伏的升压。少量的10个电子会对注入浮栅的电子数量产生很大的统计变化，从而使编程态的Vt分布宽度更宽。

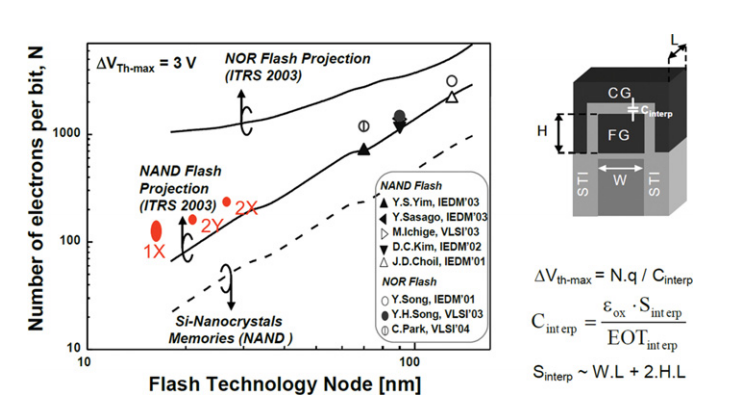


图1‑16 FG中存储电子的电子数，电子的数量随着存储单元大小的微缩而减少

据报道，在编程脉冲过程中，通过统计电子注入扩散，编程中的Vt分布宽度变宽。当在编程脉冲过程中注入少量电子时，电子注入过程由泊松统计量支配。这可以解释为在电子注入到浮栅之后的隧穿氧化物场的减少，然后降低了电子注入速率。 结果可以用蒙特卡罗模型来解释，该模型能够正确地描述编程操作背后的主要物理原理。

由于注射统计扩散，阈值分布扩大，导致一些cell vt移位超过Vstep。随着cell不断微缩，由于一个编程脉冲中电子数量的减少，使得NAND存储单元的注入统计扩散变大。

采用ramp 编程策略(ISPP：增量步进脉冲) 对ΔVt扩散进行了实验研究，如图1‑17所示。

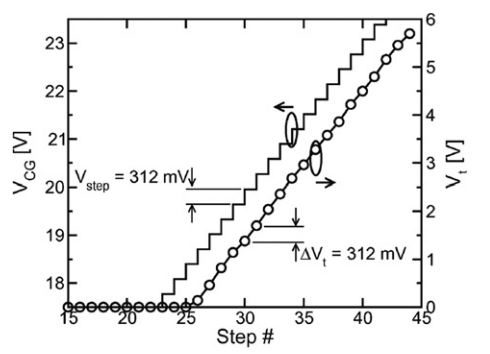


图1‑17 在60nm器件的控制栅加电压后，NAND cell被编程并产生的Vth位移。（注：在传统的NAND阵列中，只能sense到正的Vt值）

### 低掺杂浓度对FG的影响

在浮栅中，降低掺杂浓度，增强了电子注入扩散。在40纳米设计规则单元中有较低浮栅掺杂浓度的cell中观察到更大的ΔVt分布，如图1‑18所示，其中ΔVt意味着Vt从j转移到j+1加速编程脉冲[ΔVt≡Vt(j+1)−Vt(j)]。低FG掺杂的ΔVt分布比高掺杂的ΔVt分布更大，在低FG掺杂的较高ΔVt处观察到tail比特。

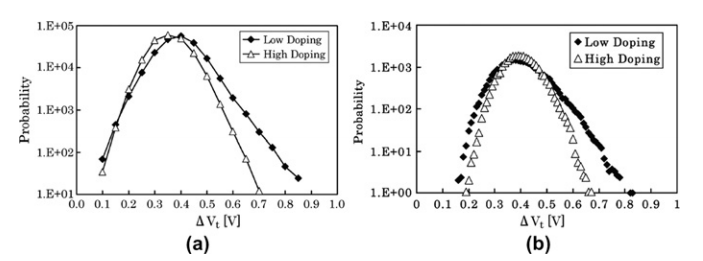
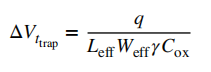


图1‑18 (a)从i到i+1阶梯编程脉冲（Vstep是400 mV）的cell ΔVt分布的测量结果。(b)计算了考虑FN隧穿统计量的影响，低和高FG磷掺杂的ΔVt分布

## 随机电报信号噪声（RTN）

### Flash Memory cell RTN

MOSFET中的随机电报噪声(RTN)是栅氧化物界面附近电荷捕获点的电子捕获和发射事件引起的漏极电流或阈值电压波动，如图1‑19所示。闪存单元中每个陷阱位置(ΔVttrap)的阈值电压波动幅度约为



式中，q为基本电荷，Leff和Weff分别为有效通道长度和宽度，Ƴ为控制栅与浮动栅的耦合比，以及Cox为栅电容。在浮栅闪存单元中，RTN的振幅通常比在CMOS逻辑器件中更大，这是因为Cox非常小，这是由于相对较厚的隧道氧化物(~10nm厚)。此外，在NAND闪存中，存储器单元的大小已经被大大缩小了；因此，L和W的尺寸，特别是W，比传统的CMOS逻辑器件要小得多。此外，由于漏电电流路径机制，RTN的振幅可以大于上述等式的预期。

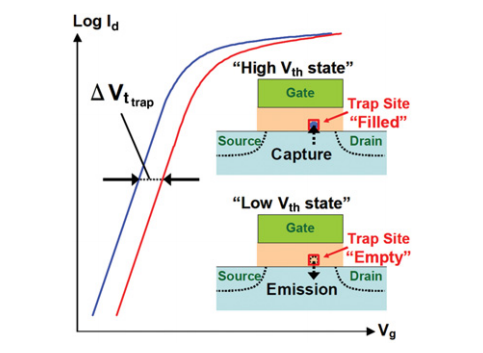


图1‑19 MOSFET中随机电报噪声(RTN)引起的阈值电压波动是由氧化物电子陷阱的电子捕获和发射事件引起的。

首次报道了在闪存中观察到由随机电报信号噪声(RTN)引起的阈值电压(Vth)波动。 图1‑20显示了在90nm cell中测量的RTN的一个例子。电流在RTN发生时与逻辑CMOS晶体管中显示相同的开关行为。因此，RTN是在缩放的NAND闪存中读取失败的一个潜在来源。

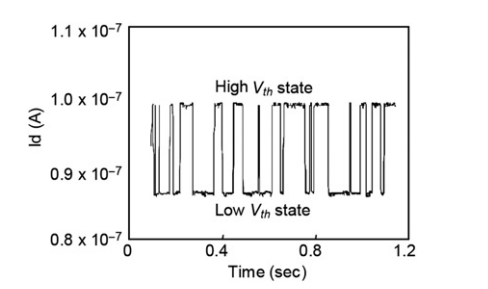


图1‑20 在90nm节点的闪存中，漏极电流的时间序列变化的一个例子

## CELL 结构挑战

随着工艺节点缩小， SA-STI cell的结构也带来挑战。SA-STI cell的关键结构是“CG formation margin”，即FGs 之间的CG的制造margin。图1‑21显示了超过2x-nm一代的FG slimming结构中CG制造margin的估计。在这个估计中，假设FG宽度和CG宽度相等。随着SA-STI cell尺寸的缩小，1X-nm cell的FG宽度和CG宽度都减小到小于10nm的宽度。在1Y-nm和1Z-nm代中，FG和CG的宽度必须控制在5nm左右，甚至ONO的厚度也要按每代×0.95的比例缩小。在编程和擦除过程中，FG和CG的耗尽(depletion)效应也必须被抑制。金属或硅化物材料将适用于未来的NAND cell中的FG和CG。

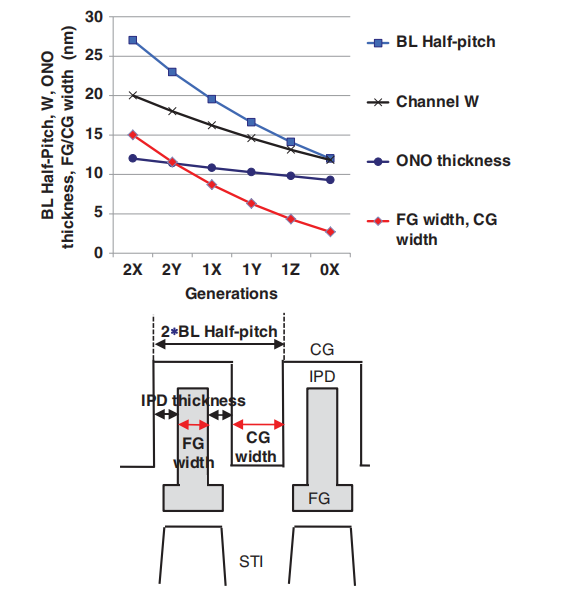


图1‑21 在FG之间的CG制造的估计margin

## 高电场限制

NAND闪存的编程和擦除电压高(~22V)，并且不能大幅降低，因为在编程和擦除过程中，FN隧道机制在隧道氧化物中需要较高的高电场(~10MV/cm)。据报道，由于编程字线(WL)与相邻的WL之间的高电场，产生了新的编程干扰现象。这种新的编程干扰是指在编程过程中相邻word line的cell 的Vth值减少。当存储单元尺寸微缩时，该编程对干扰变得更加严重，因为这种现象随着栅极间距间的减少而严重加剧。

图1‑22显示了新的编程干扰现象的测量条件。当WL(n)被编程为高Vt状态时，高编程电压(Vpgm)应用于WL(n)的控制栅，通过电压(Vpass)应用于其他控制栅，使得选定和未选择的string的单元达到编程禁止的目的。同样，NAND闪存从更接近CSL的较低字线（common source）编程到更接近BL的较高字线（bit line）。在相对较低的pass电压下，在30nm以下的存储单元中观察到这种新的编程干扰现象。

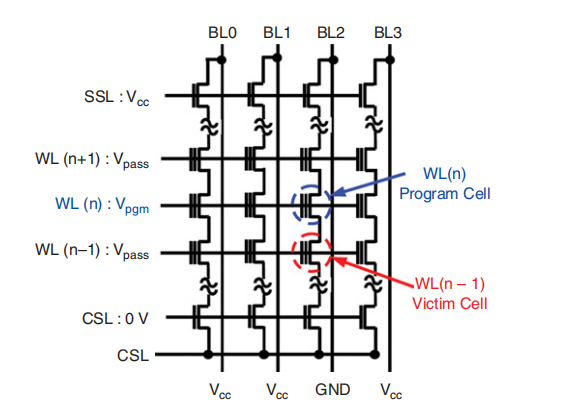


图1‑22 Self boost的基本编程条件示例

图1‑23显示了Vpgm=26V和Vpass=4.5V在WL(n) 编程stress加速条件下的受干扰的cell的Vt。在测量的40个cell中，随stress时间的增加，Vth明显减少。在MLC操作中，当所有字线都被编程到高状态时，在单元阵列中观察到Vth分布的tail bit，如图1‑24所示。Cell阵列的栅极设计规则在30nm以下。当WL(n)被编程时，图1‑24的tail bit在WL（n−1）处产生。String中最后一个word line没有upper word line, 这些单元没有Vth分布的tail bit。随着pass电压的增加，Vth分布的tail bit逐渐减小，如图1‑24所示。

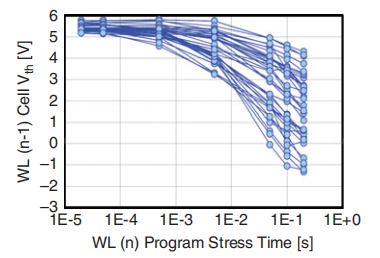


图1‑23 WL(n−1)中受干扰的40个cell的vth减少，编程stress 施加在40个cell。编程电压为26V，pass电压为4.5V。

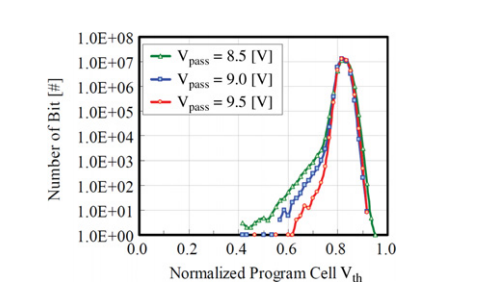


图1‑24 单元阵列编程后编程单元Vth的分布，其中Vpass分别为8.5V、9.0V和9.5V

图1‑25显示了30nm存储单元WLs的电场模拟结果。模拟是在三维结构中进行的，实际尺寸反映了在硅沟道、poly 浮栅和控制栅上的掺杂。通过控制浮动栅的电荷，从Id-Vg曲线调整浮栅的目标vt。在Vpgm=24V和Vpass=8.5V条件下，浮栅顶边缘和控制栅底边缘之间的最大电场为9.7MV/cm。确认边缘场足够大，可以在控制栅和相邻浮动栅之间产生FN隧道电流。

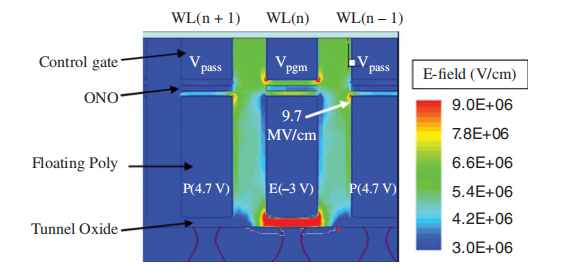


图1‑25 在t=0条件下，编程下的电场分布模拟结果，其中Vpgm为24V，Vpass为8.5V

在30nm到50nm的不同栅极设计规则下进行了评估。 图1‑26显示了这种现象的栅极spacer依赖性。相邻的WL栅极spacer对这一现象非常重要。虽然Vth的减少是根据栅极设计规则在不同的编程电压下测量的， 所有的测量结果都被简单地推广到控制栅之间的电场中。随电场的归一化结果如图1‑26b所示，其中y轴为编程应力持续0.2s后受干扰的cell Vth的减小。如图1‑25所示，WL边缘的电场易受栅极的形状和profile的影响。

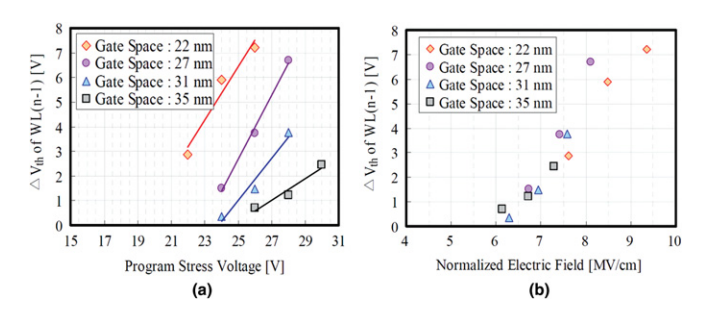


图1‑26 (a)不同栅space的样片在0.2s的编程应力后，WL(n−1)的Vth的下降值；通过电压为4.5V。(b)对于不同栅space的样片，会降低WL(n−1)的Vth的下降值。x轴为WL(n)和WL（n−1）的控制栅之间的电场。

在NAND闪存单元中，最严重的高WL问题发生在选择字线(WL)和邻居WL之间。在2X-nm cell中，所选的WL在Vpgm(~22V)中，相邻的WL在Vpass(7-10V)中，如图1‑27所示。存在三个问题：（1）电荷（电子）损失；相邻WL cell的FG中的电荷被放电到选定的WL中；（2）WL漏电或击穿；WL之间的高场（VPGM-VPASS\_n+1/n−1>10V）可能导致漏电或击穿；（3）编程干扰；电荷（电子）已从衬底注入到FG。为了缓解这些问题，优化VPASS\_n+1/n−1将会对于工艺微缩很重要。

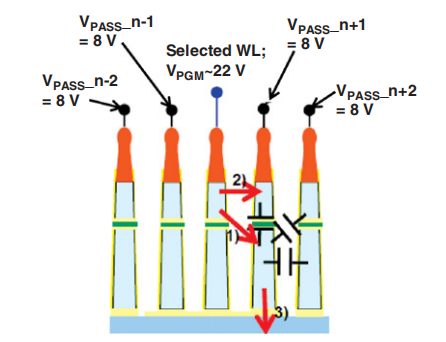


图1‑27 Word线(WL)高场问题 （1）从相邻FG到选定WL(CG)的电荷损失，（2）选定WL和相邻WL之间的漏电或击穿，以及（3）对相邻cell的编程干扰。电荷（电子）已从衬底注入到FG。

## 其他电子现象

通过缩小存储单元的大小，由于poly间介电ONO电容的降低，存储在浮栅上的电子数量显著减少。图1‑16表示了作为NAND和NOR闪存单元的技术节点的每比特的电子数(对于ΔVt为3V)的函数。预计大约有100个电子存储在1x-nm的单元中。通过进一步缩小存储单元的大小，存储电子的数量将远远小于100个电子。它将足够小，可以观察到很少的电子现象。然后研究了这些单电子现象对浮栅(FG) cell性能的影响。FG存储单元的充放电不再被认为是一种连续现象，而是一种离散随机事件的总和。这导致了数据保持时间和存储器编程窗口的离散。

对于纳米大小的存储节点，已经讨论了少量电子在存储器中充电过程的随机特性。研究还表明，由于电子电荷的泊松性质，在编程后，FG中的带电电子数存在不确定性。

图1‑28a表示每比特不同电子数(N)的数据保持分布。通过减少每比特的电子数(N)，数据保持时间TR概率密度从一个类高斯分布(当N=250时)强烈演化为一个纯指数/类泊松分布(当N=5时)。我们还可以看到，平均值周围的离散随着N的减小而增大。

图1‑28b显示了不同值下每位电子的数据保持时间TR的累积概率。对于N值较大时，累积概率演化是非常紧密的分布。另一方面，我们可以看到，随着N的减少，分布尾部的数据保持时间要短得多，这意味着放电的cell的数量将变得至关重要。

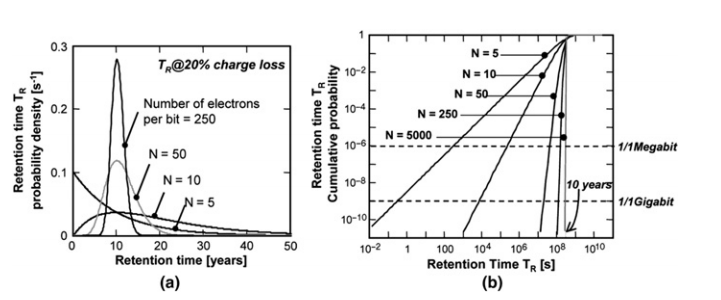


图1‑28 (a)每bit N位电子数下，存储器的数据保持时间TR的概率密度分布。平均TR固定在10年。(b)每比特电子数减少的存储器数据保持时间TR(来自(a))的累积概率

图1‑29显示了这个失效时间作为每比特电子数(N)的函数的演变。故障准则定义为在1gb阵列上的1位的数据保持时间，且数据保持对应于电荷损失的20%。我们可以看到，当每一个比特的电子数量减少时，失效时间的减少可以变得相关，并且在少数电子存储器中变得非常重要。如果我们考虑90nm的NAND Flash技术节点，一个3V的阈值电压位移相当于每位约1000个电子。因此，在这种情况下，超过1gb的不稳定位的保留时间等于6.5年。然而，如果考虑35nm的NAND Flash技术节点，每比特对应约200个电子，一个不稳定位的保留时间大大减少到3.3年，这可能是非常关键的。

此外，在MLC中，数据保持时间的操作margin将进一步减少，每位的电子数减少2bit/cell−1。图1‑29b显示了作为SLC和MLC(2位/单元) 存储技术的NAND Flash技术节点的功能的1位/1Gb的数据保持时间。该图表明，在未来的技术节点中，MLC cell将大大减少高密度内存阵列的故障时间。因此，对于35nm的内存节点，引入MLC(2位/单元)的故障时间将从3.3年减少到1年。

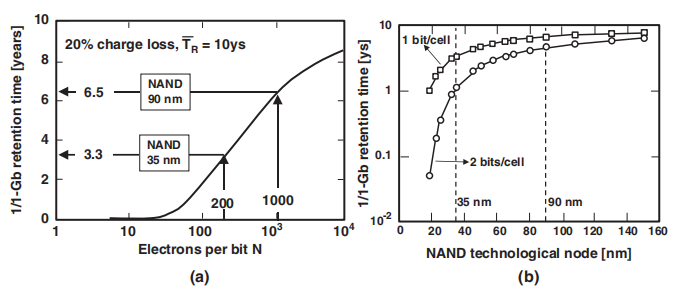


图1‑29 (a)计算了由于FG的单电子放电导致的失效时间 (b)计算出的故障时间(即1位/1gb阵列的数据保持时间)

## 图形（Patterning）限制

随着设备的最小尺寸缩小， NAND闪存单元可以很容易地跟着缩小。 由于SA-STI单元的可扩展性以及NAND Flash编程/擦除操作方案的沟道长度可微缩性，NAND闪存单元的微缩没有任何电气、操作和可靠性的限制。因此，当特征大小从0.7um减小到当前的1y-nm生成时，记忆单元的大小可以直接减小。

SA-STI cell使NAND Flash的cell大小成为理想的4∗F2。特征尺寸(F)通常由光刻工具的能力决定。目前，最先进的光刻工具是ArF浸没式(ArFi)机台。最小特征尺寸为38-40纳米。然后，特征尺寸(F)的比例被限制在38-40nm之间。为了进一步加速缩小NAND闪存单元的尺寸，Double pattern已经从3x纳米一代开始使用。在常规的Double pattern过程中，使用side-wall spacer作为图形掩模，如图1‑30中的Double pattern过程所示。由于Double pattern，特征尺寸可以从38-40nm缩小到19-20nm。此外，Quadruple pattern已经使用超过20nm，如图1‑30中的Quadruple pattern过程所示。受到严重的物理限制，使用ArFi可以将特征尺寸(F)从19-20nm缩小到9.5-10nm。对于超过9.5-10nm的比例，需要新的工具或技术来制作精细的图案。一个候选的是EUV（极紫外线）光刻工具。

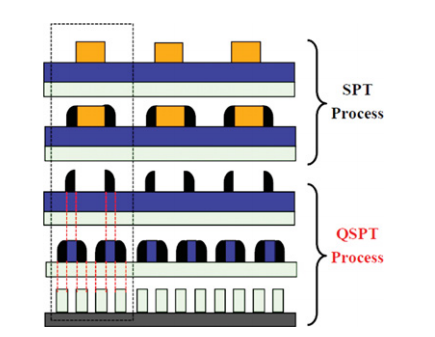
、

图1‑30 Double Pattern和Quadruple pattern关键制造步骤示意图

## 工艺波动

微缩记忆单元大小的障碍之一是工艺波动，工艺波动强烈影响NAND闪存单元的阈值电压(Vt)分布，影响其性能和可靠性。

为了工艺波动效应，提出了NAND闪存阵列的紧凑模型。该模型包括3个32个单元的与NAND String，带有选择的晶体管，并考虑了浮栅电容耦合干扰效应。只模拟了String的中心，而两个相邻的NAND string为单元间的静电耦合设置了边界条件。

特别是，由于电荷的离散性质，我们考虑了工艺过程引起的cell几何波动和更基本的（内在的）波动。Cell几何形状的过程引起波动包括W、L、隧道和poly介电厚度波动(分别用WF、LF、TOXF、IPDF表示)以及浮栅耦合系数的控制波动。后者的基本（本征）因素解释了随机掺杂(RDF)和氧化物陷阱波动(OTF)。通过改变器件参数(W、L等)，将工艺引起的波动直接代入到紧凑模型中，根据从工艺数据中提取其扩散的高斯分布。基本（本征）贡献的实现进行如下：RDF对Vt的影响，而Vt变化由于OTD可由下式表示：

，其中，Kox和Qox为工艺参数分布数据。

图1‑31显示了模拟结果，包括cell 的Vt的分布和在41nmNAND闪存阵列页面上测量的实验Vt分布。该结果表明，测量和仿真之间有很好的一致性，支持工艺波动模型的正确性。对扩散的轻微低估可能是由于编程Vt分布的软擦除操作造成的。

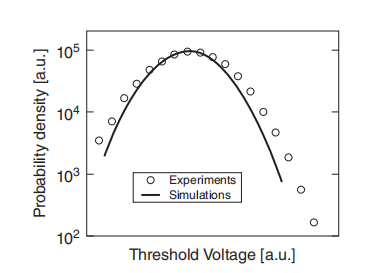


图1‑31 一个41nmNAND闪存阵列页面的Vt分布及相应的仿真结果

图1‑32显示了Vt分布Vt作为NAND flash技术节点从100-25nm规则的函数的模拟和实验标准偏差。结果表明，由于影响阵列功能的所有波动因素的退化，随着存储单元的微缩，Vt工艺波动增加。图1‑33显示了详细的相对权重的主要波动因素函数（技术节点作为变量），代表RDF（随机掺杂剂波动），OTF（氧化物陷阱波动），以及W和L的波动。结果表明，工艺波动是由几个因素，而不是由单一因素造成。

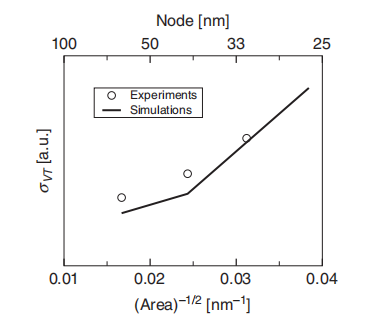


图1‑32 Vt作为技术节点函数的建模结果与实验数据的比较。

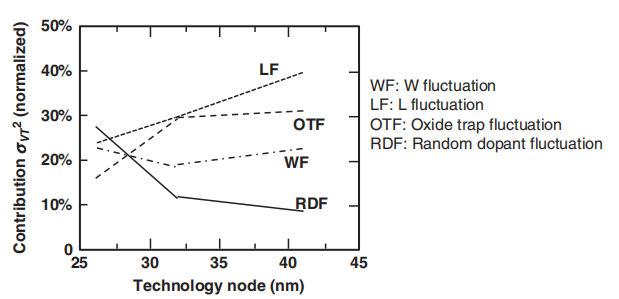


图1‑33 对不同技术节点的cell VT2（归一化）的最重要贡献

该模型，用于研究cell Vt的行为，被扩展到模拟编程和擦除操作。在擦写操作过程中，在VT中起着基本作用的最重要的参数之一是控制栅到浮动栅耦合系数(G)。该参数与浮栅定义所采用的结构有关，其波动取决于几何参数的分布，如图1‑34a所示。这些因素对G中传播的贡献如图1‑34b所示。在这个模拟结果中，tFG和tR的波动对所有技术节点起着主要作用。

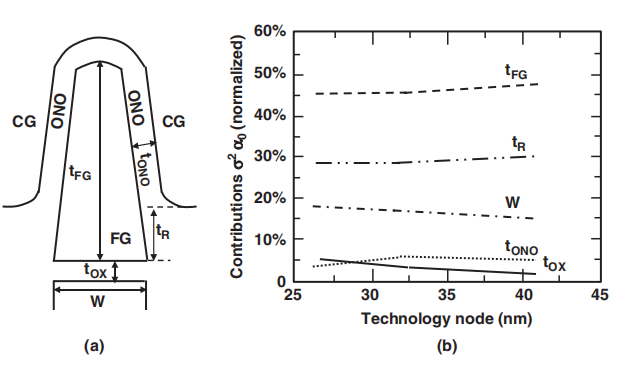


图1‑34(a)记忆单元沿W方向的横截面示意图视图，显示浮动栅极几何形状。(b)不同的cell对不同技术节点的耦合系数G的扩散的贡献。

在小于25nm的存储cell中，RDF（随机掺杂剂波动）是导致cell Vt变化的主要因素，如图1‑33所示。随着存储单元的缩小，每个单元的掺杂原子数量减少，导致阈值电压的标准偏差更大。

## 微缩对数据保持的影响

数据保持特征具有相邻单元的数据模式依赖性。当相邻单元处于擦除状态时，编程单元比在编程状态时具有更大的阈值电压(Vt)损失。在同一位线和字线上的单元格对Vt损失的加速有类似的影响。这一现象可以用相邻cell中电荷的影响来解释，因此相邻cell中存储的电荷会对栅极和目标cell 有源区域周边的的隧道氧化层电场产生影响。

采用60nm技术的cell阵列来分析单cell数据保持的电学表征。这些阵列允许在NAND cell阵列的中心部分的三个相邻的WLs和BLs的任意偏置条件，如图1‑35a所示。图中的阴影圆圈显示了将VGS应用于选中的WL，所有其他WL加Vpass电压时，选定的单元cell在gate stress实验。

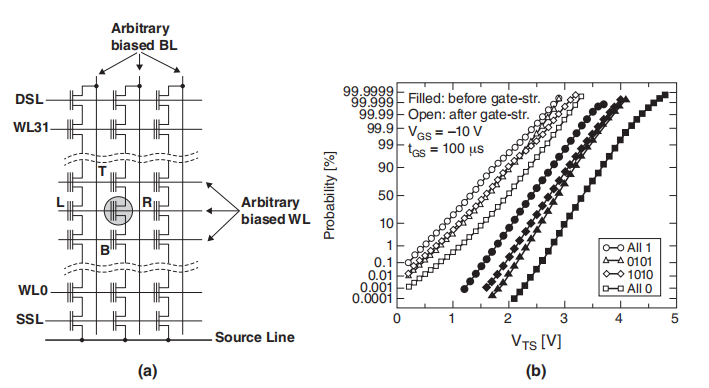


图1‑35 (a)在NAND阵列中的cell连接示意图 (b)在VGS=−10V进行100us负压应力实验后，在70nmNAND测试芯片上采用不同背景图形下测量的累积VTS分布

为了研究相邻单元数据图形的依赖性，我们进行了3D TCAD模拟，如图1‑36所示。3D TCAD模拟的模板NAND器件结构如图1‑36a所示。在60~70nm设计的NAND闪存单元中，All0（左）和All1（右）数据的相邻单元上计算出的隧道电流密度如图1‑36b所示。在有源区域的角落产生较大的尖峰电流，源/漏结与单元浮栅重叠。这些峰在All1情况下更高，因为当相邻的浮栅带正电荷时，有源区边缘的静电分布很强。Cell有源区拐角处沿BL方向和WL方向的电流密度分布有尖峰。结果证实，在负栅应力条件下，较大的隧穿电流流过源/漏结。

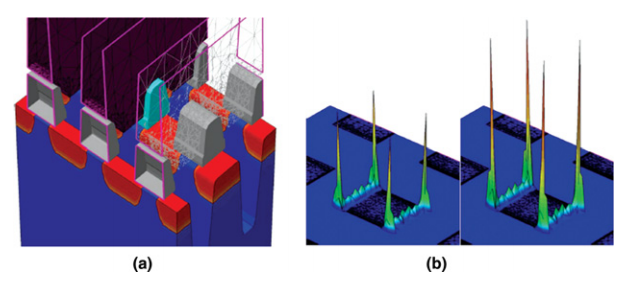


图1‑36 (a)NAND器件的三维TCAD（技术计算机辅助设计）模拟结构。(b)计算了All0和All1背景模式在选定的cell 有源区边缘的隧道电流密度。