

3D NAND Flash存储器模型模拟技术

在对 3D NAND 存储器器件单元特性进行详细介绍后,本章将对 3D NAND 存储器模型的模拟技术进行概述,从而更好地推动 3D NAND 存储器技术的发展。本章首先介绍仿真工具,并分析了纳米尺度器件模拟过程中使用的模型,最后分别对 2D NAND Flash 器件和 3D NAND Flash 器件模型和模拟方法进行归纳总结。

5.1 仿真工具简介

由于 3D NAND 存储器结构复杂,使用仿真工具进行仿真可以加快研发速度。商用仿真工具收录模型全面,覆盖领域广泛,性能稳定可靠,在半导体产业界有着重要的应用。目前主流商用仿真工具包括 Synopsys 公司的 Sentaurus 计算机辅助设计技术(Technology Computer-Aided Design, TCAD)工具和 Silvaco 公司的 Silvaco 仿真工具。下面对 Sentaurus TCAD 模拟仿真工具以及其在 3D 电荷俘获存储器中的应用进行介绍。

5.1.1 Sentaurus TCAD

TCAD 是指利用计算机仿真技术进行半导体工艺模拟以及半导体器件模拟的技术,所使用的软件就是 TCAD 仿真工具。TCAD 仿真工具通过对基本的物理模型方程进行计算求解(例如求解漂移扩散方程等),还原半导体器件在制造或者工作过程中发生的物理过程。TCAD 工具主要有以下几方面的应用。

(1) 对现有的 IC 工艺流进行模拟,对其进行监测、分析和优化,分析每一步工艺波动对器件的影响。

(2) 对现有器件的电学特性进行模拟,帮助改进器件设计和优化器件的操作方式。

(3) 对新工艺新器件进行模拟,提供具有参考价值的预测性结果。

如图 5.1 所示,TCAD 模拟已广泛应用于存储器件、CMOS 逻辑器件、光电器件、射频器件以及高压器件等半导体领域。随着半导体器件的持续微缩,技术变得更加复杂,TCAD 作为辅助性设计工具已成为半导体行业不可或缺的组成部分。

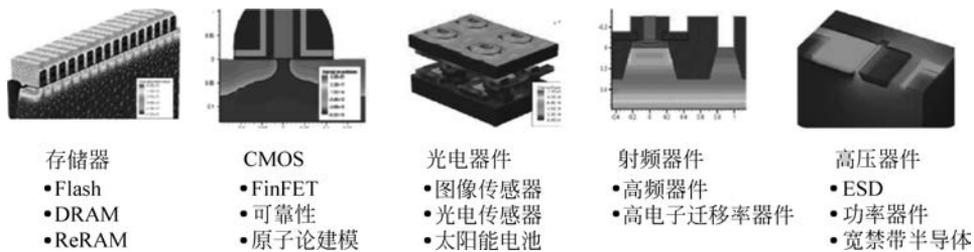


图 5.1 TCAD 工具的应用领域

作为一种常用的 TCAD 工具, Sentaurus TCAD 仿真工具是 Synopsys 公司在收购和整合了多家公司的 TCAD 仿真软件之后形成的完整的 TCAD 仿真组件集合。其中, 最为核心的是 Sentaurus 模拟仿真工作平台(Sentaurus Workbench, SWB)^[1]。如图 5.2 所示, SWB 包含众多组件, 主要由工艺仿真模块(Sentaurus Process, S-Process)、器件结构仿真模块(Sentaurus Structure Editor, SDE)和器件物理特性仿真模块(Sentaurus Device, S-Device)等模块构成。基于 SWB, 可以使用其他各种仿真组件, 并且不同仿真组件间可以相互调用。

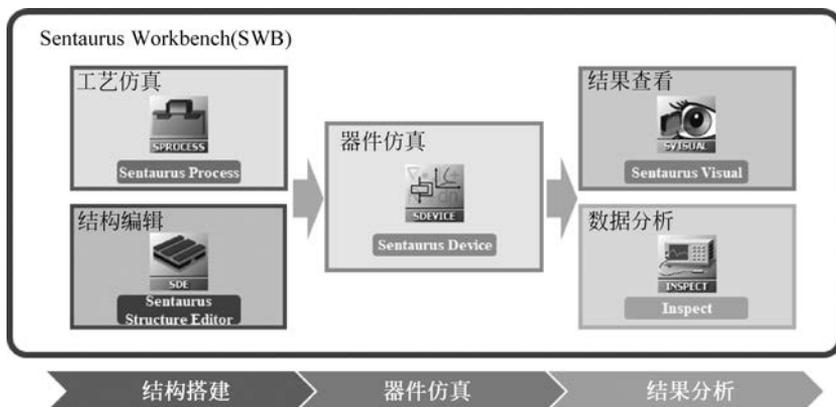


图 5.2 Sentaurus TCAD 工具组成

5.1.2 Sentaurus WorkBench

SWB(Sentaurus WorkBench)将各种 Sentaurus 仿真工具整合到一个环境中。用户可以通过图形用户界面(Graphical User Interface, GUI)直观地设计、组织和运行各个模块, 进行半导体工艺和器件的仿真模拟。仿真流程通常包括多个模块, 图 5.3(a)展示出已经整合到 SWB 里的仿真模块, 包括工艺仿真、器件仿真、输出和分析模块。将所有的模块整合到一个平台后, SWB 可以自动管理各个模块间的信息流。其中, 信息流包括输入程序及参数的预处理、参数的传递、各个模块执行顺序、各个模块的输入输出、仿真结果的查看等。使用 SWB 平台, 用户可以自定义参数和变量, 从而灵活地进行参变量分析^[1]。其结果可导出到处理软件中, 也可以直接输出到 GUI 的表格中, 如图 5.3(b)所示。

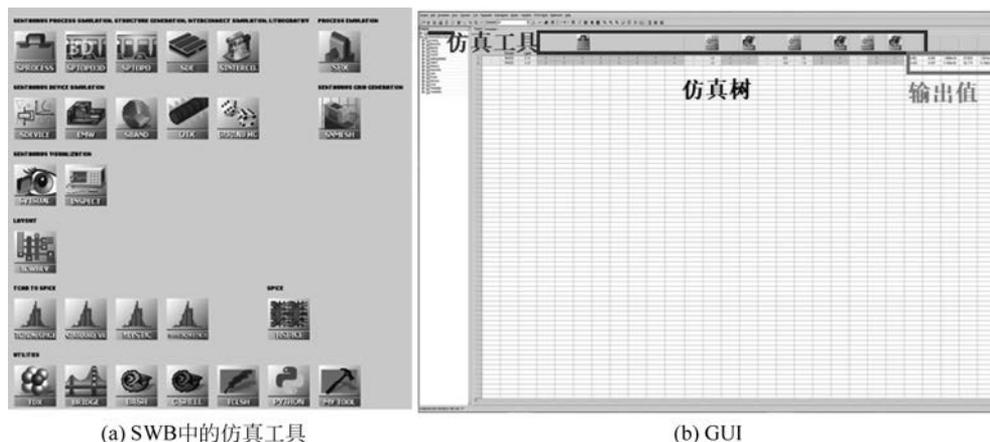


图 5.3 SWB 中的仿真工具及 GUI

5.1.3 Sentaurus Process

Sentaurus Process(S-Process)是一个完整的、高度灵活的多维工艺模拟工具。它以现代化软件体系结构为基础,通过已验证的校准方法对大量的实验数据进行模拟。S-Process 工艺仿真对现代硅和其他半导体的工艺提供了很好的预测能力^[2]。

通过求解不同物理方程,S-Process 可以仿真不同的工艺步骤,如离子注入、杂质扩散、氧化、刻蚀、沉积、热退火等。其中离子注入仿真可以分为蒙特卡罗法和解析查表法。模拟的晶圆或器件离散化地划分为网格,在每一个网格内通过有限元计算求解物理方程。相比于接下来介绍的结构编辑工具 Sentaurus Structure Editor,使用 S-Process 构建仿真结构的优势在于:能够准确计算注入和退火扩散后的杂质分布;可以仿真氧化过程中 SiO_2 的结构(比如 LOCOS 工艺中的鸟嘴结构)及结构变化引入的机械应力。

图 5.4 中列举了一个简单的 2D MOSFET 的工艺仿真流程。利用 S-Process 工艺仿真工具,可以清晰准确地提取出器件在每一步工艺过程中的结构变化和杂质的再分布情况。为集成工艺的优化提供了参考,同时为之后器件电学特性的仿真奠定了基础。

图 5.5(a)所示为 LOCOS 氧化工艺中产生的鸟嘴结构。虽然在先进的工艺制程中已经不再使用 LOCOS 工艺,但这个例子有助于理解使用 S-Process 工具进行仿真的优势,其可以准确地仿真出在氧化过程中硅材料逐渐被氧化的过程。图 5.5(b)所示是 MOSFET 局部结构,S-Process 可以准确地仿真在氧化工艺后漏端上的氧化层、栅氧化层、栅侧壁上氧化层及交界处的厚度,并且 S-Process 还可以计算由于结构的变化引入的局部应力变化。由于器件结构和应力的改变都会引起器件电学特性的变化,S-Process 能够为后面进行器件仿真提供准确的支持。

5.1.4 Sentaurus Structure Editor

SDE 是 2D 和 3D 器件结构搭建及编辑工具。SDE 不进行工艺的仿真过程,而是将工艺步骤转换成几何操作,从而完成器件结构的搭建。相比于 S-Process,SDE 可以灵活地编

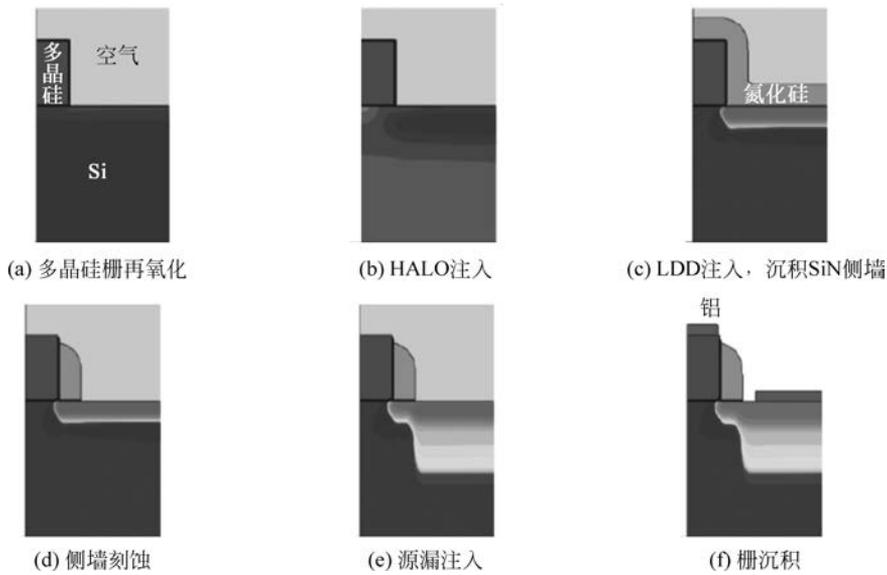


图 5.4 简单 2D MOSFET 的工艺仿真流程

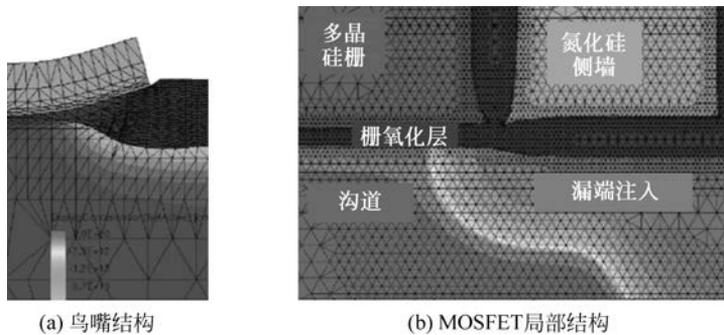


图 5.5 LOCOS 氧化工艺产生的鸟嘴结构以及 MOSFET 的局部结构

辑器件结构,例如进行圆滑、旋转、摆动和结构之间的混合等结构编辑功能。SDE 可与 S-Process 联合使用,以弥补各自的不足。SDE 支持图形用户界面,以交互的方式可视化地生成、编辑器件结构。SDE 工具在工作时的输入输出文件如图 5.6 所示。SDE 输入可以是命令行或文本方式组织的语句,也可以是 S-Process 生成的结构边界文件和带有网格的杂质分布文件。SDE 输出为带有具体结构、掺杂分布和网格的文件^[3]。

为了对器件进行离散化处理以完成仿真计算,需要将器件分割成若干小的三角形。这些小的三角形称为网格,这一分割过程就称为网格划分^[3]。网格的划分对仿真过程的收敛性和仿真结果的准确性起到了至关重要的影响。网格数目太少,会降低仿真结果的准确性,而划分得过多会大幅增加计算量,同时会导致仿真的收敛性变差。此外,SWB 平台还有专门的网格生成工具 Sentauros Mesh,同时 SDE 也支持网格生成策略,即调用 Sentauros Mesh,以生成基于四叉树或八叉树的网格,既可以产生与坐标轴对齐的普通网格结构,也可以沿着材料边界方向定义出特殊的网格。SDE 中进行的网格优化是为后面的器件仿真做铺垫的重要步骤。所以使用 SDE 生成网格时,需要在电场强度大或者掺杂浓度高的地方加

密网格,以保证器件仿真具有足够的收敛性。

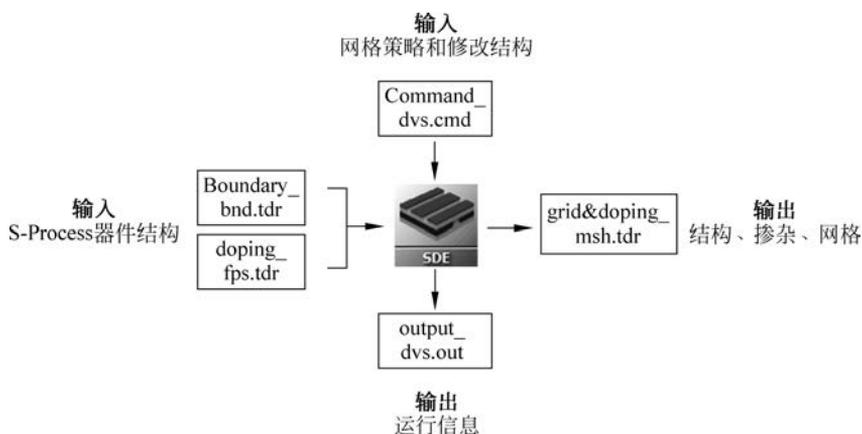


图 5.6 SDE 工具的输入与输出流程

5.1.5 Sentaurus Device

Sentaurus Device(S-Device)是模拟各种半导体器件的电学、热学、光学等物理特性的仿真工具。作为业界领先的半导体器件模拟器,S-Device 可以基于 2D 和 3D 的器件结构,进行物理特性的仿真。此外,还可以结合紧凑模型(compact modeling)添加一些电学元件,从而进行混合电路仿真。S-Device 的强大之处在于其包含一个复杂的物理模型库集,可以对多种器件和材料进行物理特性的模拟与预测^[4]。通过使用 S-Device 工具,可以对 S-Process 或 SDE 工具生成的结构进行电学特性模拟,可以得到器件在工作过程中各个时刻的多种物理量,包括电场、电势、电流、载流子浓度和能带结构等。

S-Device 可以对不同类型的半导体器件进行仿真。例如,在进行 CMOS 逻辑器件仿真时,在根据二次离子质谱(Secondary Ion Mass Spectroscopy, SIMS)实验结果校准后,可以准确地模拟出阈值电压 V_{th} 、roll-off 特性和开态电流 I_{on} 的大小。S-Device 还可以实现基于阻抗场法(Impedance Field Method, IFM)的变异性(variability)分析技术,从而能够用来分析线边缘粗糙度(line edge roughness)可能引起的器件性能的变异性。在存储器仿真领域,S-Device 也有着重要应用,可以应用在不同种类的存储器中,如 SRAM、DRAM、NOR Flash、NAND Flash、RRAM、PCM 等。此外,S-Device 还可以应用在射频器件领域和功率器件领域。

如图 5.7 所示,S-Device 仿真工具的输入为之前 S-Process 或 SDE 输出的具有结构、掺杂分布和网格的文件,以及物理模型、求解所使用的迭代算法及电极电压要求等在内的指令文件。同时还要输入参数文件,仿真中涉及的材料特性参数需要在参数文件中指出,否则将使用默认参数。S-Device 输出的分析文件分为两种:一种是器件的物理特性曲线,比如转移特性曲线(I_d - V_g);另一种是包含了各种物理量(如电场、电势、载流子浓度等)的可视化结构文件。

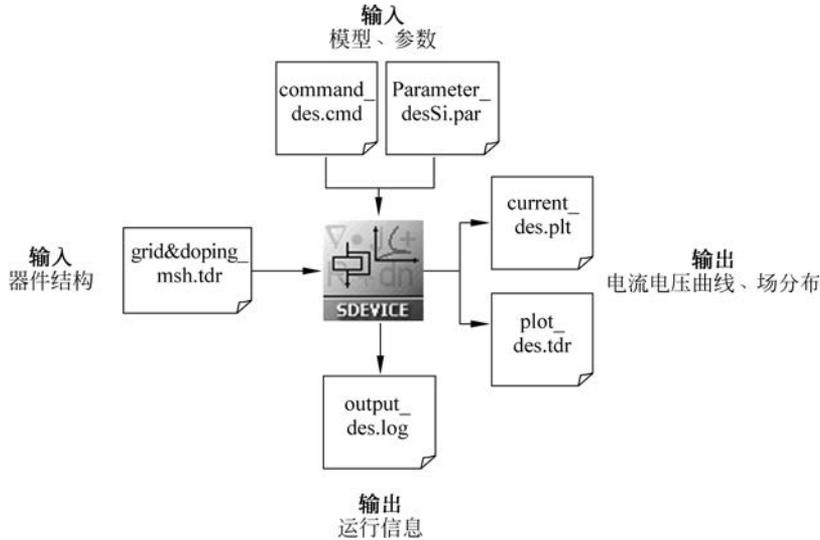


图 5.7 S-Device 工具的输入与输出流程

5.2 纳米尺度器件模拟

5.2.1 纳米尺度 MOS 器件输运特性简介

随着集成电路制造工艺的不断改进,集成电路产业得到了迅猛的发展,如今,5nm 工艺制程已进入大规模量产阶段,单个芯片上集成的晶体管数量 10^{10} 以上。为了满足对产品的设计要求,对器件的设计、制造和分析需要综合考虑其详细的器件物理特性。

回望 MOSFET 器件的发展,自 MOS 器件的尺寸进入纳米尺度后,面临着如图 5.8 所示的一系列问题。

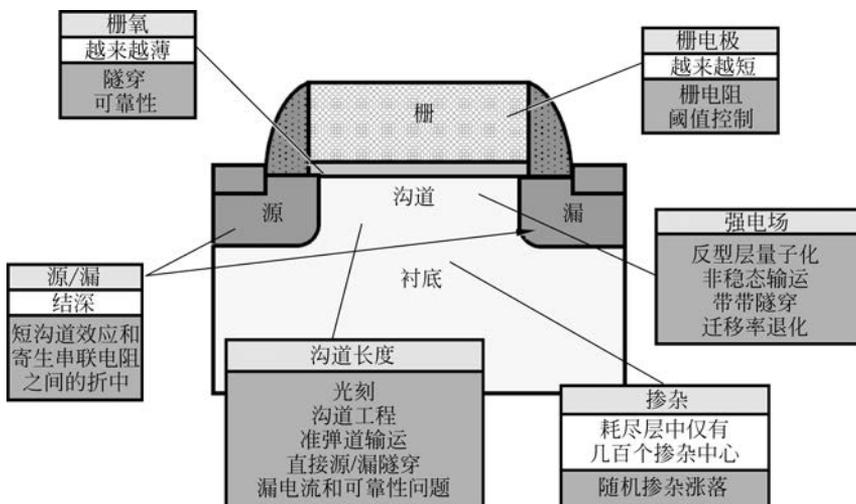


图 5.8 纳米尺度 MOSFET 器件遇到的问题

(1) 随着栅氧化层厚度的不断减薄,栅氧隧穿和可靠性问题变得越来越严重,要考虑隧穿漏电流、可靠性和多晶硅栅的耗尽等问题。

(2) 由于尺寸变小,沟道量子化效应及沟道变短、带带隧穿(Band To Band Tunneling, BTBT)效应、迁移率降低、非平衡态的输运,栅电极变得越来越短,栅电阻变得越来越大,阈值控制也越来越难。

(3) 由于沟道中强电场的影响,反型层载流子量子化和非稳态输运变得更显著,高电场诱发的带带隧穿和沟道中的载流子迁移率退化严重地影响了器件性能。

(4) 由于器件尺寸的缩小,耗尽层中仅有几百个掺杂中心,随机掺杂涨落的影响越来越大。

(5) 沟道长度的不断减小使得源漏隧穿现象和准弹道输运变得显著。

(6) 为了减小短沟道效应和源漏串联电阻,需要对源漏的结深做优化设计。

以上问题都成为制约器件尺度进一步缩小的因素,如何解决这些问题,如何将这些问题的影响降至最小,成为目前亟待解决的主要问题。MOSFET 器件是构成 MOSFET 集成电路的基本单元,输运则是 MOSFET 器件特性的基础。在以上各种纳米尺度 MOSFET 器件遇到的问题中,输运方面的问题将直接决定纳米尺度器件特性,纳米尺度器件要求从根本上对输运特性做新的研究和认识,分析其对器件特性的影响,从而建立正确的器件模型,为实验制备提供更好的支持,同时为电路设计提供坚实的基础。

纳米尺度 MOSFET 器件中的输运行为主要有三个特点:非稳态输运、准弹道输运和量子输运。

随着器件尺寸不断缩小,器件内的电场越来越大,强电场效应的影响已经无法忽略。在器件内电场不是很大的情况下,载流子分布趋向于平衡分布,载流子的平均速度与电场强度成正比,其比值是迁移率。迁移率的大小在微观上来说取决于各种不同的散射机制的影响。在纳米尺度 MOSFET 器件内,电场强度急剧增大,载流子分布已经远远偏离了平衡分布,根据载流子输运与电场强度的关系,漂移速度首先从线性关系经过过渡区进入饱和区。由于起始阶段载流子能量弛豫时间大于动量弛豫时间,可能产生漂移速度过冲现象。研究表明,稳态输运不能描述此时器件中载流子的输运过程,必须使用非局部、非稳态载流子输运模型。图 5.9 给出了纳米尺度 MOSFET 器件内不同部分的载流子分布。图 5.9(a)和图 5.9(b)分别为源端和漏端的分布,可以看出,由于源端和漏端的电场强度较低,此时的分布趋向于平衡分布。图 5.9(c)和图 5.9(d)分别为沟道内靠近源端和靠近漏端的分布,强电场使沟道内的载流子分布已远远偏离平衡位置。在沟道不同位置,中间圆圈非常小的点,只位于导带底或者价带顶。当移动到其他地方时,中间的圆圈会很大,导带底在中间,距离越远,K 空间的波矢便越远离导带底或者价带顶,为非平衡态的输运。高能载流子能量越高,对可靠性损害越大,不仅损坏钝化之后的悬挂键,甚至会进入氧化层中,打断 Si-O 键,导致器件的进一步退化。

对于纳米尺度 MOSFET 器件,载流子在渡越沟道的过程中将受到极少的散射。在这种情况下,载流子将实现准弹道输运甚至弹道输运。由于散射次数显著减少,载流子将以较大的速度渡过沟道,器件的特性也将会有很大的提高。同时由于准弹道输运已不属于散射现象占主导地位的输运,漂移扩散(Drift Diffusion, DD)模型和流体动力学(HydroDynamics, HD)模型不再适用,如何预测纳米尺度 MOS 器件的准弹道输运特性以及对器件特性的影响变得尤为重要。

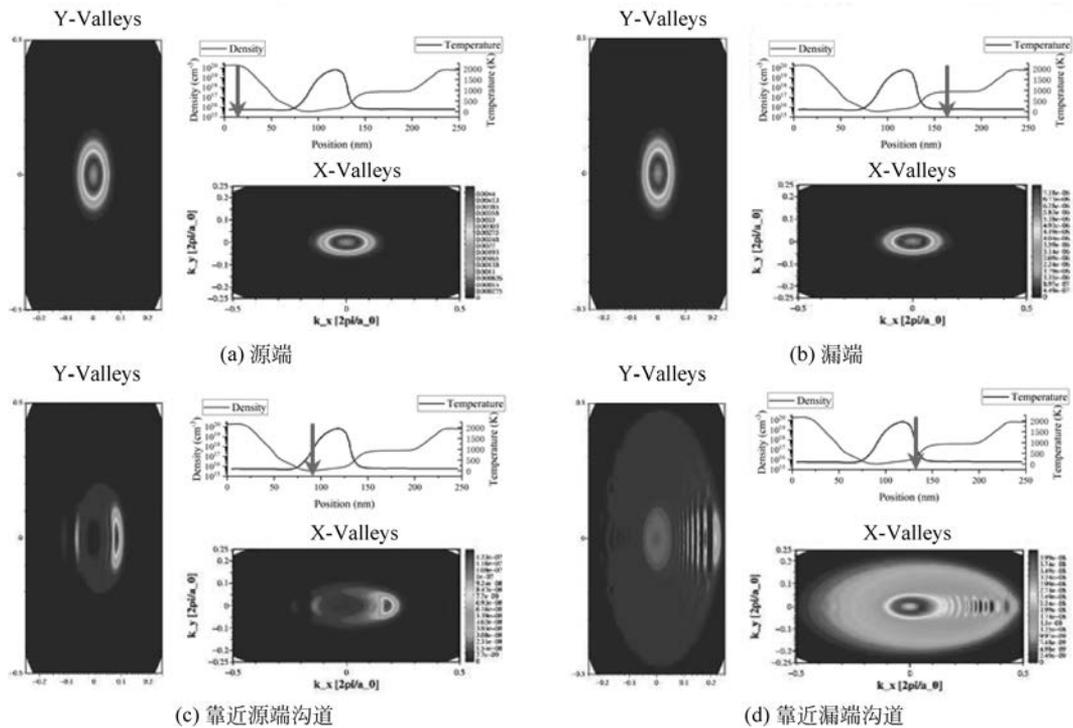


图 5.9 纳米尺度 MOSFET 器件内不同部分的载流子分布

由于波函数可以透射、散射、反射等性质,载流子会出现图 5.10 所示的行为。在纳米尺度下,能级会分裂,原来是 3D 的电子,在这里会变成 2D 电子气,能级就会有阶梯形的能级效应,反映到具体的器件中,沟道中能级会量子化,能级量子化会带来一些结果。传统情况下,载流子从沟道表面向沟道里逐渐减少,但是由于能级量子化,峰值会平移,此平移会带来一些问题,如实际和等效的氧化层厚度变厚,同时由于变厚,阈值电压也会漂移,量子化的阈值电压会高一些。逻辑中对阈值电压的要求是不变的,但是在存储器中,阈值电压可以通过编程/擦除操作进行调节。

5.2.2 半导体仿真模拟概述

半导体器件模拟是指通过电子计算机的模拟计算,得到半导体器件或集成电路的性能参数。

早期的器件模拟主要是基于漂移扩散模型或流体动力学模型^[5-10]。其历史可以追溯到 H. K. Gummel 在 1964 年的开创性工作^[11],他在一维情况下求解了漂移扩散方程,采用的器件结构是双极型晶体管。20 世纪 70 年代后期,出现了一些专用模拟程序,比如维也纳大学的 MINIMOS(针对 MOS 器件)^[12],斯坦福大学的 SEDAN(针对双极器件)^[13]。实际上,当时已经出现了通用、非平面结构及多维的模拟程序,其中最著名是 IBM 公司的 FIELDAY^[14]。真正具有影响力的 2D 通用模拟器是斯坦福大学于 20 世纪 80 年代开发的 PISCES-II 程序^[15],它能够模拟两种载流子和非平面结构,除了 Si 以外还能够模拟 Ge、GaAs 等其他材料。20 世纪 90 年代,PISCES-II 的改进版本 PISCES-2ET 能够模拟异质结

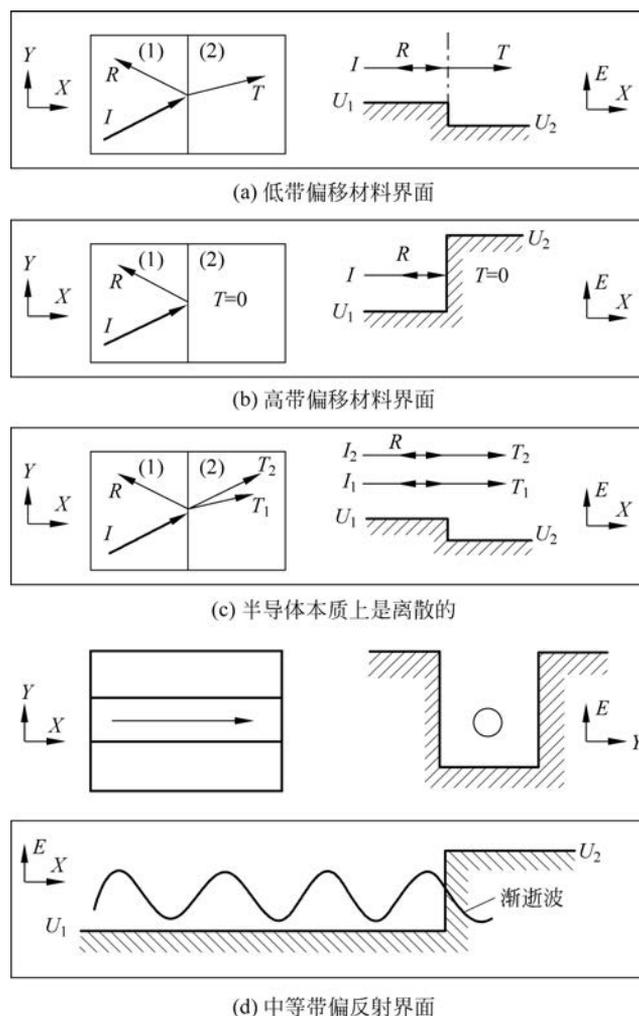


图 5.10 波函数效应

器件^[16],同时在原有的漂移扩散模型的基础上增加了能量平衡方程,很大程度地丰富了器件的模拟范围,最终 PISCES 发展成为 MEDICI^[17],后期被 Synopsys 收购。不过由于软件主要基于经典情况下的漂移扩散模型和流体动力学模型,无法准确模拟纳米尺度 MOS 器件的输运特性。而另外一些模拟方法如 Wigner 函数^[18]和 Green 函数方法^[19]虽然在理论上是最精确的模拟方法,但是由于其自身的复杂性,无论是精度还是速度都无法满足器件模拟的需要。图 5.11 给出了器件模拟所采用的各种方法。

蒙特卡罗方法是应用于半导体中的电荷传输的用于求解半经典玻尔兹曼方程的统计数值方法。它包括依次模拟空间中的粒子运动的单粒子蒙特卡罗模拟和作为一个整体的多粒子蒙特卡罗模拟。这种方法首先由 Jacoboni 和 Reggiani 于 1983 年用于散装材料研究^[20],经过大约十年后被应用于器件仿真^[21-22]。在过去的 20 年中,对蒙特卡罗器件仿真技术进行了许多项改进,例如全频带色散关系的数值描述^[21],实现与泊松方程的自洽,模拟热电子现象(碰撞电离、栅极电介质注入等)等。

近似				简单近似
	方法、模型	特点	局限	
准经典	漂移-扩散模型	适合 $0.5\mu\text{m}$ 以上器件, 包括 $\mu(E)$	无法处理热载流子效应	
	流体动力学模型	热载流子和速度过冲	高估强场下载流子速度	
	玻尔兹曼方程 蒙特卡罗法	经典理论下最准确	费时	
量子	Wigner函数 密度矩阵	准确的单粒子描述	难以适用于多维模拟	
	Green函数 路径积分	至今最为准确的方法	至今最为困难的方法	
准确				困难

图 5.11 器件模拟的各种方法

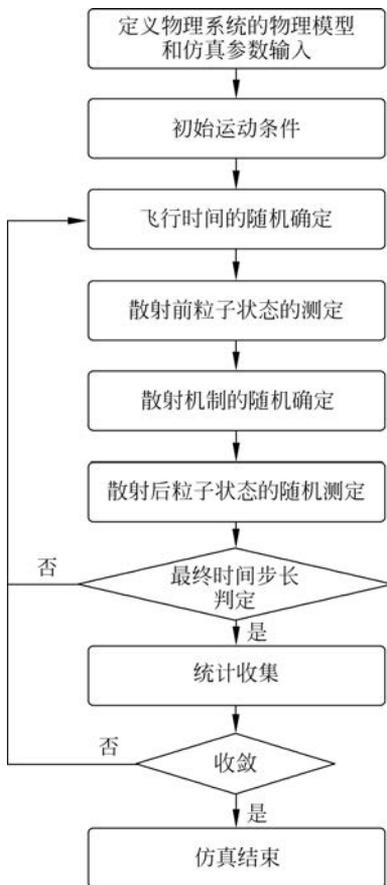


图 5.12 蒙特卡罗仿真流程

蒙特卡罗方法基于随机数的生成,用于随机确定自由飞行持续时间并选择散射后的状态。在载体运动期间,将执行基于图 5.12 的基本步骤。

(1) 散射率评估,为了确定自由飞行时间的中断频率。

(2) 自由飞行时长随机确定。

(3) 载体运动开始,遵循经典运动规律。

(4) 自由飞行被散射事件打断。

(5) 基于该载体状态的单个机制的散射率,以一种偶然的方式选择负责自由飞行停止的散射机制;一旦确定了散射机制,就可以随意选择粒子状态,即确定最终的动量。

重复以上步骤,直到到达时间步长终点,此时可确定总体的相位状态。

图 5.13 给出了使用不同模拟方法模拟 20nm 沟长器件的结果对比图。从图中可以看出,由于此时的漂移扩散模型已不再符合纳米尺度 MOSFET 器件的载流子输运状态,漂移扩散模型的结果与其他方法的结果相差较大。由于非平衡格林函数方法(Non-Equilibrium Green's Function Formalism, NEGF)^[23]忽略了散射的影响,模拟结果又比蒙特卡罗方法偏小。当去掉蒙特卡罗方法中的散射事件时,蒙特卡罗的结果和 NEGF 模拟结果偏差变小。由此可

见,在实际的纳米尺度 MOSFET 器件输运特性模拟中,随机散射仍然具有重要的影响,蒙特卡罗模拟方法是精确模拟纳米尺度 MOSFET 器件输运的首选方法。

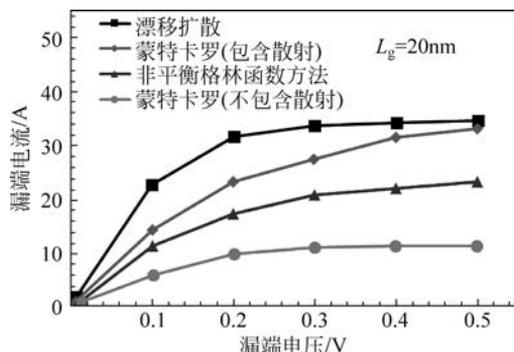


图 5.13 各种不同模拟方法的结果对比

5.2.3 半导体器件模型介绍

半导体器件仿真基本模型主要包含以下几种主要模型。

1. 泊松方程

在半导体物理中,泊松方程用来描述空间电荷区中电荷密度与电场之间的关系:

$$\nabla^2 \phi = -\frac{\rho}{\epsilon\epsilon_0} \quad (5.1)$$

其中, ϵ 和 ϵ_0 分别为相对介电常数和真空介电常数; ρ 代表空间电荷区中电荷密度; ϕ 代表电势。一维情况下,式(5.1)可简化为

$$\frac{d^2 \phi}{dx^2} = -\frac{\rho}{\epsilon\epsilon_0} \quad (5.2)$$

2. 漂移扩散模型/流体力学模型

漂移扩散模型/流体力学模型的最大差异是电子-空穴的温度项。假设载流子是一个平衡态,只在导带底或者价带顶,便是漂移扩散模型。

$$\nabla(\epsilon\epsilon_0) \nabla \phi = -q(p - n + N_D^+ - N_A^-) \quad (5.3)$$

$$\frac{\partial n}{\partial t} = G_n - R_n + \frac{1}{q} \nabla J_n \quad (5.4)$$

$$\frac{\partial p}{\partial t} = G_p - R_p - \frac{1}{q} \nabla J_p \quad (5.5)$$

式(5.3)是泊松方程,式(5.4)和式(5.5)分别是电子和空穴的连续性方程;另外还需要两个附加的条件:

$$J_n = q\mu_n n E_n + qD_n \nabla n \quad (5.6)$$

$$J_p = q\mu_p p E_p - qD_p \nabla p \quad (5.7)$$

随着半导体技术的不断发展,器件尺寸越来越小,为了更精确地描述载流子的输运特性,开始引入流体力学模型,该模型引入了载流子的能量,同时可以模拟高电场下的载流子传输。

3. 迁移率退化模型

迁移率退化模型主要包含电离杂质散射模型、表面退化模型和高场饱和模型。由于半导体沟道会进行掺杂,在晶格中总是存在一些磷或者硼,作为正电中心或负电中心,可以通过库仑作用影响载流子的运动,该种现象称为电离杂质散射,由于是电离中心的库伦作用造成的影响,也称为库仑散射,库仑散射与掺杂是强相关的。对于表面退化模型,沟道中的载流子会因为表面粗糙度受到散射,由于表面散射作用,迁移率会显著降低。而对于高电场饱和模型,当漏端电场比较大时,会发生高电场饱和现象。迁移率如果是恒定值,电场很大,速度 $v = \mu E$ 也将变得很大,在常温下,硅中载流子的漂移速度是 10^7 cm/s,因此电场增强,载流子迁移率便会下降,产生这一现象的原因是光学声子散射。当光学声子的能量比较高时,可以与载流子发生强烈碰撞,因此,当载流子的动能大于光学声子的能量时,便会失去部分动能,载流子速度出现饱和现象。

4. 产生复合模型

产生复合模型过程描述的是导带和价带之间交换载流子的过程,这在器件物理学中非常重要,大多数模型是局部的,因为它们实现不涉及电荷的空间传输。对于每个单独的产生复合过程,所涉及的电子和空穴在同一位置出现或消失。其中的例外是 SRH(Shockley-Read-Hall)模型和带带隧穿模型。接下来介绍这两类产生复合模型。

1) SRH 模型

SRH 模型通过禁带中的深陷阱能级进行间接复合,通常称为 SRH 复合^[24]。SRH 模型的复合率为

$$R_{\text{net}}^{\text{SRH}} = \frac{np - n_{i,\text{eff}}^2}{\tau_p(n + n_1) + \tau_n(p + p_1)} \quad (5.8)$$

$$n_1 = n_{i,\text{eff}} \exp\left(\frac{E_{\text{Trap}}}{kT}\right) \quad (5.9)$$

$$p_1 = p_{i,\text{eff}} \exp\left(\frac{-E_{\text{Trap}}}{kT}\right) \quad (5.10)$$

其中, E_{Trap} 为陷阱能级与本征费米能级的差值; τ_n 和 τ_p 分别为电子和空穴的寿命。在 SRH 模型中,载流子寿命与掺杂浓度、电场增强以及温度有紧密的联系,因此,在仿真过程中需根据实际需求选择产生复合影响因素。

2) 带带隧穿模型

隧穿是一种量子力学过程,其中电子移动穿过能量势垒。当电子从价带通过禁带跃迁到导带(或反之亦然)时,即发生带带隧穿。在平面 MOSFET 中,BTBT 路径沿着垂直沟道和平行沟道方向传导^[25]。结中的电场既受栅极到漏极电位的控制,也受漏极到体电位的控制^[26]。

能带间的隧穿过程主要包含两种形式:局部隧穿和非局部隧穿。对于局部隧穿模型,不会发生真正通过势垒的载流子传输,而是在充分假设的前提下将隧穿过程的非局部属性转换为局部变量,然后将通常取决于电场或费米势的梯度的产生复合项添加到连续性方程中。因此,此模型仅使用离散化网格的每个顶点处的局部变量,而忽略了隧穿效应的距离影响。

与局部隧穿模型相比,非局部隧穿模型引入了通过势垒的实际空间载流子传输。两种方法的区别如图 5.14 所示。对于局部隧穿模型,电子在同一位置复合空穴,而在非局部模型中,电子和空穴在不同的位置复合,因此存在真正的载流子通过势垒。

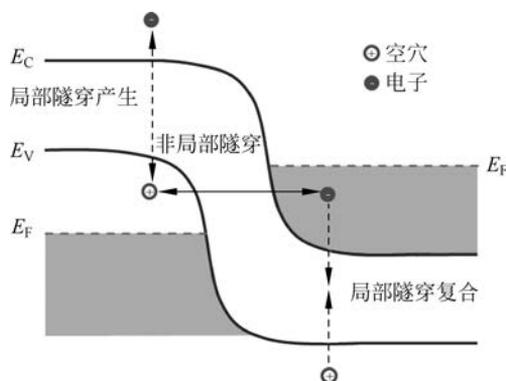


图 5.14 局部和非局部隧穿模型的空间行为之间的差异^[27]

S-Device 仿真工具可以仿真 1D、2D 及 3D 结构,并且包含多种隧穿模型。Schenk 模型是基本的声子辅助隧穿模型,没有对非局部依赖性区域进行计算,即不执行对导带等于价带的区域的计算。使用 Schenk 模型的产生率为

$$G_{\text{BTBT}} = AE^{3.5} \frac{np - n_{i,\text{eff}}^2}{(n + n_{i,\text{eff}})(p + n_{i,\text{eff}})} \left[\frac{E_C^{-1.5} \exp\left(\frac{-E_C^{\mp}}{E}\right)}{\exp\left(\frac{\hbar\omega}{kT}\right) - 1} + \frac{E_C^{-1.5} \exp\left(\frac{-E_C^{\mp}}{E}\right)}{1 - \exp\left(\frac{\hbar\omega}{kT}\right)} \right] \quad (5.11)$$

而 Hurkx 模型同样不支持基本的非局部隧穿的带带隧穿模型,Hurkx 模型没有考虑任何非局部依赖性,但它的优势是在计算带隙时考虑了温度依赖性,该模型表达式为

$$G_{\text{BTBT}} = DA \left(\frac{E}{1\left(\frac{\text{V}}{\text{cm}}\right)} \right)^P \exp \left[\frac{-BE_G T^{1.5}}{E_G (300\text{K})^{1.5} E} \right] \quad (5.12)$$

$$D = \frac{np - n_{i,\text{eff}}^2}{(n + n_{i,\text{eff}})(p + n_{i,\text{eff}})} (1 - |\alpha|) + \alpha \quad (5.13)$$

S-Device 包括一个简单的 Kane 公式,用于带带隧穿^[28],其简单模型为

$$G_{\text{BTBT}} = AE^P \exp\left(\frac{-B}{E}\right) \quad (5.14)$$

隧穿概率取决于沿着整个隧穿路径的能带分布。根据 GIDL 效应的物理机制,由电场引起的足够的能带弯曲可以使价带中某个位置的电子通过直接隧穿或者声子辅助隧穿到达导带,此时模型为

$$R_V = |\nabla E_V(0)| C_d \exp\left(-2 \int_0^l k dx\right) \left[\left(1 + \exp\left(\frac{\epsilon - E_{Fn}(l)}{kT(l)}\right)\right)^{-1} - \left(1 + \exp\left(\frac{\epsilon - E_{Fp}(l)}{kT(0)}\right)\right)^{-1} \right] \quad (5.15)$$

在 S-Device 中,已经开发了一种适用于涉及非均匀电场和突变/渐变异质结的任意隧

穿势垒的非局部带带隧穿模型^[29]。采用动态非局部路径带带隧穿模型,该模型实现了由直接和声子辅助的带间隧穿过程引起的电子和空穴的非局部产生^[30]。如果导带波谷之间的能量差很小,则直接和声子辅助隧穿过程都很重要。隧穿路径是根据能带的梯度动态确定的。该模型考虑了直接隧穿和声子辅助隧穿过程,在均匀电场内,它精确地简化为 Kane 和 Keldysh 的模型^[30-31]。式(5.15)显示了空穴直接隧穿的复合率,与传统的非局部隧穿模型相似。

5. 量子效应模型(修正载流子密度)

经典半导体器件方程式表明,移动载流子、电子和空穴的行为类似于半导体中的经典粒子。对于较大的器件尺寸,此假设给出了很好的结果,但是对于较小的器件几何形状,量子效应(如前文所述的量子隧穿)和量子力学限制变得非常重要。后一种效应导致 Si-SiO₂ 界面附近电子和空穴的允态减少。在使用经典的 DD 模型器件仿真中,经计算的 N 型沟道 MOSFET 的沟道中电子浓度的峰值被计算为直接位于 Si-SiO₂ 界面。该计算方式是不准确的,因为在界面附近,允态的数量急剧减少,因此载流子浓度的峰值距离界面几埃。

1) 量子约束

对于 NBTI 的建模,量子限制模型的使用降低了靠近 Si-SiO₂ 界面的载流子浓度,并且可能对所使用的 NBTI 模型产生重大影响。在经典的器件模拟器中,通过使用其他量子校正模型来解决量子限制问题。这些模型局部地改变载体的态密度^[32-33]或修改靠近界面处的材料导带边缘^[34]。

在经典器件仿真中,建模时材料的态密度(Density Of States, DOS)在整个器件中为恒定值。为了描述在考虑量子效应限制后 Si/SiO₂ 界面处 DOS 与距离的关系,提出以下模型^[32,35]:

$$h_{\text{cor}} = 1 - \exp\left(-\frac{(z+z_0)^2}{\zeta^2 \lambda_{\text{TH}}^2}\right) \quad (5.16)$$

其中, z 是到 Si-SiO₂ 界面距离; z_0 是相对于所述 Si-SiO₂ 界面整个函数偏移量,是一个新引入的参数,可以进行 λ_{TH} 变化以进行校准。符号 λ_{TH} 表示热波长,且

$$\lambda_{\text{TH}} = \frac{\hbar}{\sqrt{2m^* k_B T}} \quad (5.17)$$

其中, \hbar 为约化普朗克常数; m^* 是有效载流子质量; k_B 玻尔兹曼常数, T 为温度。且有

$$N_c = N_{c,0} \cdot h_{\text{cor}} \quad (5.18)$$

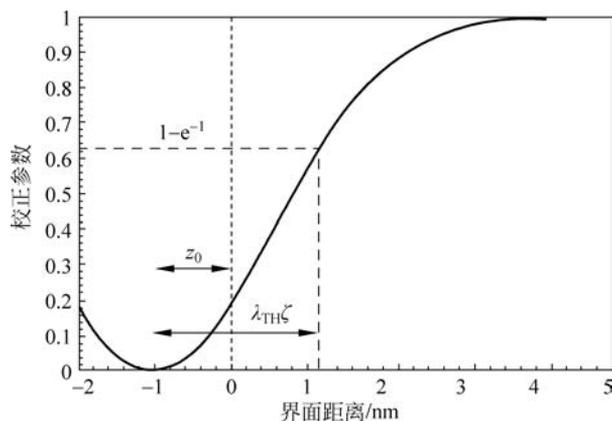
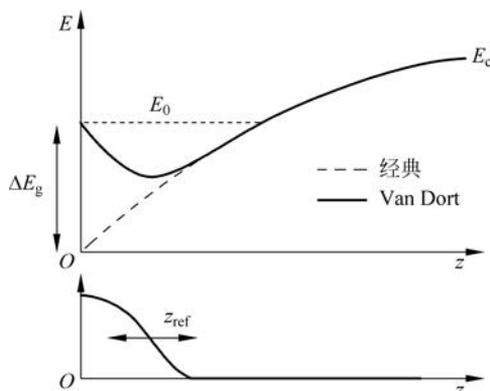
图 5.15^[36] 中示意性地描述了不同参数的相互作用。参数 z_0 很重要,当 $z_0=0$, DOS 在界面处变为零。这将导致数值问题并降低数值求解器的收敛性。由于考虑了波函数的渗透,正数 z_0 将校正函数向介质移动。 $\zeta \lambda_{\text{TH}}$ 定义了校正的有效深度。当 $\zeta > 1$ 可以实现的高值导致 DOS 的减少,即使在衬底深处也是如此^[36]。

2) 导带边缘校正

基于 E_0 三角形能量阱的替代方法的一个特征值如图 5.16 所示,该模型由 Van Dort^[34] 提出:

$$\Delta E_g = E_0 - E_c(0) = \frac{13}{9} \beta \left(\frac{\epsilon_{\text{si}}}{4qk_B T} \right)^{1/3} |E_n|^{2/3} \quad (5.19)$$

其中, $\beta = 4.1 \times 10^{-8} \text{ eV} \cdot \text{cm}$ 是从高掺杂水平下观察到的阈值电压偏移中找到的比例因子^[33]; ϵ_{si} 是硅的介电常数; E_n 是垂直于 Si-SiO₂ 界面的界面处的电场。

图 5.15 DOS 校正参数 h_{cor} ($z_0=1$ 和 $\zeta=1$, 温度 $T=300\text{K}$)图 5.16 Si/SiO₂ 界面处的带边缘弯曲, 用因子 $\Delta E_g F(z)$ 校正经典的能带边缘^[36]

ΔE_g 值乘以距离相关的权重函数(该函数已由 Selberherr^[37] 引入)用于建模 MOSFET 中的表面粗糙度。该函数具有以下形式:

$$F(z) = \frac{2 \exp \left[- \left(\frac{z}{z_{\text{ref}}} \right)^2 \right]}{1 + \exp \left[- 2 \left(\frac{z}{z_{\text{ref}}} \right)^2 \right]} \quad (5.20)$$

其中, z_{ref} 是界面距离的比例因子。因此, 用 Van Dort 对经典能带边能量 E_{class} 进行量子校正得到的带边能量如下:

$$E_c = E_{\text{class}} + F(z) \Delta E_g \quad (5.21)$$

图 5.17 描绘了 FinFET 中“鳍”的一维剖视图, 显示了在相同偏置条件下不同模型的载流子浓度。可以发现 DOS 校正模型定性地提供了更好的结果。

虽然 DOS 校正模型可以产生合理的结果, 但是由于它不能解决平带弯曲问题, 因此必须针对每个偏置点进行校正。Van Dort 的模型无法重现沟道中的载流子浓度, 可能是由于假设了三角能阱。对于极细的沟道, 此假设是过于粗略的估计。因此, 这些模型可以很好地用于描述非常薄的沟道器件中的电流减小。

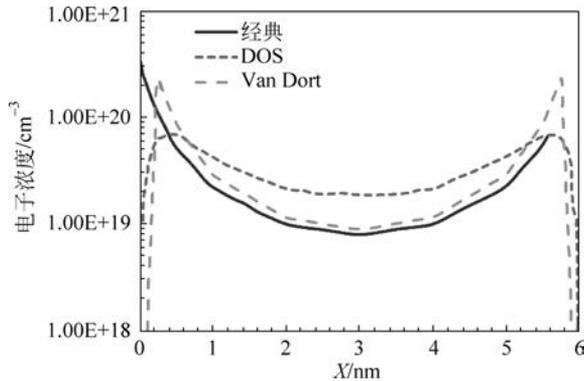


图 5.17 使用经典的器件模拟和约束校正模型, 整个鳍片上的电子浓度^[36]

5.3 2D NAND Flash 器件模拟

5.3.1 2D NAND Flash 器件模型

关于 NAND Flash 模拟仿真, 基础模型主要包含基本器件模拟所考虑的泊松方程、能带、介电常数、漂移扩散模型等, 根据实际情况还需考虑如迁移率模型、产生-复合模型等。NAND Flash 器件仿真中还增加了一些模型, 如一般浮栅存储器中的浮动金属栅、浮动半导体栅、电荷守恒等模型。电荷在浮栅中会导致表面电场的再分布, 同时要考虑陷阱模型。陷阱是局域态, 可以表现为界面的悬挂键, 也可以是体材料中的悬挂键、断键、键扭曲、陷阱空或者满、电子陷阱或空穴陷阱、电子或空穴捕获率、陷阱能级、浓度、陷阱类型等。

根据陷阱位置, 可分为界面陷阱模型、体陷阱模型以及电荷俘获存储器的氮化硅中的转移特性模型。在具体的操作过程中还会涉及隧穿模型、退化模型、NBTI 模型以及物理模型接口 (Physical Model Interface, PMI) 模型等。虽然已经在一定程度上进行了简化, 但对于 NAND Flash 来说, 仿真还是十分复杂的。

1. 隧穿模型

Flash 器件可以分为 NOR 型和 NAND 型, NOR 型器件使用沟道热载流子注入 (Hot Carrier Injection, HCI) 方式进行编程操作^[38]; 而 NAND 型则使用 F-N 隧穿方式实现编程和擦除操作^[38]。在 S-Device 仿真模块中, 隧穿分为越过势垒的隧穿和热载流子注入隧穿。图 5.18 列出了 S-Device 仿真模块中的各种隧穿模型^[39]。

对于 NOR 型存储器的编程操作仿真需要使用热载流子注入隧穿模型。势垒隧穿模型要比热载流子注入模型复杂, 包括 P-F 发射 (Poole-Frenkel Emission) 模型^[39] 和非局部 (Nonlocal) 隧穿模型^[40]。P-F 发射模型主要应用于电荷俘获型存储器中保持特性的仿真。Nonlocal 隧穿模型为用途最广的隧穿模型, 应用于浮栅存储器的仿真、电荷俘获型存储器仿真、肖特基接触仿真、异质结仿真等。直接隧穿模型包含于 Nonlocal 隧穿模型内, 主要应用于栅漏电流的仿真。F-N 隧穿是简单的隧穿模型也包含于 Nonlocal 隧穿模型内, 常应用于三角形势垒隧穿仿真。

Nonlocal 隧穿模型同时考虑了载流子温度, 可以提供多种不同的隧穿概率近似计算方

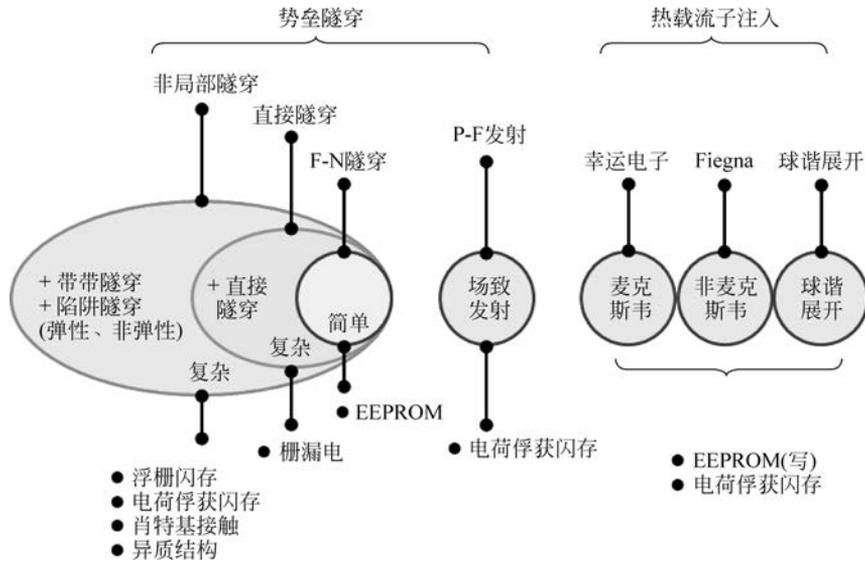


图 5.18 隧穿模型及主要应用介绍

法。Nonlocal 隧穿模型适用于各种势垒的隧穿情况，例如，肖特基接触、异质结、薄栅漏电、叠层等发生的隧穿^[41]。图 5.19 展示了 Nonlocal 隧穿模型适用的多种隧穿过程。对绝缘体势垒的隧穿包括 F-N 隧穿、直接隧穿、带带隧穿、陷阱辅助的弹性隧穿和陷阱辅助的非弹性隧穿^[42-48]。这些隧穿主要发生在存储器编程和擦除操作过程中。半导体内部的带带隧穿按照隧穿方向分为垂直沟道方向和平行沟道方向。电荷泵测量操作时会发生垂直沟道的带带隧穿，此时界面态被俘获的电荷会先发生 F-N 发射。平行沟道方向的带带隧穿发生在 3D NAND Flash 中，这会给 3D NAND Flash 带来影响，既要利用又要抑制这种隧穿。在 BiCS 结构中使用 GIDL 擦除方式时，由于漏极空穴为少数载流子，需要利用带带隧穿产生大量空穴实现擦除操作。另外，在编程抑制时，沟道电势发生自抬升，在上下选择管发生带带隧穿，导致沟道电势下降。这种带带隧穿可以是直接带带隧穿也可以是陷阱辅助带带隧穿。

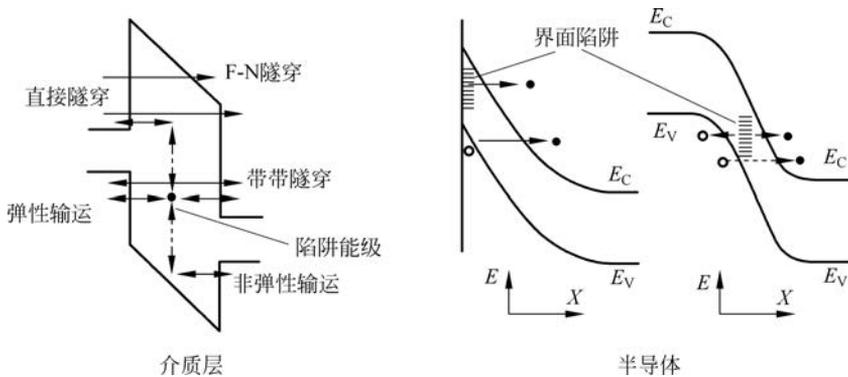


图 5.19 Nonlocal 隧穿模型适用的各种隧穿过程

电荷俘获存储器的栅堆栈结构包括半导体的衬底、半导体或金属的栅电极和具有高缺陷密度的存储层。这三个功能层通过隧穿氧化层和阻挡层进行隔离，电荷从衬底进入存储层及从存储层到栅电极都是通过隧穿的方式进行。以载流子从衬底通过隧穿氧化层进入

存储层为例,无论是电子还是空穴,在栅介质叠层结构中都是通过势垒隧穿对器件进行编程和擦除操作,可能发生的隧穿类型包括 F-N 隧穿、直接隧穿(Direct Tunneling,DT)、修正的 F-N(Modified Fowler-Nordheim, MFN)隧穿和陷阱辅助隧穿(Trap Assisted Tunneling, TAT)。当介质内的电场满足特定的条件时发生相应的隧穿,如图 5.20 所示, d_{ox} 为隧穿氧化层厚度, q 为电子电荷, $q\phi_1$ 为隧穿氧化层与衬底之间的导带漂移差, $q\phi_2$ 为隧穿氧化层与存储层之间的导带漂移差。当隧穿氧化层的电场很大时,高电场会导致形成三角形势垒,发生 F-N 隧穿,电荷穿过氧化层。当电场继续减小,电荷的隧穿路径由三角形转变为梯形,形成直接隧穿。电场继续减小,电荷的隧穿路径将变为两层——第一层隧穿氧化层的梯形势垒和第二层存储层的三角势垒,则形成修正的 F-N 隧穿。当电场足够低,而且第二层介质层较厚且存在可以辅助隧穿的陷阱能级时,将发生陷阱辅助隧穿。

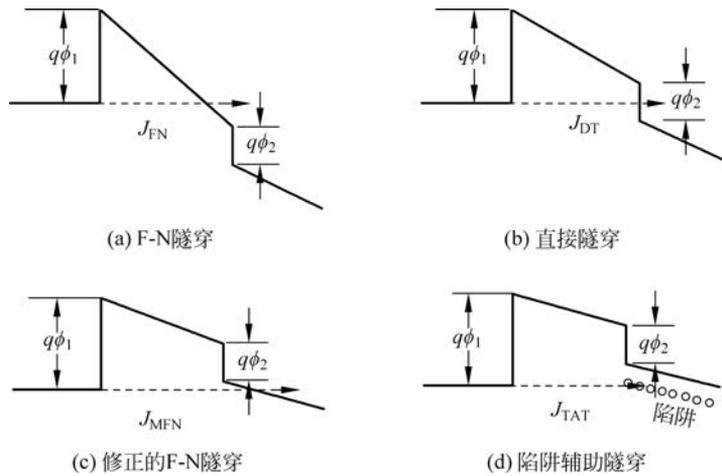


图 5.20 SONOS 结构在不同电场下的导带图

1) WKB 近似

WKB(Wentzel-Kramers-Brillouin)近似是一种广泛应用于大规模数值计算的求解隧穿概率的方法,下面介绍其在 F-N 隧穿和直接隧穿计算中的应用。

WKB 近似是对一维薛定谔方程的一种近似求解方法,在半导体数值模拟计算中通常用来对隧穿电流进行计算求解。通常认为能量为 E 的粒子穿过势能为 $V(x)$ 的区域时, $V(x)$ 保持恒定。当 $V(x)$ 不是一个常量,但相比于振荡函数的波长变化非常缓慢时,对于给定区域仍可以假设势能基本保持不变,且波函数仍然保持正弦形式。对于电荷俘获存储器而言,在栅堆栈隧穿势垒变化较为平缓的情况下,采用 WKB 近似是首选隧穿电流计算的近似方法。在该近似中,隧穿系数可表示为

$$TC(\xi) = \exp\left(-\frac{2}{\hbar} \int_{x_1}^{x_2} \sqrt{2m(W(x) - \xi)} dx\right) \quad (5.22)$$

其中, ξ 表示隧穿电荷的能量; $W(x)$ 为在介质中隧穿距离和电场的函数,积分仅当 $W(x) - \xi > 0$ 是有效。因为 $W(x)$ 和电场有关,所以在 F-N 隧穿、直接隧穿以及修正的 F-N 隧穿中, $W(x)$ 的变化导致了隧穿系数的变化。

2) F-N 隧穿

当隧穿介质层上所加的电场比较大的情况下: $E_{ox} \geq \phi_1/d_{ox}$,电荷的隧穿区域为三角形

势垒,如图 5.20(a)中所示, $W(x) > q\phi_1 - qE_{ox}d_t$,积分的上限 $x_2 = (q\phi_1 - \epsilon)/qE_{ox}$,因此,F-N 隧穿概率可表示为

$$TC(\xi) = \exp\left(-\frac{2}{\hbar} \int_0^{x_2} \sqrt{2m_{ox}(q\phi_1 - qE_{ox}x - \xi)} dx\right) \quad (5.23)$$

对式(5.23)积分可得

$$TC(\xi) = \exp\left(-4 \frac{\sqrt{2m_{ox}}}{3q\hbar E_{ox}} (q\phi_1 - \xi)^{\frac{2}{3}}\right) \quad (5.24)$$

隧穿电流密度 J_{FN} 正比于传输系数,可表示为

$$J_{FN} = \frac{m_0 q^3}{8\pi m_{ox} q h \phi_1} E_{ox}^2 \exp\left(-4 \frac{\sqrt{2m_{ox}}}{3q\hbar E_{ox}} q\phi_1^{\frac{2}{3}}\right) \quad (5.25)$$

其中, m_0 为自由电子的质量; h 为普朗克常数; m_{ox} 为隧穿氧化层中电子的有效质量。

3) 直接隧穿

当电场强度 $(\phi_1 - \phi_2)/d_{ox} < E_{ox} < \phi_1/d_{ox}$,如图 5.20(b)所示,电荷发生隧穿的势垒为梯形,隧穿系数为

$$TC(\xi) = \exp\left(-\frac{2}{\hbar} \int_0^{d_{ox}} \sqrt{2m_{ox}(q\phi_1 - qE_{ox}x - \xi)} dx\right) \quad (5.26)$$

进一步积分得到

$$TC(\xi) = \exp\left(-4 \frac{\sqrt{2m_{ox}}}{3q\hbar E_{ox}} \left((q\phi_1 - \xi)^{\frac{3}{2}} - (q\phi_1 - qE_{ox}d_{ox} - \xi)^{\frac{3}{2}}\right)\right) \quad (5.27)$$

直接隧穿电流

$$J_{DT} = \frac{m_0 q^3 \exp\left(-4 \frac{\sqrt{2m_{ox}}}{3q\hbar E_{ox}} \left((q\phi_1 - \xi)^{\frac{3}{2}} - (q\phi_1 - qE_{ox}d_{ox} - \xi)^{\frac{3}{2}}\right)\right)}{8\pi m_{ox} h (\sqrt{q\phi_1} - \sqrt{q\phi_1 - qE_{ox}d_{ox}})^2} E_{ox}^2 \quad (5.28)$$

4) 修正的 F-N 隧穿

当电场 $(\phi_1 - \phi_2)/(d_{ox} - \gamma d_N) < E_{ox} < (\phi_1 - \phi_2)/d_{ox}$ 时,发生的隧穿为修正的 F-N 隧穿,其中 $\gamma = k_{SiN}/k_{SiO_2}$, γ 为存储层与隧穿层介电常数之比, d_N 为存储层厚度。如图 5.20(c)所示,当发生修正的 F-N 隧穿时,电子隧穿通过的势垒为隧穿氧化层的梯形势垒和存储层的三角势垒,而后进入存储层的导带。修正的 F-N 隧穿发生时,隧穿系数近似等于直接隧穿通过氧化层的隧穿系数与通过存储层三角形势垒的 F-N 隧穿系数的乘积。

$$TC(\xi) = \exp\left\{-4 \frac{\sqrt{2m_{ox}}}{3q\hbar E_{ox}} \left[(q\phi_1 - \xi)^{\frac{3}{2}} - (q\phi_1 - qE_{ox}d_{ox} - \xi)^{\frac{3}{2}}\right]\right\} \cdot \exp\left\{-4 \frac{\sqrt{2m_{ox}}}{3q\hbar E_{ox}} \left[\frac{\epsilon_n}{\epsilon_{ox}} \sqrt{\frac{m_n}{m_{ox}}} (-qE_{ox}d_{ox} - \xi)^{\frac{3}{2}}\right]\right\} \quad (5.29)$$

隧穿电流为

$$J_{MFN} = \frac{m_0 q^3 \exp\left\{-4 \frac{\sqrt{2m_{ox}}}{3q\hbar E_{ox}} \left[(q\phi_1 - \xi)^{\frac{3}{2}} - (q\phi_1 - qE_{ox}d_{ox} - \xi)^{\frac{3}{2}}\right] + \frac{\epsilon_n}{\epsilon_{ox}} \sqrt{\frac{m_n}{m_{ox}}} (-qE_{ox}d_{ox} - \xi)^{\frac{3}{2}}\right\}}{8\pi m_{ox} h \left[\sqrt{q\phi_1} - \sqrt{q\phi_1 - qE_{ox}d_{ox}} + \frac{\epsilon_n}{\epsilon_{ox}} \sqrt{\frac{m_n}{m_{ox}}} (-qE_{ox}d_{ox} - \xi)\right]^2} E_{ox}^2 \quad (5.30)$$

5) 陷阱辅助隧穿

当电场满足 $E_{ox} \leq (\phi_1 - \phi_2 - \phi_t) / d_{ox}$ 时, 发生陷阱辅助隧穿, 如图 5.20(d) 所示。其中 ϕ_t 表示陷阱能级的深度。在通常的电荷俘获存储器件编程/擦除操作过程中, 电场强度一般不会引发陷阱辅助隧穿, 且其需要考虑的计算过程非常复杂, 在模拟中将其忽略并不会对模拟结果造成太大的影响。

6) P-F 发射模型

P-F 发射模型经常用于解释电介质和非晶薄膜中的传输效应^[49]。该模型预测带电陷阱中心发射概率的变化, 其中介质层的势垒高度会随着外部电场的增强而降低。图 5.21 展示了在 MIS 结构中, 当半导体端施加正压, 由于介质层的能带发生倾斜, 介质层中被俘获的电荷将会发射出来。随着电场增强, P-F 系数逐渐增大(见图 5.22), 发射概率增大。

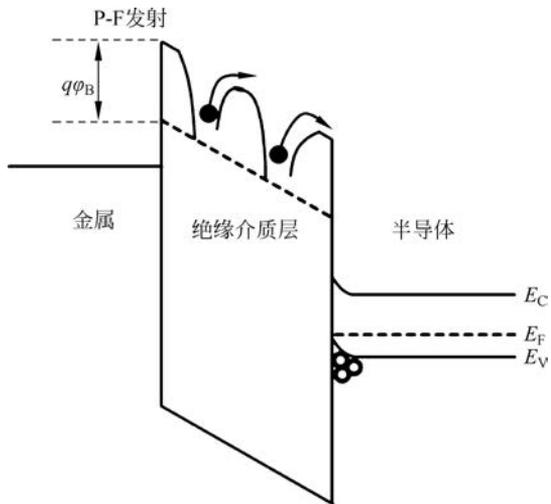


图 5.21 P-F 发射模型

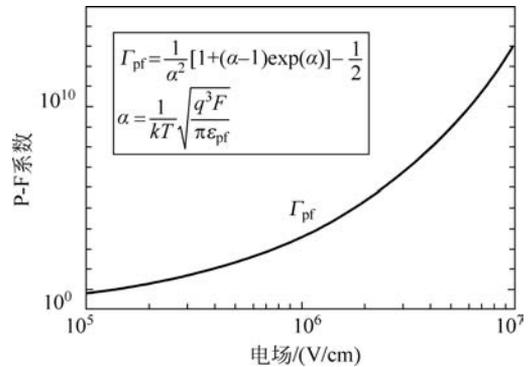


图 5.22 P-F 系数与电场间的关系

在 Flash 仿真中, 该模型应用于载流子在氮化硅存储材料中运输过程的模拟。P-F 过程是存储器编程机制开始时的传导限制机制, 因为隧穿提供了更高的电流。随着捕获的电子增加, 开始屏蔽电场, 修正的 F-N 隧穿成为主要限制机制。此时, 氧化物-氮化物界面处的俘获电荷密度与流过它的 P-F 电流的积分成正比^[50]。随着存储器编程和擦除循环次数的增加, 由于氮化物的体电导率增加, 保持特性将会出现退化。

2. 陷阱模型

在 Flash 中, 电荷存储在单元的电荷存储层中。氮化硅是目前应用最为广泛的电荷存储层材料。在制造过程中, 氮化硅材料内部会产生大量的缺陷。这些原子结构上的缺陷会在氮化硅的禁带中引入陷阱, 电荷存储层中的电荷就存储在这些陷阱中。因此, 氮化硅材料中电荷陷阱的性质对于 3D NAND Flash 器件的性能有着重要影响。本节对在 3D NAND Flash 器件的模拟和仿真中使用的电荷陷阱模型进行介绍。

1) 电荷陷阱的能量分布

从能带的角度来看, 陷阱实际上是一个电子态, 其能级位于氮化硅材料的禁带之中, 不

同缺陷所引起的陷阱能级在禁带中的位置存在差异。在氮化硅材料(禁带宽度约为 5eV)中,如果陷阱能级较浅,与导带底的距离小于 1eV,这样的陷阱属于浅能级陷阱(shallow trap)。当陷阱能级与导带的距离大于 1eV 时,这样的陷阱则归类为深能级陷阱(deep trap)。在实际的电荷存储层中,由于不同类型的缺陷在数量上存在很大差异,因此不同能级的陷阱的浓度也不同。电荷陷阱的能量范围以及陷阱浓度(即态密度 DOS)在能量范围内的变化情况,共同构成了电荷陷阱的能量分布。

电荷陷阱的能量分布对于 Flash 器件的性能有极其重要的影响。首先,根据费米统计分布,不同能级的陷阱被电荷占据的概率不同,因此陷阱的能级分布最终决定了编程后存储单元俘获电荷的浓度。其次,陷阱能级与导带的距离决定了其俘获的电荷进入导带或价带所需能量的大小,电荷陷阱的能量分布会影响存储电荷在电荷存储层内部的运动。最后,陷阱能级一定程度上决定了电荷在隧穿过程中面对的势垒,因此电荷陷阱的能量分布会影响隧穿电流的大小。总之,电荷陷阱的能量分布对于 Flash 器件的编程/擦除、保持特性、耐久性等方面都有着重要的影响。

但是,由于在氮化硅材料的制造工艺和陷阱的表征方法上存在差异,因此对于氮化硅材料电荷陷阱能量分布的研究并没有统一的结论。在电荷俘获型存储单元发展的早期,Kapoor 等通过光电效应的方法,对 LPCVD 形成的氮化硅材料电荷陷阱的能级分布进行了研究,发现在氮化硅中存在 5 种不同的陷阱能级^[51]。Toshiyuki 等则利用雪崩电荷注入和电容-电压曲线测量的方法,通过对不同厚度的结构进行研究,发现电子陷阱分布在距导带 0.9~1.7eV 范围内^[52]。此外,还有研究者使用深能级瞬态谱方法对氮化硅和 SONOS 栅堆叠结构中的电荷陷阱进行类似的研究^[53-54]。

除了实验,还有研究者利用理论计算与计算机仿真技术对氮化硅材料电荷陷阱的能级进行了研究。Gritsenko 等通过理论计算分析了与氢结合的氮空位(V_N-H)这一缺陷引起的陷阱能级^[55],并将理论计算结果与其他文献中的实验数据进行了对比,如图 5.23 所示。左侧为实验数据,右侧为理论计算结果,包括电子陷阱和空穴陷阱两种情况^[55]。J. Wu 等则通过第一性原理计算的方法,研究了氮化硅材料中 6 种主要类型的缺陷引起的电荷陷阱的能级^[56]。这些陷阱包括:①氮空位(V_N);②硅取代氮的替位原子(Si_N);③与氢结合的氮空

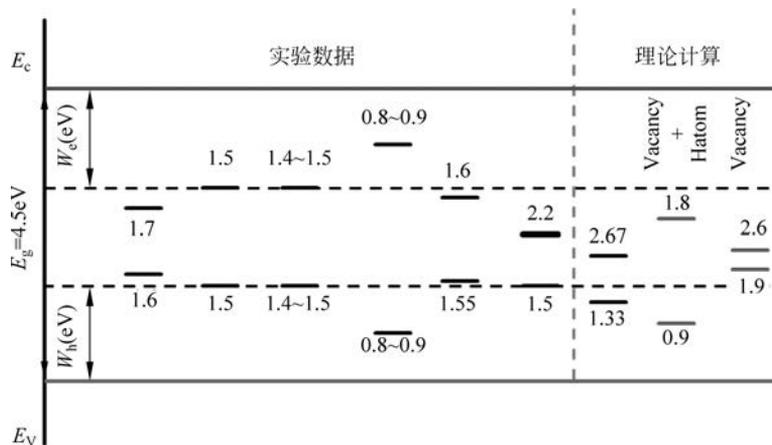


图 5.23 由 V_N-H 缺陷引起的电子陷阱和空穴陷阱的能级^[55]

位(V_N-H)；④与氢结合的硅替位原子(Si_N-H)；⑤与氧结合的氮空位(V_N-O)；⑥与氢氧结合的氮空位(V_N-OH)。图 5.24 给出了这些缺陷的结构,图 5.25 是与这些缺陷相对应的陷阱能级。研究表明,不同的缺陷所引起的陷阱能级差异较大,同时有些缺陷可以引起多个陷阱能级。其中,存在一些距离导带非常近的超浅能级陷阱,这些陷阱可能会引起存储电荷的快速电荷泄漏(fast charge loss)问题^[57]。

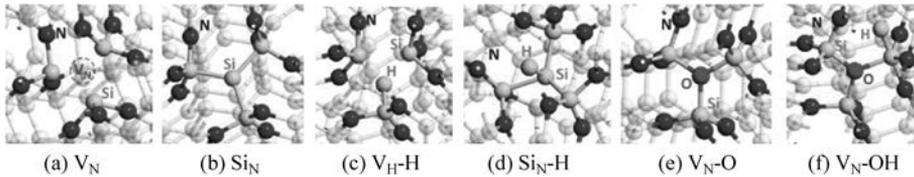


图 5.24 氮化硅材料中的 6 种主要缺陷^[56]

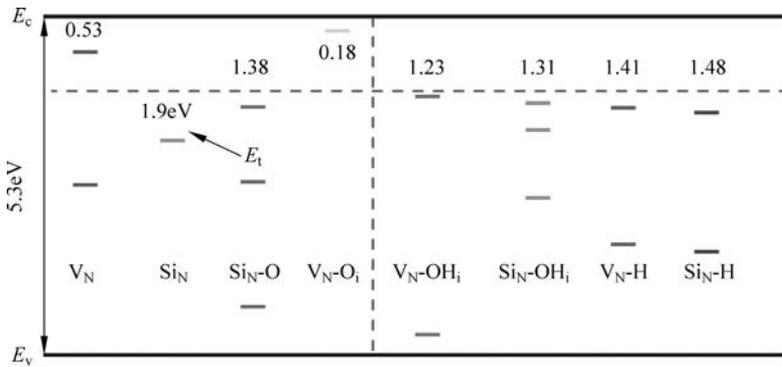


图 5.25 氮化硅中不同陷阱能级分布情况^[56]

以上这些研究结果反映出,在氮化硅材料中,由于缺陷有多种类型,缺陷的产生又与工艺过程密切相关,因此电荷陷阱的能量分布具有多样性和复杂性。在 Flash 器件的模拟和仿真中,对于电荷陷阱的能量分布,往往采用一定的简化模型进行处理,以兼顾仿真准确性与仿真速度。根据由简单到复杂的顺序,这些电荷陷阱能量分布模型包括单一能级分布、多单一能级分布、均匀分布、指数分布和高斯分布。单一能级分布中,陷阱只分布在一个固定的能级上,多单一能级分布则由多个单一能级分布组合而成。这两种分布属于离散分布。均匀分布、指数分布和高斯分布均属于连续分布。陷阱的态密度在一定的连续的能量范围发生变化,它们的差异就在于态密度随能量变化的规律不同。均匀分布中,陷阱在整个能量范围上的态密度为定值,而另外两种分布中,态密度分别遵循指数规律和高斯规律变化,具体的分布可以通过调整相关的参数进行控制。图 5.26 给出了除多单一能级分布以外的其他 4 种分布的示意。

在仿真中,理论上越复杂的模型所仿真出来的结果就越准确,但考虑到仿真的速度,使用较为简单的分布模型也可以得到满足要求的仿真结果。在一般的仿真中,单一能级分布、多单一能级分布和均匀分布是应用最为广泛的。

2) 陷阱的电性模型

陷阱在俘获电荷后,自身的电性会产生相应的变化。在 Flash 的模拟和仿真中,对于陷阱的电性有两种处理方式,即存在两种电性模型,分别称为单性陷阱模型和两性陷阱模型。

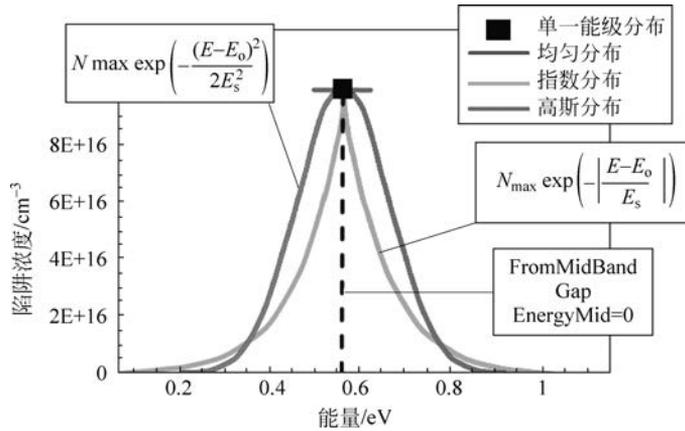


图 5.26 4 种陷阱能量分布

在单性陷阱模型中,陷阱被划分为两种类型:一种在俘获电子时带负电,不被电子占据时不显电性,称为电子陷阱;另一种在俘获空穴(不被电子占据)时带正电,不被空穴占据(俘获电子)时不显电性,称为空穴陷阱,因此电子陷阱又称为受主陷阱(acceptor),空穴陷阱又被称为施主陷阱(donor)。两种陷阱俘获和发射电荷的过程以及电性变化的情况如图 5.27 和图 5.28 所示。其中,电子陷阱的两种电性的状态分别用 A_0 和 A_- 来表示,空穴陷阱的两种电性的状态分别用 D_0 和 D_+ 来表示。

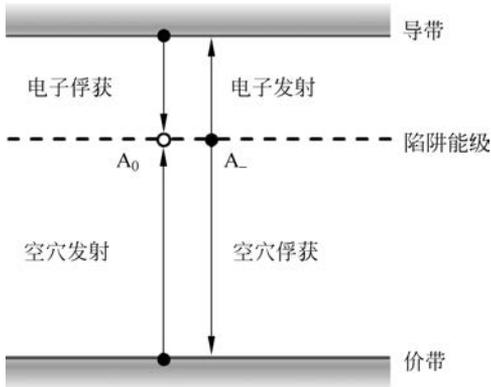


图 5.27 电子陷阱对电荷的俘获和发射以及电性变化

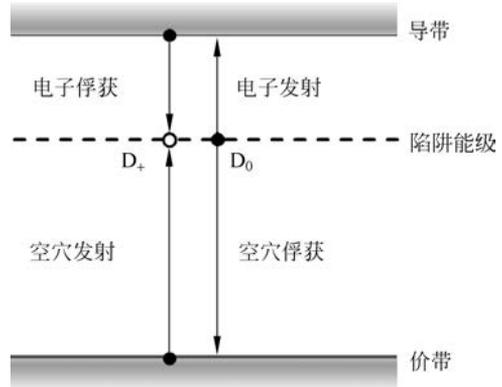


图 5.28 空穴陷阱对电荷的俘获和发射以及电性变化

而在两性陷阱模型中,不再按照电性的变化将陷阱分为两类进行处理,而是统一为一类陷阱,对陷阱的行为统一进行描述。在两性陷阱模型当中,每一个陷阱均存在三种不同的电性状态,分别用 D^- 、 D^0 和 D^+ 来表示。 D^- 状态表示陷阱带一个负电荷, D^0 状态表示陷阱能级为电中性状态, D^+ 状态表示陷阱能级带一个正电荷。 D^- 态的陷阱释放一个电子或者俘获一个空穴可以使其变为 D^0 态,反之亦然。同样 D^0 和 D^+ 态之间也可以进行类似的转化。三者的相互转化关系可以用图 5.29 进行表示。

对于两性陷阱来说,其既可以与一个电子作用,也可以与一个空穴作用。在 D^0/D^+ 状态之间转换的过程即相当于单性陷阱模型中的施主陷阱,而在 D^0/D^- 状态之间转换的过程即相当于受主陷阱。但与单性陷阱模型中的施主和受主陷阱不同的是,两性陷阱模型中的

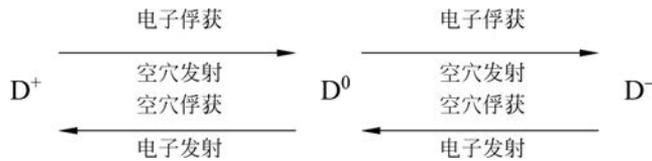


图 5.29 两性陷阱模型中三种状态的转化

陷阱同时具有施主和受主陷阱的性质。当一个中性状态(D^0)的陷阱变为 D^- 态的同时,一个施主型的陷阱也同时消失了;反之,当一个中性状态(D^0)的陷阱变为 D^+ 的同时,一个受主型的陷阱也同时消失了。也就是说,在两性陷阱模型当中,施主型陷阱和受主型陷阱的数量不是独立的,而是会互相转化。在电荷存储层中,电荷的运动需要考虑复合过程,包括导带中电子与陷阱中空穴的复合、价带中空穴与陷阱中电子的复合以及陷阱中的电子和陷阱中的空穴之间的复合,同时要考虑复合过程对器件编程特性的影响。在单性陷阱模型中,需要 Shockly-Read-Hall 模型来描述电荷的复合过程,但是对于两性陷阱模型来说,其本身就将电子与空穴的相互影响和耦合考虑在内,所以不需要考虑额外的复合效应。

5.3.2 2D NAND Flash 器件操作及可靠性模拟

对于 2D 电荷俘获存储单元的编程/擦除操作,需要重点考虑的是能带调制工程,编程时的能带图如图 5.30(a)所示,擦除时如图 5.30(b)所示。编程时电子的隧穿经过隧穿层后被电荷存储层俘获,但也会有一部分电子可能会继续隧穿通过氧阻挡层。当进行擦除操作时,空穴从沟道中进入,绝大部分会被电荷存储层俘获实现擦除,但也有一部分会继续隧穿通过阻挡氧层。对于电荷俘获存储单元,最理想的情况是,隧穿层和阻挡层中不存在陷阱,电荷存储层有比较深的陷阱能级,这样就可以最大限度地实现存储的功能。

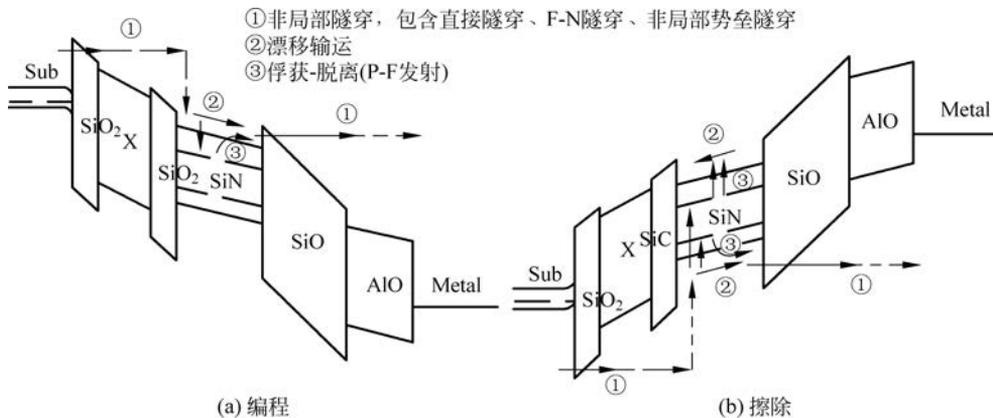


图 5.30 CTF 的编程和擦除操作

电荷俘获存储单元的关键是能带的调控。实际应用中,调控的重点是隧穿层,如果隧穿层采用纯氧化层,空穴面临的势垒高度和宽度都非常大,因此擦除操作的实现就变得非常困难。考虑采用三明治结构,把中间 X 的势垒降低,这样空穴需要穿过的势垒显著降低,擦除效率可以显著提高。

对于半导体而言,主要的隧穿机制是直接隧穿中的带带隧穿;对于绝缘介质层而言,主要的隧穿机制是间接隧穿中的陷阱辅助隧穿,该隧穿会造成 GIDL 效应。对于金属硅化物

(silicide)而言,肖特基隧穿是主要的隧穿方式,也是主要漏电机理,在肖特基模型中,除了上述所说的隧穿模型之外还有一个很显著的现象:费米能级钉扎效应,由于界面存在很多陷阱,会导致费米能级被钉扎,也会增加能带调制的难度。

对于浮栅器件,在沟道和隧穿层界面、氧化层和氮化硅中都存在陷阱,其中,界面陷阱(N_{it})主要影响亚阈值摆幅、阈值电压等,此外,一部分陷阱会引发库仑散射,从而导致沟道电流下降。同时,这些陷阱还会影响到高温存储(High Temperature Storage, HTS),在高温下,一部分界面态会被修复,影响保持特性。氧化物陷阱(N_{ot})会导致阈值电压变化,此外,编程/擦除的阈值电压的窗口也会减小甚至关闭,单元的电流也会下降,并伴有 HTS 的问题。

器件耐久性的退化主要源于编程/擦除期间电荷隧穿对栅介质的损伤。研究显示^[58],造成耐久性退化的机制主要分为以下 5 种:①电子注入隧穿氧化层;②空穴注入氧化层;③器件阳极附近电子-空穴对的产生,其中一部分空穴注入隧穿层导致空穴陷阱的产生;④在氧化层中注入的电子和被陷阱束缚的空穴复合,这会产生电子陷阱;⑤空穴脱离陷阱束缚时释放的氢原子则会运动到界面处导致 Si-H 键的断裂,产生界面陷阱^[59]。

接下来考虑电荷俘获存储单元的数据保持特性,电荷俘获存储单元的数据保持特性模型和浮栅单元类似,主要的电荷流失路径依旧是隧穿层以及阻挡层,如图 5.31 所示共有 4 种方式:①P-F 的陷阱俘获和发射模型;②漂移扩散;③包括 DT、F-N、修正的 F-N 等势垒隧穿;④陷阱辅助隧穿。数据保持特性和温度、陷阱能级等都息息相关。一般来说,随着温度的升高,数据保持特性会明显降低,陷阱越深,保持特性越好。

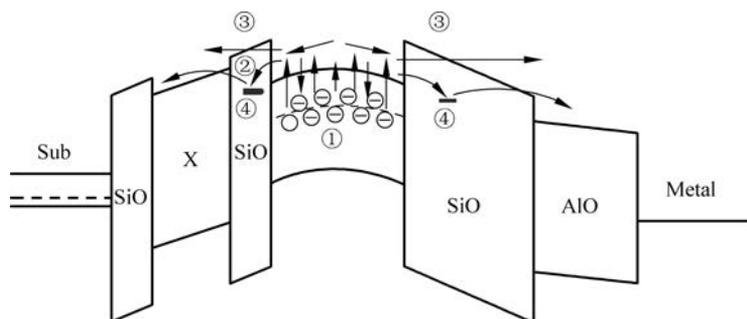


图 5.31 CTF 的数据保持特性

在数据保持特性中,电荷分别经过隧穿层和阻挡层两条路径流失。因此,可以通过仿真研究具体的栅介质叠层结构中电荷流失的主要路径,从而探索降低电子流失的解决方法。仿真结果显示如图 5.32 所示,电荷的流失主要是通过隧穿层的损耗,因此需要进一步优化隧穿层结构。

在 2D NAND Flash 模拟中,还需要考虑 STI 结构。STI 形状不同,电场的分布也不同,这便需要借助仿真来进行分析。根据高斯定理,该点的曲率越小,电场就会更强,电子的聚集就越明显。对于 $ACD=10\mu\text{m}$ 和 $ACD=100\text{nm}$,从图 5.33 中可以看出,尺寸对电性的影响十分明显,尺寸越小,性能越差。随着器件尺寸的缩小,出现了越来越多的问题,这时就需要考虑 3D 存储器模型。

氧化层的陷阱的产生主要有以下几种方式。

(1) 键的断裂,比如高能载流子轰击。

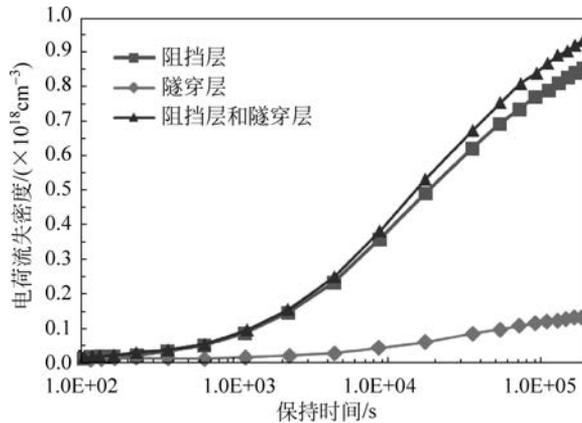


图 5.32 通过隧穿层与阻挡层的电荷流失对比

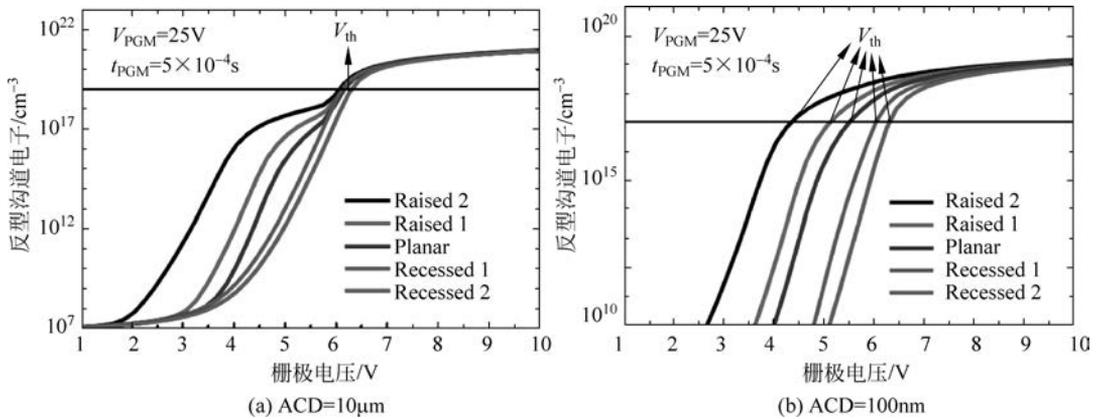


图 5.33 ACD=10 μ m 和 100nm 时,沟道电子聚集速度对比

(2) 阳极空穴注入, 高能的载流子引入空穴和共价键结合, 由此降低键能, 产生陷阱。

(3) 在存储器工作的过程中也会产生陷阱, 这是由于高电场或高能载流子的引入产生的。

陷阱的产生会对存储器的编程/擦除操作有很大影响, 如电子陷阱的产生会使能带向上弯曲, 如图 5.34 所示。会降低电子的隧穿概率, 使单元阈值电压下降, 同时由于其所带负电荷的影响, 会造成阈值电压上升, 因此在编程操作中, 阈值电压基本保持不变。但对擦除操作而言, 阈值电压会增加, 总阈值电压窗口会缩小, 反复编擦之后存储窗口会逐渐变小, 保持特性会受到很大影响。

由于存在陷阱, 因此在考虑载流子的传输特性时, 除了漂移扩散模型, 还需要考虑陷阱的俘获和发射模型。当电子隧穿过势垒后被陷阱俘获, 还会存在发射现象, 在经历漂移扩散后, 可能还会被俘获, 这就需要进一步考虑陷阱的发射俘获率。电子从陷阱发射的原理是当外加电压时势垒倾斜, 电场越强, 势垒的倾斜程度越大, 电子面对的势垒明显减小, 电子便越容易隧穿。

除了电场的影 响, 发射效率和温度以及陷阱所处的能级也存在关系, 如图 5.35 所示, 随着温度升高, 发射效果变强。同样地, 能级越浅, 发射效果越好。

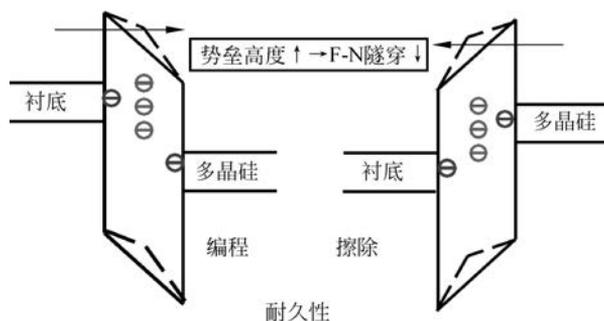


图 5.34 陷阱的产生对编程/擦除操作的影响

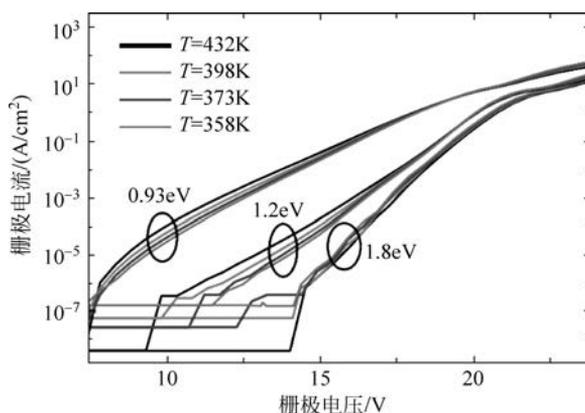


图 5.35 发射效率和温度以及能级的关系

从前文的分析了解到,在电荷俘获存储器中,陷阱是离散化的,如图 5.36(a)所示。为了描述方便,引入陷阱浓度的概念,用连续性陷阱描述问题。一般来说,陷阱的分布大约是服从泊松分布的。值得注意的是,电子可能会经过多个陷阱隧穿,这在应力感应漏电流 (Stress Induced Leakage Current, SILC)效应中会产生尤为明显的尾巴。图 5.36(b)所示,虽然这种情况发生的概率要远低于直接隧穿和单陷阱隧穿,但不能忽略这种现象。

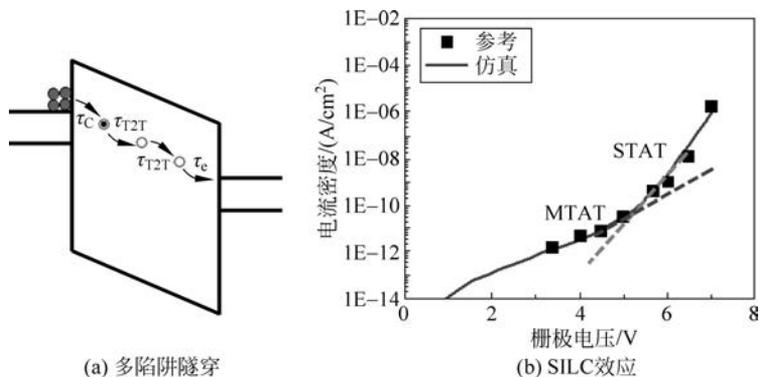


图 5.36 多陷阱隧穿及 SILC 效应中的“尾巴”

前文提到了陷阱的发射概率和温度的关系,接下来关注室温情况。对浮栅单元,在浮栅中存储大量电子,由于电子的存在,能带发生倾斜,如图 5.37 所示,在该情况下,在讨论保持特性时就需要同时考虑陷阱自身的电荷失去和 SILC 效应的影响。此时的保持特性和隧穿

层厚度、存储单元阈值电压分布、反复编程/擦除次数、隧穿层存储层及阻挡层(ONO)的厚度、读取电压等都息息相关。当进行读取操作时,由于外加电压,能带发生弯曲,特别是对擦除态单元进行读取操作时,隧穿层的能带弯曲非常剧烈,这会造成严重的电荷丢失,产生读干扰,在阈值电压分布中产生“尾巴”。

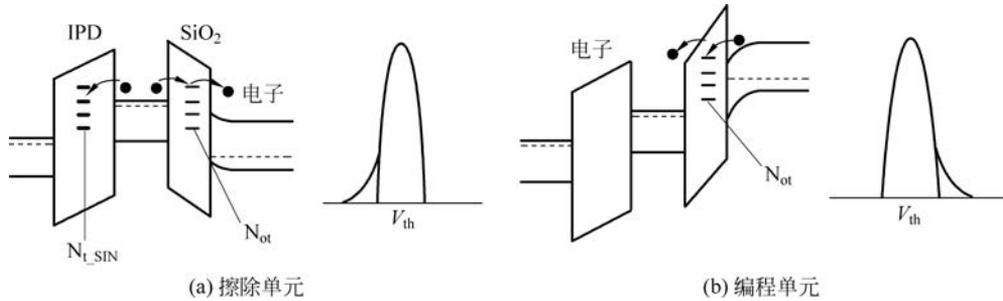


图 5.37 室温下和读操作时的能带特征及阈值电压分布

5.4 3D NAND Flash 器件模拟

5.4.1 3D NAND Flash 器件模拟介绍

随着工艺尺寸的缩小,2D NAND Flash 面临严峻挑战,逐步转向 3D NAND Flash,3D NAND Flash 中,存储单元串的结构如图 5.38 所示。图 5.38(a)为 3D 结构图,图 5.38(b)为剖面图。与 2D NAND Flash 相比,3D NAND Flash 在结构上有着很大的不同,最主要的区别包括以下几点。

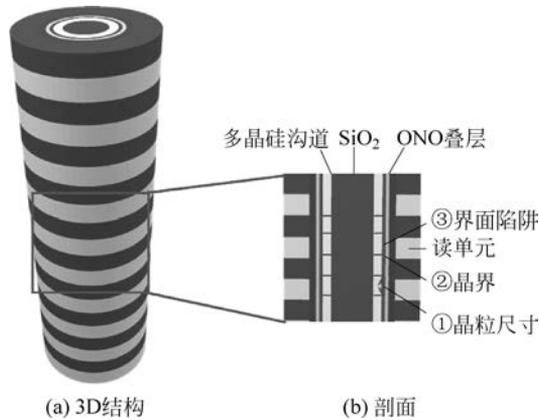


图 5.38 3D NAND Flash 存储器的器件结构

- (1) 3D NAND Flash 的单个存储单元是一个环栅晶体管。
- (2) 3D NAND Flash 的沟道在单元间是连续的,整个沟道是非掺杂的状态。受到工艺限制,单元之间的沟道部分无法进行高浓度掺杂,因此 3D NAND Flash 属于无结器件。
- (3) 3D NAND Flash 的沟道为多晶硅材料。
- (4) 3D NAND Flash 的电荷存储层也是连续的,没有进行分割。

这些结构上的变化为 3D NAND Flash 器件的工作性能和可靠性带来了许多变化。首先,环栅结构改变了原来平面结构晶体管所具有的电场分布,对编程和擦除性能有很大影响。由于单元之间的沟道没有进行高掺杂,因此在沟道导通时,该部分沟道的开启依赖于相邻单元对其产生的边缘电场,这会使沟道的导电能力受到影响,同时单元之间的距离(也称为栅间距)成为影响 3D NAND Flash 性能的重要因素。由于 3D NAND Flash 的沟道材料为多晶硅,这为沟道电势的传导、随机电报噪声、导电性能、阈值电压的温度特性等都带来了一系列的问题。此外,由于 3D NAND Flash 的电荷存储层在单元间是连续的,更使一个单元存储的电荷会向单元两侧运动,甚至进入相邻的单元中,从而影响存储单元的保持特性。

在 3D NAND Flash 存储器的研发过程中,为了更好地研究这些问题,并提高其工作性能,TCAD 仿真技术逐渐成为一种重要的研究手段。对于 3D NAND Flash 来说,由于其复杂的器件结构和阵列结构,以及使用材料的多层次和多样性,2D NAND Flash 中某些常见实验的成本增加,实验的难度和复杂度都有所上升。在这种情况下,TCAD 仿真的重要性更加突出。

例如,为了提高存储密度,需要不断增加存储单元的层数。在这个过程中,会有一些非理想效应出现,比如进行读取操作时,沟道电流会经过所有的存储单元,因此,存储单元的增加会造成沟道电流的减小。利用 TCAD 仿真技术,就可以预测当 3D NAND Flash 层数增加时,沟道电流的变化情况。如图 5.39 所示,随着存储单元层数增加,沟道电流出现大幅下降。类似地,利用 TCAD 仿真技术,就可以在进行下一步的制造与研发之前,对 3D NAND Flash 器件的性能进行预测,从而提高研发效率。

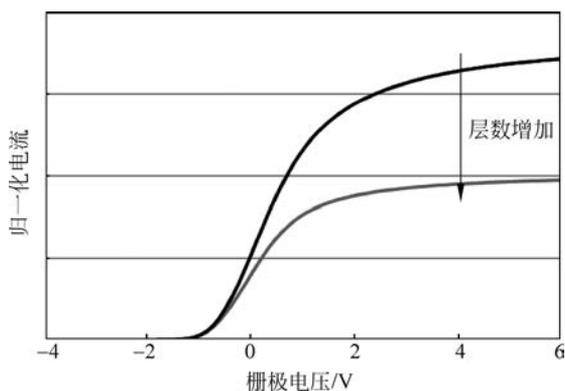


图 5.39 存储单元层数的增加与沟道电流的减小

与 2D NAND Flash 一样,3D NAND Flash 器件的模拟同样是基于物理模型完成的,其在基本物理机制上,与 2D NAND Flash 器件是一致的。有所不同的是,3D NAND Flash 器件的结构更加复杂,仿真中需要考虑更多的因素,3D NAND Flash 中涉及的主要物理模型如图 5.40 所示。在器件的仿真中,除了存储单元本身,更多地需要关注存储单元串的性能以及单元与单元之间的影响。本节将继续补充 3D NAND Flash 器件模拟中重要的物理模型,并介绍 3D NAND Flash 器件模拟技术在实际问题研究中的应用以及常见仿真工具的基本情况。

在 3D NAND Flash 仿真中,影响存储单元性能的陷阱包括以下几种。

(1) 存储层氮化硅中含有大量的陷阱,在多晶硅沟道中也同样含有大量体陷阱。



图 5.40 3D NAND Flash 涉及的主要物理模型

(2) 多晶硅沟道中含有大量晶粒,而晶粒与晶粒间的界面即晶界处同样含有大量的界面陷阱。

(3) 多晶硅沟道与隧穿层界面以及与沟道内侧填充介质层的界面处均含有大量的陷阱。

多晶硅陷阱模型如下。

在 3D NAND Flash 中,沟道材料为多晶硅,相对于单晶硅沟道,多晶硅沟道引入了大量的陷阱,这些陷阱主要存在于晶粒间界面处^[60]。如图 5.41 所示,在 3D NAND Flash 中,多晶硅沟道的陷阱主要分为 3 类:沟道与隧穿层界面处的前界面陷阱(Front Interface Traps, FIT)、晶界陷阱(GB traps)、沟道与沟道孔内填充介质界面处的背界面陷阱(Back Interface Traps,BIT)^[60]。

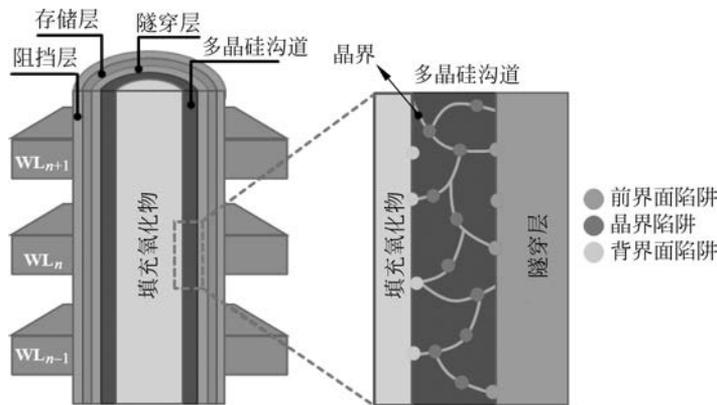


图 5.41 多晶硅沟道陷阱:前界面陷阱、晶界陷阱和背界面陷阱

前界面和背界面陷阱是多晶硅与氧化硅层的界面陷阱,这些界面陷阱主要是 Si 悬挂键。而晶界陷阱主要包含 Si-Si 键扭曲形成的浅能级陷阱和 Si 悬挂键形成的深能级陷阱^[61-62]。浅能级陷阱主要影响沟道开态电流和跨导,而深能级陷阱主要影响亚阈值摆幅和阈值电压^[61-62]。

从 2D NAND Flash 到 3D NAND Flash,沟道材料由单晶硅变为了多晶硅,受晶界陷阱的影响,多晶硅展现出不一样的温度特性,如图 5.42 所示。随着温度的升高,受声子散射影

响,电子迁移率降低,导致单晶硅沟道开态电流降低;相反,由于晶界处陷阱的影响,受主陷阱会形成势垒,阻碍电子在沟道中传输。随着温度升高,晶界陷阱导电能力增强,克服了声子散射造成的迁移率降低问题,进而出现开态电流随温度升高而增加的现象^[63]。

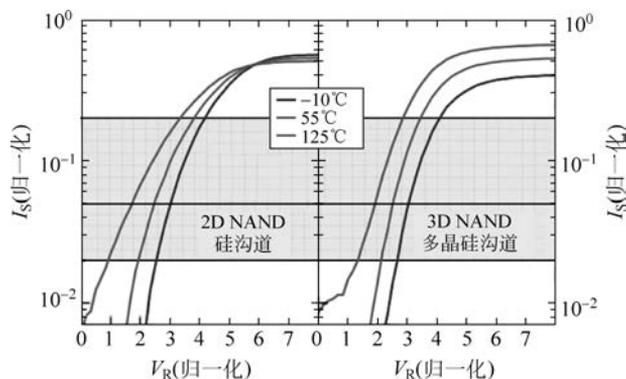


图 5.42 2D NAND Flash 与 3D NAND Flash 不同温度下,开态电流对比^[63]

5.4.2 3D NAND Flash 器件沟道模拟

1. 多晶硅沟道晶粒尺寸对数正态分布对器件性能影响

在 3D NAND Flash 中,多晶硅作为主流沟道材料受到了广泛的关注。要从根本上降低晶界陷阱对器件性能波动性的影响,需要进一步优化多晶硅的结构,增大多晶硅晶粒尺寸,减小晶界数量。为了优化多晶硅沟道,首先利用透射电子显微镜及旋进电子衍射技术统计出多晶硅晶粒尺寸分布图,可以清晰地发现,在多晶硅沟道中,晶粒尺寸满足对数正态分布,如图 5.43 所示,该分布形式的产生原因是多晶硅成核以及晶粒生长。随着晶粒的生长,成核位点耗尽,此时分布如图 5.44(a)所示,成核过程结束,随着进一步晶粒的生长如图 5.44(b)所示,最终形成了对数正态分布的晶粒尺寸分布形式^[64]。

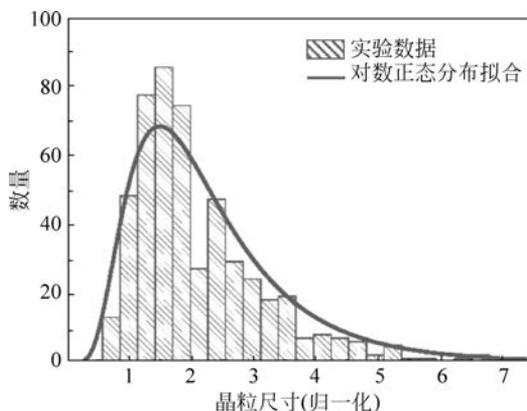


图 5.43 多晶硅晶粒尺寸对数正态分布图

为了研究多晶硅晶粒尺寸对数正态分布对器件性能的影响,利用 TCAD 2D 器件仿真,搭建了对数正态分布仿真平台,根据给定的对数正态分布平均值和标准差在多晶硅沟道中生成对数正态分布的晶粒尺寸列表,并根据该列表将多晶硅沟道进行分割,随后利用

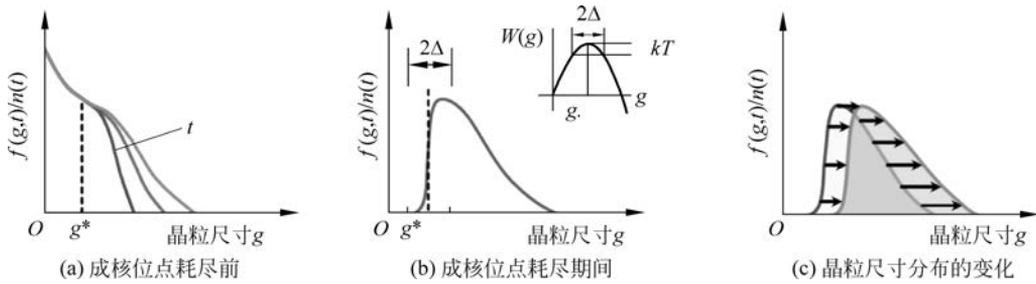


图 5.44 晶粒尺寸满足对数正态分布

S-Device 工具对该结构进行电学特性仿真。图 5.45 为 TCAD 仿真阈值电压统计分布图，对数正态分布晶粒尺寸对器件性能的影响是晶界位置和晶界数量的综合影响。当晶粒尺寸比较小时，即远小于存储单元时，阈值电压会随着晶界数量的增加而增大；相反，随着晶粒尺寸的增大，晶界数目会明显减小，并且阈值电压也会显著减小。但是，随着晶界数量的减小，晶界位置对阈值电压的影响变得更加敏感。

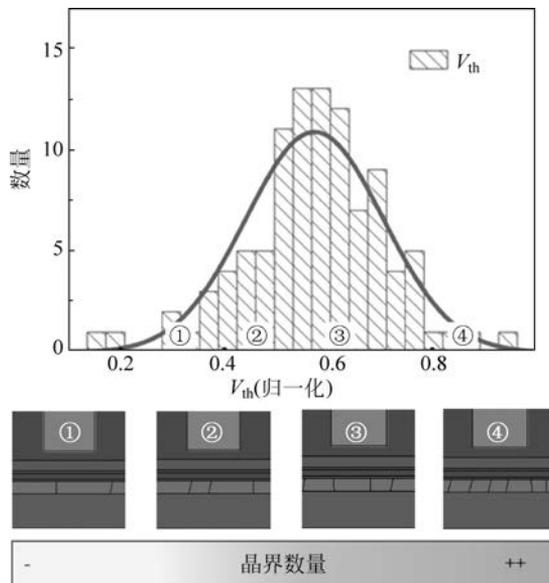


图 5.45 TCAD 仿真阈值电压统计分布图

2. 多晶硅晶界陷阱导致的读取不稳定性问题

与单晶硅沟道相比，多晶硅沟道最根本的特点在于其内部含有大量的晶界陷阱。这些晶界陷阱的产生与修复，以及其对电荷的俘获，会对 3D NAND Flash 器件的性能造成很大的影响。因此，多晶硅沟道会导致读取操作的不稳定性问题。这种读取操作的不稳定性在于，当对单元进行读取时，读电流会在短时间内出现下降。图 5.46 中对比了单晶硅沟道器件(S1、S2)和多晶硅沟道器件(P0、P1)在读取操作开始后 $2\mu\text{s}$ 到 1s 时间内读电流的变化情况。当沟道为多晶硅时，读电流会出现瞬态的变化，在短时间内出现大幅下降，而单晶硅沟道的器件则不存在这种现象。同时，这种读电流的变化与读取时所加的偏压有关。随着读取操作栅极偏压的增加，读电流的快速变化过程逐渐变得不明显直至消失。

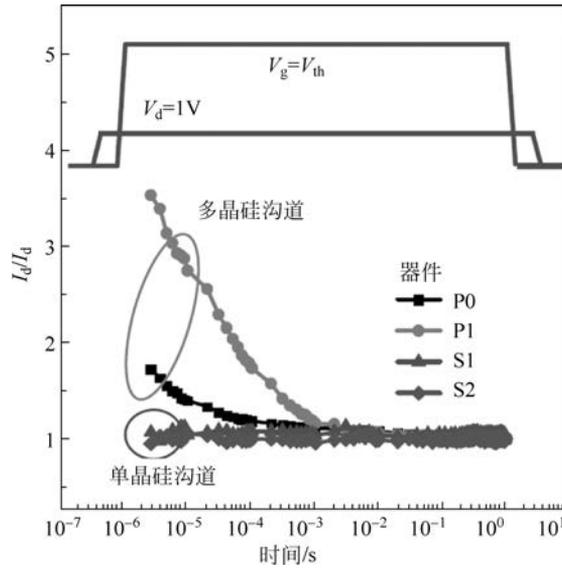


图 5.46 多晶硅沟道和单晶硅沟道读电流的变化

由于晶界陷阱导致的现象都可以用晶界陷阱的占据模型予以解释,如图 5.47 所示,在晶界陷阱占据模型中,晶界陷阱有着三种占据状态。除了晶界陷阱,还存在着隧穿氧化层与多晶硅沟道间的界面陷阱(interface trap)。对栅极施加一个小的读取电压(V_{g1}),使器件处

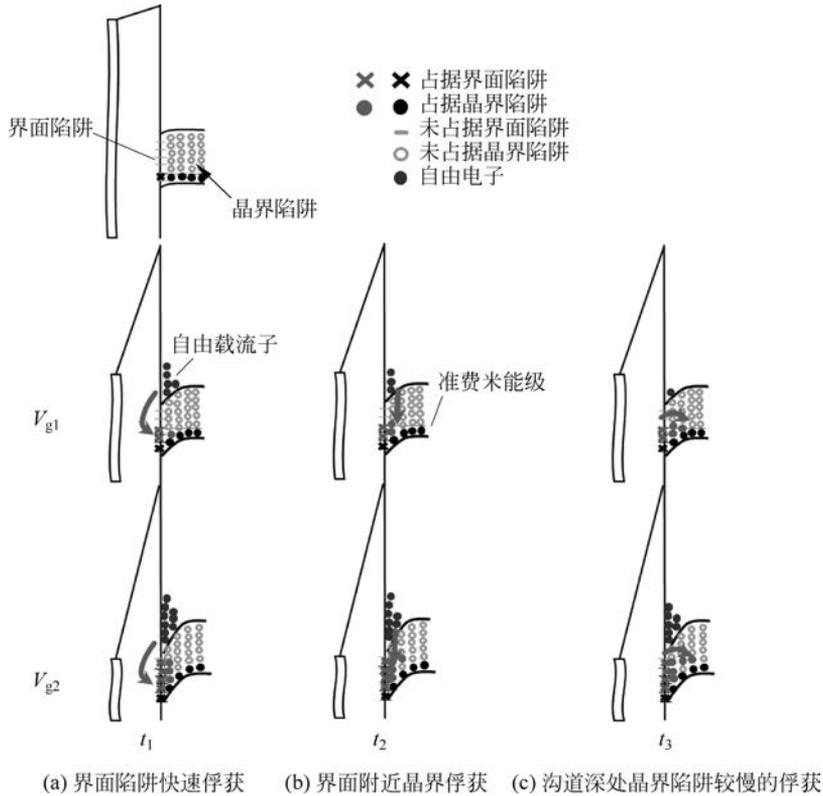


图 5.47 晶界陷阱占据模型

于弱反型时,首先在费米能级以下的界面陷阱会很快地被自由电子占据(t_1)。之后,在很短的时间内(约 $10\mu\text{s}$),在 SRH 过程的作用下,邻近界面陷阱的这一部分晶界陷阱会被电子占据(t_2)。而对于远离界面处,在多晶硅内部深处的晶界陷阱,则需要较长的时间(约 200ms)才会被电子占据(t_3)。由于这些电荷填充过程的存在,将会导致在读取过程中,阈值电压上升,读取电流出现下降,即产生如图 5.46 所示的现象。

类似地,在栅极施加一个较大的读取电压(V_{g2}),使器件处于强反型状态时,也会出现类似的情况。在强反型的状态下,反型层电荷的浓度很高,能带弯曲程度也更大,因此,晶界陷阱会更快地被填充。由于器件处于强反型状态,此时的沟道电流对阈值电压的变化不如在弱反型区那么敏感。当栅极偏压增加时,沟道电流的变化的幅度减小,这便是不同栅压下产生读取操作不稳定的原因。

在 TCAD 仿真中,可以方便地调整晶粒的尺寸以及晶粒的均匀性。如图 5.48 中的 6 种晶粒分布情况,分布(1)~分布(3)的晶粒直径为 30nm ,分布(4)~分布(6)的晶粒直径为 10nm ,均匀性有所差异。研究表明,晶粒尺寸越大,晶粒尺寸分布越均匀,晶界陷阱对 3D NAND Flash 器件性能的影响越不显著。

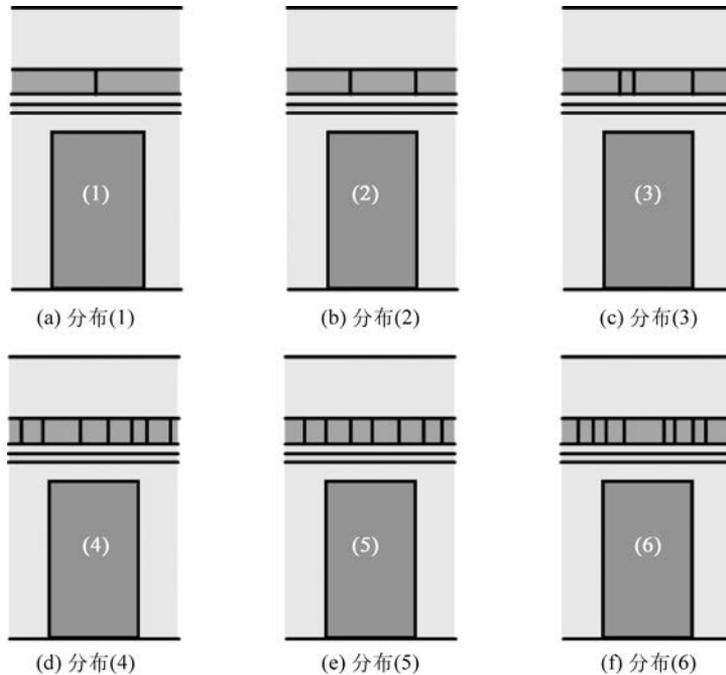


图 5.48 6 种晶粒分布情况

3. 多晶硅沟道对上选择管阈值电压的影响

在 3D NAND Flash 的存储单元串中,除了存储单元,还包括一部分功能单元,如上选择管、冗余管(dummy)和下选择管(Bottom Select Gate,BSG),可以实现对存储单元的操作并改善性能。上选择管控制着编程和读取时存储单元串的选择,其阈值电压的稳定对于器件的正常工作非常重要。

如图 5.49 所示,在对存储单元进行反复编程/擦除操作的过程中,存储单元串的上选择管的阈值电压会显著上升,其阈值电压分布会出现非对称右移。研究表明,这种上选择管阈

值电压的变化仅与对存储单元擦除的工作过程有关,而与对存储单元编程的工作过程无关。此外还发现,在操作结束一段时间之后,上选择管阈值电压又会逐渐恢复,如图 5.50 所示。

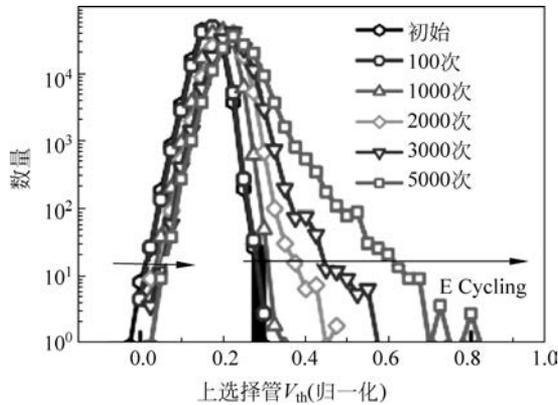


图 5.49 工作过程中上选择管阈值电压分布的变化

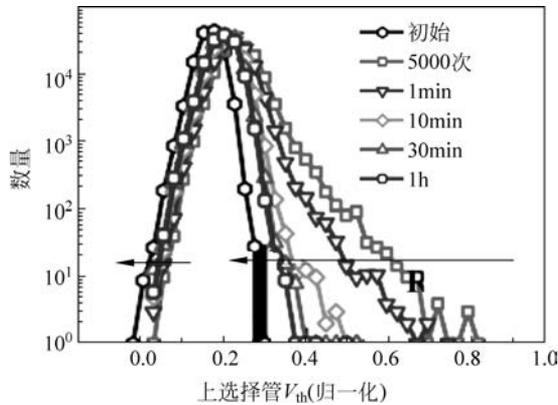


图 5.50 上选择管阈值电压的恢复

在 3D NAND Flash 的擦除过程中,会对存储单元串末端的 P 阱施加一个高电压,使沟道电压上升,同时 P 阱中的空穴注入沟道中提供单元擦除所需的空穴,以完成擦除操作。由于受到多晶硅沟道中晶界陷阱的散射,沟道电势的传导缓慢,在单元栅极间结构电容的作用下,P 阱的电势会率先传导到位线端,这时 P 阱的电势还没有完全传导到沟道上端,因此会造成位线端和顶部单元的沟道间出现巨大的电势差。由于与位线相连的沟道部分进行了 N 型重掺杂,而单元部分的沟道为未掺杂,沟道的顶部处会出现一个近似 PN 结。由于电势差的作用,这个 PN 结处于反偏的状态。

利用 TCAD 仿真技术,可以更好地分析这一过程。图 5.51 给出了随着 P 阱所加电压的增加,沟道电势的分布情况。无论 P 阱所加电压的大小如何,位线端都能耦合出相应的电势,当 P 阱上的电压超过 9V 时,上选择管处的沟道电压会显著上升。在仿真中,可以进一步提取出当 P 阱电压为 9V 时,沟道中的电势分布和 PN 结处带带隧穿效应的强弱。如图 5.52 所示,由于电势差在沟道的顶端产生了一个巨大的电场,且由于 PN 结的存在,还会引起强烈的带带隧穿现象,产生大量的电子-空穴对。其中电子被位线收集,

而空穴则在强电场的作用下,被加速至很高的能量,成为了热空穴(Hot-Hole, HH),足以在上选择管附近打断多晶硅中的硅-氢键($\equiv\text{Si-H}$)产生晶界陷阱^[67],从而引起上选择管阈值电压的上升。而在对器件的操作结束之后,硅-氢键($\equiv\text{Si-H}$)又会发生钝化,从而恢复阈值电压。

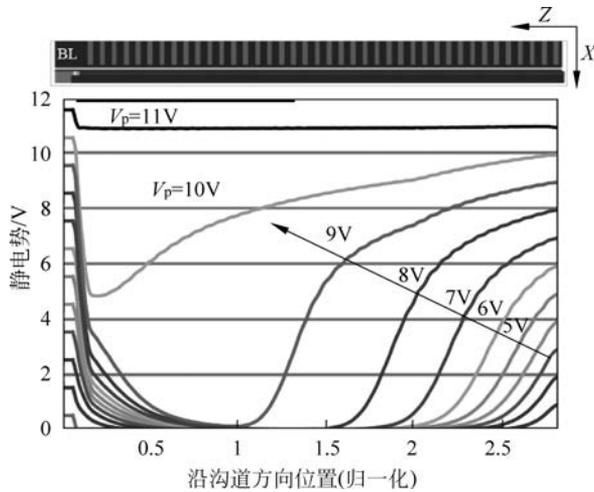


图 5.51 擦除操作中沟道电势传导的仿真

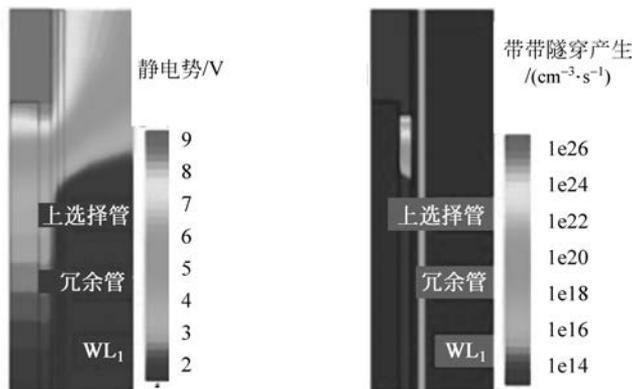


图 5.52 P 阱电压为 9V 时的电势分布和带带隧穿的仿真结果

5.4.3 3D NAND Flash 器件操作及可靠性模拟

当 2D NAND Flash 转变到 3D NAND Flash 后,与需要操作的存储串的相邻存储串数量增多,存储串的环境变得复杂,如图 5.53 所示。根据位线和上选择管的组合方式,可以分为四类,其中一类是编程状态,被编程的单元为图中所框单元,另外的相邻存储单元分为 Case A、Case B 和 Case C 三类。三类的主要区别在于上选择管的关断状态不同,上选择管端的漏电情况会影响编程抑制时沟道电势抬升的效果。如图 5.54 所示,Case C、Case A 和 Case B 的上选择管端漏电流依次降低,沟道电势的抬升效果逐渐增强,编程抑制效果更好。

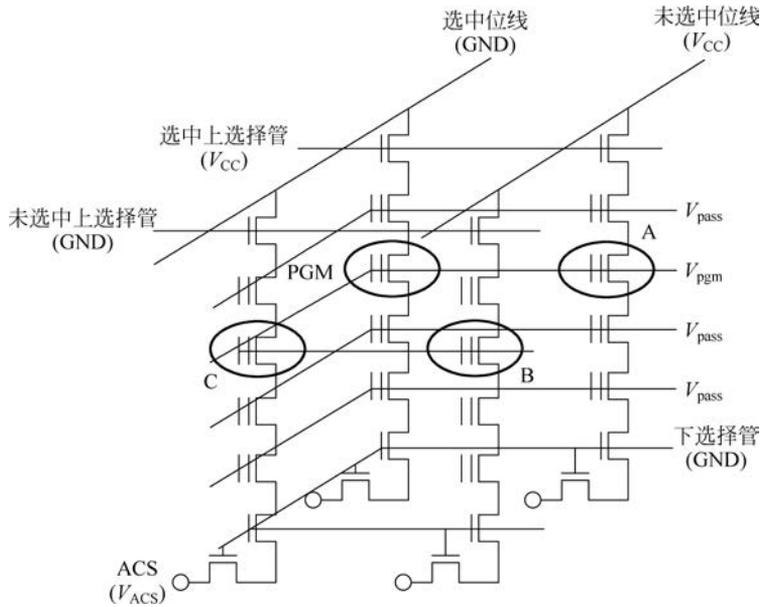


图 5.53 3D NAND Flash 编程状态

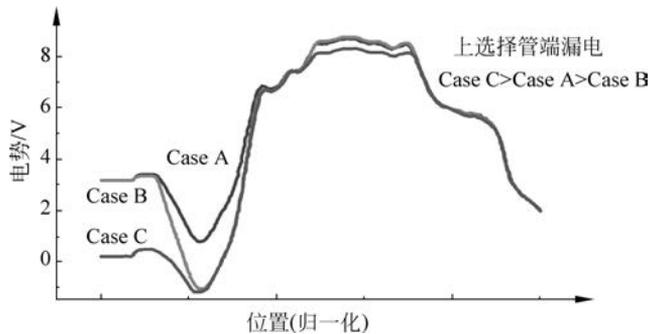


图 5.54 不同编程抑制的沟道电势情况

在 3D NAND Flash 中,存储单元多为 CTF 型,相邻单元间存储层是连续的,因此,电荷不仅失去在垂直方向上通过隧穿层和阻挡层流失,还会出现存储的电荷在存储层中横向扩散,进而影响数据保持特性,如图 5.55 所示^[68-69]。基于 TCAD 仿真工具,并结合 P-F 发射模型和陷阱辅助隧穿等模型,可以将垂直方向电荷流失和横向扩散电荷流失进行独立模拟,更有助于研究其失效机制。研究表明,单元间不同状态组合,对横向扩散具有不同的影响,当存储单元间为图 5.56(a)所示的编程-编程-编程模式(Program-Program-Program Pattern, PPP Pattern)时,电荷失去得最少,电荷分布展宽最少,而图 5.56(b)所示的擦除-编程-擦除模式(Erase-Program-Erase Pattern, EPE Pattern)时,电荷分布展宽严重,电荷失去严重,数据保持特性最差^[68-71]。此外,研究结果表明,存储层中浅陷阱有助于电荷的横向扩散,不利于数据保持特性,因此,在材料方面的优化在于减少存储层中浅陷阱数量,增加深陷阱数量^[56],从而实现数据保持特性的进一步提高。

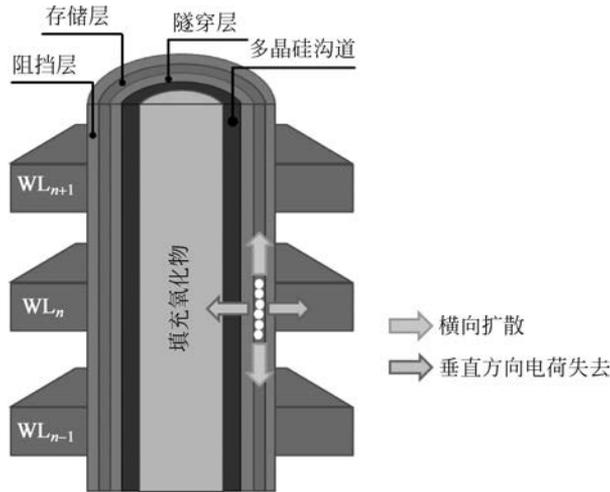


图 5.55 3D NAND Flash 垂直方向和横向电荷失去示意图

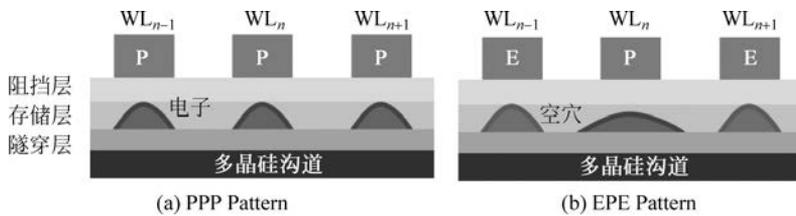


图 5.56 PPP Pattern 和 EPE Pattern 的电子-空穴分布图

本章小结

本章主要介绍了 3D NAND Flash 模型模拟技术,首先介绍了当前主流的仿真工具 Sentaurus TCAD 及其主要仿真模块。随后,详细介绍纳米尺度器件模拟,以及随着 MOS 器件的发展,MOS 器件的尺寸进入纳米尺度后面临着的系列问题。为了研究纳米尺寸器件的各项性能,对半导体仿真模拟技术进行了介绍,并针对关键器件模型进行了介绍。2D NAND Flash 器件的模拟在传统逻辑器件仿真的基础上还增加部分模型,如浮栅存储器中的浮动金属栅、浮动半导体栅、电荷守恒等模型。同时要考虑陷阱模型,陷阱是局域态,可以表现为界面的悬挂键、也可以是体材料中的悬挂键、断键、键的扭曲,界面陷阱的占据状态、陷阱状态、不同类型陷阱的俘获率、陷阱能级等。相比于 2D NAND Flash 器件模拟,3D NAND Flash 器件模拟增加了多晶硅沟道相关模拟、复杂阵列中的操作模拟以及连通存储层带来的横向电荷流失模拟等。

习题

- (1) NAND Flash 器件仿真中,势垒隧穿机制有哪些?
- (2) 在 Sentaurus 中实际建立一个三层的 3D NAND Flash,模拟不同的多晶硅晶粒尺寸对器件特性的影响。

- (3) 相比于单晶硅,多晶硅在电学仿真中,需要考虑哪些物理模型?并解释原因。
- (4) 在进行 3D NAND Flash 模拟中,详细介绍需要考虑的运输模型。
- (5) 3D NAND Flash 模拟中,横向电荷失去产生的原因有哪些?需要重点使用哪些仿真模型?
- (6) 3D NAND Flash 与 2D NAND Flash 在仿真过程中的主要差别有哪些?
- (7) 相比于 MOSFET 仿真,NAND Flash 仿真需要考虑更多的物理模型,介绍相关模型。
- (8) 陷阱仿真需要考虑哪些因素?如何有效地确定仿真中的陷阱设置?
- (9) 半导体和介质材料间的载流子输运有哪些特点?
- (10) 使用 WKB 近似值,计算当电势为①平坦和②下坡时,动能为 1eV 的电子从左侧入射的隧穿概率,如图 5.57 所示。可以使用 Wolfram-alpha。其中,

$$T \approx e^{-2\gamma}, \quad \gamma = \int_0^a \frac{1}{\hbar} \sqrt{2m [V(x) - E]} dx$$

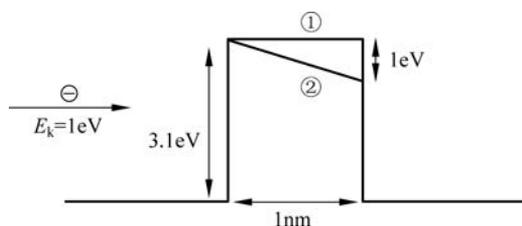


图 5.57 习题(10)图

- (11) 硅衬底上 SiO₂ 厚度为 5nm,顶部采用 Pt 探针,当探针上施加 0~6V 电压时,请计算流过 SiO₂ 的 F-N 隧穿电流。

$$I = A_{\text{eff}} \frac{m_0 q^3}{8\pi m_{\text{ox}} q \hbar \phi_B} E_{\text{ox}}^2 \exp\left(-4 \frac{\sqrt{2m_{\text{ox}}}}{3q\hbar E_{\text{ox}}} q\phi_B^{\frac{3}{2}}\right)$$

其中, I 为 F-N 隧穿电流; $E_{\text{ox}} = V_{\text{ox}}/d_{\text{ox}}$ 为电场; $A_{\text{eff}} = 50\text{nm}^2 = 5 \times 10^{-17} \text{m}^2$ 为探针与样品间的有效面积; $q = -1.602 \times 10^{-19}$ 电荷量; $\hbar = 6.58 \times 10^{-16} \text{eVs} = 1.055 \times 10^{-34} \text{J}$ 为约化普朗克常数; $m_{\text{ox}} = 0.37m_e$ 为氧化硅中电子有效质量; $m_e = 5.10 \times 10^3 \text{eV}/c^2 = 9.1 \times 10^{-31} \text{kg}$ 为电子有效质量; $d_{\text{ox}} = 5\text{nm}$ 为 SiO₂ 厚度; $\phi_B = 3.5\text{eV} = 5.61 \times 10^{-19} \text{J}$ 为 SiO₂ 势垒高度。

参考文献

- [1] Synopsys. Sentaurus™ Workbench User Guide[EB/OL]. <https://www.doc88.com/p-181690531371.html>.
- [2] Synopsys. Sentaurus™ Process User Guide[EB/OL]. <https://www.doc88.com/p-7764823673172.html>.
- [3] Synopsys. Sentaurus™ Structure Editor User Guide[EB/OL]. <https://www.doc88.com/p-6631629724317.html>.
- [4] Synopsys. Sentaurus™ Device User Guide[EB/OL]. https://www.docin.com/p-351831400.html?_t_t_t=0.7626968347126073.
- [5] Stratton R. Diffusion of hot and cold electrons in semiconductor barriers[J]. Physical Review, 1962, 126(6): 2002.
- [6] Blotekjaer K. Transport equations for electrons in two-valley semiconductors[J]. IEEE Transactions on Electron Devices, 1970, 17(1): 38-47.

- [7] Tang T-W. Extension of the Scharfetter—Gummel algorithm to the energy balance equation[J]. *IEEE Transactions on Electron Devices*,1984,31(12): 1912-1914.
- [8] Chen D, Yu Z, Wu K-C-C, et al. Dual energy transport model with coupled lattice and carrier temperatures[M]. *Simulation of Semiconductor Devices and Processes*. Springer, 1993: 157-160.
- [9] Li Y. A Novel Approach to Carrier Temperature Calculation for Semiconductor Device Simulation using Monotone Iterative Method. Part I: Numerical Algorithm[C]//3rd WSEAS Symposium on Mathematical Methods Computing Technology Electrical Engineering,2001.
- [10] Li Y, Sze S, Chao T-S. A practical implementation of parallel dynamic load balancing for adaptive computing in VLSI device simulation[J]. *Engineering with Computers*,2002,18(2): 124-137.
- [11] Gummel H K. A self-consistent iterative scheme for one-dimensional steady state transistor calculations[J]. *IEEE Transactions on Electron Devices*,1964,11(10): 455-465.
- [12] Selberherr S, Fichtner W, Pötzl H. MINIMOS—A program package to facilitate MOS device design and analysis[J]. *Numerical Analysis of Semiconductor Devices and Integrated Circuits*,1979,1: 275-279.
- [13] Cottrell P E, Yu Z. Velocity saturation in the collector of Si/Ge/sub x/Si/sub 1-x//Si HBT's[J]. *IEEE Electron Device Letters*,1990,11(10): 431-433.
- [14] Buturla E, Cottrell P. FIELDAY-Finite Element Device Analysis Program[J]. IBM TR 190356,1975.
- [15] Pinto M R, Rafferty C S, Dutton R W. PISCES II: Poisson and continuity equation solver[EB/OL]. <https://www.docin.com/p-283656253.html>.
- [16] Yu Z, Chen D, So L, et al. PISCES-2ET—Two-Dimensional Device Simulation for Silicon and Heterostructures[J]. *Integrated Circuits Laboratory, Stanford University, Stanford, California*,1994.
- [17] Francis P, Terao A, Flandre D, et al. Modeling of ultrathin double-gate nMOS/SOI transistors[J]. *IEEE Transactions on Electron Devices*,1994,41(5): 715-720.
- [18] Mari A, Eisert J. Positive Wigner functions render classical simulation of quantum computation efficient[J]. *Physical Review Letters*,2012,109(23): 230503.
- [19] Do V-N. Non-equilibrium Green function method: theory and application in simulation of nanometer electronic devices [J]. *Advances in Natural Sciences: Nanoscience and Nanotechnology*, 2014, 5(3): 033001.
- [20] Jacoboni C, Reggiani L. The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials[J]. *Reviews of Modern Physics*,1983,55(3): 645.
- [21] Fischetti M V, Laux S E. Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects[J]. *Physical Review B*,1988,38(14): 9721.
- [22] Venturi F, Smith R K, Sangiorgi E C, et al. A general purpose device simulator coupling Poisson and Monte Carlo transport with applications to deep submicron MOSFETs[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*,1989,8(4): 360-369.
- [23] Aeberhard U. Theory and simulation of quantum photovoltaic devices based on the non-equilibrium Green's function formalism[J]. *Journal of Computational Electronics*,2011,10(4): 394-413.
- [24] Aberle A G, Glunz S, Warta W. Impact of illumination level and oxide parameters on Shockley-Read-Hall recombination at the Si-SiO₂ interface[J]. *Journal of Applied Physics*,1992,71(9): 4422-4431.
- [25] Gundapaneni S, Kottantharayil P, Ganguly P J. Investigation of junction-less transistor (JLT) for CMOS scaling[D]. 2012.
- [26] Caillat C, Beaman K, Bicksler A, et al. 3D NAND GIDL-assisted body biasing for erase enabling CMOS under array (CUA) architecture[C]//IEEE International Memory Workshop,2017.
- [27] Hermle M, Letay G, Philipps S, et al. Numerical simulation of tunnel diodes for multi-junction solar cells[J]. *Progress in Photovoltaics: Research and Applications*,2008,16(5): 409-418.

- [28] 刘明军. 隧穿场效应晶体管的新型器件结构及优化设计研究[D]. 成都:电子科技大学,2019.
- [29] Synopsys. 3D NAND Flash cell device simulation[R]. Synopsys Inc,2016.
- [30] Kane E O. Theory of tunneling[J]. Journal of applied Physics,1961,32(1): 83-91.
- [31] Keldysh L. Behavior of non-metallic crystals in strong electric fields [J]. Soviet Journal of Experimental and Theoretical Physics,1958,6: 763.
- [32] Hänisch W, Vogelsang T, Kircher R, et al. Carrier transport near the Si/SiO₂ interface of a MOSFET [J]. Solid-State Electronics,1989,32(10): 839-849.
- [33] van Dort M J, Woerlee P H, Walker A J, et al. Influence of high substrate doping levels on the threshold voltage and the mobility of deep-submicrometer MOSFETs[J]. IEEE Transactions on Electron Devices,1992,39(4): 932-938.
- [34] Van Dort M, Woerlee P, Walker A. A simple model for quantization effects in heavily-doped silicon MOSFETs at inversion conditions[J]. Solid-State Electronics,1994,37(3): 411-414.
- [35] Paasch G, Übensee H. A modified local density approximation. Electron density in inversion layers [J]. Physica Status Solidi (b),1982,113(1): 165-178.
- [36] Entner R. Modeling and simulation of negative bias temperature instability[D]. Austria: Vienna University of Technology,2007.
- [37] Selberherr S. MOS device modeling at 77K[J]. IEEE Transactions on Electron Devices, 1989, 36(8): 1464-1474.
- [38] Cappelletti P, Golla C, Olivo P, et al. Flash memories[M]. Berlin: Springer Science & Business Media,2013.
- [39] Yeagan J, Taylor H. The Poole-Frenkel effect with compensation present[J]. Journal of Applied Physics,1968,39(12): 5600-5604.
- [40] Biswas A, Dan S S, Le Royer C, et al. TCAD simulation of SOI TFETs and calibration of non-local band-to-band tunneling model[J]. Microelectronic Engineering,2012,98: 334-337.
- [41] Jeong M, Solomon P M, Laux S, et al. Comparison of raised and Schottky source/drain MOSFETs using a novel tunneling contact model[C]//IEEE International Electron Devices Meeting,1998.
- [42] Kameda E, Matsuda T, Emura Y, et al. Fowler-Nordheim tunneling in MOS capacitors with Si-implanted SiO₂[J]. Solid-State Electronics,1998,42(11): 2105-2111.
- [43] Yeo Y-C, King T-J, Hu C. Direct tunneling leakage current and scalability of alternative gate dielectrics[J]. Applied Physics Letters,2002,81(11): 2091-2093.
- [44] Schenk A. Rigorous theory and simplified model of the band-to-band tunneling in silicon[J]. Solid-State Electronics,1993,36(1): 19-34.
- [45] Gundapaneni S, Bajaj M, Pandey R K, et al. Effect of band-to-band tunneling on junctionless transistors[J]. IEEE Transactions on Electron Devices,2012,59(4): 1023-1029.
- [46] Houssa M, Tuominen M, Naili M, et al. Trap-assisted tunneling in high permittivity gate dielectric stacks[J]. Journal of Applied Physics,2000,87(12): 8615-8620.
- [47] Jiménez-Molinos F, Palma A, Gamiz F, et al. Physical model for trap-assisted inelastic tunneling in metal-oxide-semiconductor structures[J]. Journal of Applied Physics,2001,90(7): 3396-3404.
- [48] Jiménez-Molinos F, Gámiz F, Palma A, et al. Direct and trap-assisted elastic tunneling through ultrathin gate oxides[J]. Journal of Applied Physics,2002,91(8): 5116-5124.
- [49] Schroeder H. Poole-Frenkel-effect as dominating current mechanism in thin oxide films—An illusion?! [J]. Journal of applied physics,2015,117(21): 215103.
- [50] Sze S M, Li Y, Ng K K. Physics of semiconductor devices[M]. New York: John Wiley & Sons,2021.
- [51] Kapoor V, Turi R. Charge storage and distribution in the nitride layer of the metal-nitride-oxide semiconductor structures[J]. Journal of Applied Physics,1981,52(1): 311-319.

- [52] Mine T, Fujisaki K, Ishida T, et al. Electron trap characteristics of silicon rich silicon nitride thin films[J]. Japanese Journal of Applied Physics, 2007, 46(5S): 3206.
- [53] Seo Y J, Kim K C, Kim H D, et al. Study of hole traps in the oxide-nitride-oxide structure of the SONOS flash memory[J]. Journal of the Korean Physical Society, 2008, 53(6): 3302-3306.
- [54] Cho H-Y, Kim W-S, Oh J, et al. The Origin of Trap and Effect of Nitrogen Plasma in the Oxide-Nitride-Oxide Structures for Non-Volatile Memory[J]. Journal of the Korean Physical Society, 2010, 57(2): 255-259.
- [55] Gritsenko V, Nekrashevich S, Vasilev V, et al. Electronic structure of memory traps in silicon nitride [J]. Microelectronic Engineering, 2009, 86(7-9): 1866-1869.
- [56] Wu J, Han D, Yang W, et al. Comprehensive investigations on charge diffusion physics in SiN-based 3D NAND flash memory through systematical Ab initio calculations [C]//IEEE International Electron Devices Meeting, 2017.
- [57] Chen C-P, Lue H-T, Hsieh C-C, et al. Study of fast initial charge loss and its impact on the programmed states V_t distribution of charge-trapping NAND Flash [C]//IEEE International Electron Devices Meeting, 2010.
- [58] Lee J-D, Choi J-H, Park D, et al. Degradation of tunnel oxide by FN current stress and its effects on data retention characteristics of 90nm NAND flash memory cells[C]//IEEE International Reliability Physics Symposium Proceedings, 2003.
- [59] El Hdiy A, Salace G, Petit C, et al. Relaxation of interface states and positive charge in thin gate oxide after Fowler-Nordheim stress[J]. Journal of Applied Physics, 1993, 73(7): 3569-3570.
- [60] Arreghini A, Banerjee K, Verreck D, et al. Improvement of conduction in 3-D NAND memory devices by channel and junction optimization[C]//IEEE International Memory Workshop, 2019.
- [61] Ikeda H. Evaluation of grain boundary trap states in polycrystalline-silicon thin-film transistors by mobility and capacitance measurements[J]. Journal of applied physics, 2002, 91(7): 4637-4645.
- [62] 邹兴奇. 3D NAND 存储器串列功能单元的可靠性研究[D]. 北京:中国科学院微电子研究所, 2019.
- [63] Resnati D, Goda A, Nicosia G, et al. Temperature effects in NAND Flash memories: A comparison between 2-D and 3-D arrays[J]. IEEE Electron Device Letters, 2017, 38(4): 461-464.
- [64] Bergmann R B, Shi F G, Queisser H J, et al. Formation of polycrystalline silicon with log-normal grain size distribution[J]. Applied Surface Science, 1998, 123: 376-380.
- [65] Lin W-L, Tsai W-J, Cheng C, et al. Grain boundary trap-induced current transient in a 3-D NAND flash cell string[J]. IEEE Transactions on Electron Devices, 2019, 66(4): 1734-1740.
- [66] Yan L, Jin L, Zou X, et al. Investigation of Erase Cycling Induced TSG V_t Shift in 3D NAND Flash Memory[J]. IEEE Electron Device Letters, 2019, 40(1): 21-23.
- [67] Nakagawa H, Yano H, Hatayama T, et al. Hot Carrier Effect in Low-Temperature Poly-Silicon p-Channel Thin-Film Transistors[C]//Solid State Phenomena, 2003.
- [68] Kang H-J, Choi N, Joe S-M, et al. Comprehensive analysis of retention characteristics in 3-D NAND flash memory cells with tube-type poly-Si channel structure [C]//IEEE Symposium on VLSI Technology, 2015.
- [69] Padovani A, Pesic M, Kumar M A, et al. Understanding and variability of lateral charge migration in 3D CT-NAND flash with and without band-gap engineered barriers [C]//IEEE International Reliability Physics Symposium, 2019.
- [70] Oh D, Lee B, Kwon E, et al. TCAD simulation of data retention characteristics of charge trap device for 3-D NAND flash memory[C]//IEEE International Memory Workshop, 2015.
- [71] Park J, Shin H. Modeling of lateral migration mechanism of holes in 3D NAND flash memory charge trap layer during retention operation[C]//IEEE Silicon Nanoelectronics Workshop, 2019.