

# MOS 器件三维数值模拟

陈大同 吴启明 王泽毅 李志坚  
(清华大学微电子所、计算机系)

1986年3月3日收到

## 一、引言

随着半导体器件尺寸的不断缩小和结构的日趋复杂，常用的解析分析模型往往不能满意地描述器件特性。因此，七十年代以来，二维及三维数值模拟逐渐成为半导体器件研究中的一个重要工具。

本文介绍一个用于 MOS 器件二维及三维稳态分析的有限差分数值模拟程序 TD MOS-I。为了减少计算时间，在理论分析和试算的基础上，我们对算法进行了一些改进。其中主要有：对 Newton-Raphson 迭代次数加以合理控制；在 ICCG 法中选用适当的位置集合；以二维分析结果为三维分析提供初值等等。实算表明，这些改进不仅大大节省了计算时间，而且在一定程度上解决了 Gummel 迭代法普遍存在的开启条件下迭代次数急剧增加的问题。在 VAX<sub>11</sub>-750 小型计算机上，TD MOS-I 进行二维分析，平均每个电压点约需 20 秒左右，进行三维分析，平均每个电压点约需 5 分钟。

此外，为了克服差分法不易处理复杂边界形状的缺点，TD MOS-I 中采用了一种新颖的区域延拓方法。这种方法在保证差分方程组具有良好性质的前提下，可以简便地处理各种复杂的界面及边界形状，从而给用户和程序编制者带来极大的方便。本文不包括这部分内容。

## 二、器件模型及基本方程

TD MOS-I 可用于 MOS 电容，MOS 存贮器和 MOSFET 等各种 MOS 器件的模拟(见图 1)。在半导体内部，电学特性可用五个基本方程加以描述

$$\text{泊松方程: } \nabla^2\phi = -\frac{q}{\epsilon_{Si}}(p - n + N_D^+ - N_A^-) \quad (1a)$$

$$\text{电子连续性方程: } \nabla \cdot J_n - q \frac{\partial n}{\partial t} = q \cdot R \quad (1b)$$

$$\text{空穴连续性方程: } \nabla \cdot J_p + q \frac{\partial p}{\partial t} = -q \cdot R \quad (1c)$$

$$\text{电子电流方程: } J_n = -q \cdot \mu_n \cdot n \cdot \nabla \phi + q \cdot D_n \cdot \nabla n \quad (1d)$$

$$\text{空穴电流方程: } J_p = -q\mu_p p \nabla \psi - qD_p \nabla p \quad (1c)$$

其中  $\psi$  为电位,  $R$  为载流子的净复合率,  $\epsilon_{Si}$  为 Si 的介电常数,  $N_D^+$  和  $N_A^-$  分别为电离的施主和受主杂质浓度,  $J_n$  和  $J_p$  分别为电子和空穴电流密度,  $\mu_n$  和  $\mu_p$  分别为电子和空穴的迁移率,  $D_n$  和  $D_p$  分别为电子和空穴扩散系数,  $q$  为电子电量,  $n$  为电子浓度,  $p$  为空穴浓度:

$$n = n_i \exp(q(\psi - \phi_n)/kT), \quad (2)$$

$$p = n_i \exp(q(\phi_p - \psi)/kT). \quad (3)$$

$n_i$  为本征载流子浓度,  $\phi_n$  为电子准费米势,  $\phi_p$  为空穴准费米势。

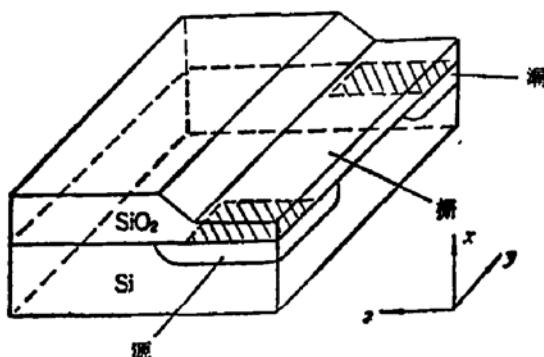


图 1 MOSFET 结构示意图

对于 MOS 器件可做如下假设:

- 1) 忽略产生、复合项, 即:  $R=0$ .
- 2) 仅沟道载流子对电流有贡献.
- 3) Einstein 关系成立, 即:  $D = \frac{kT}{q} \cdot \mu$ .

这样, 对 n 沟 MOS 器件, 半导体内需求解的归一化稳态方程为:

$$\nabla^2 \psi = \Phi_n \cdot \exp(\psi) - \Phi_p [\exp(-\psi)] - N_d. \quad (4)$$

$$\nabla [\mu_n \exp(\psi) \cdot \nabla \Phi_n] = 0. \quad (5)$$

其中:  $N_d = N_D^+ - N_A^-$ ,  $\Phi_n = \exp(-\phi_n)$ ,  $\Phi_p = \exp(\phi_p) = \text{常数}$ .

对于氧化层作如下处理: 若 Si-SiO<sub>2</sub> 界面是单一平面的简单结构 (如图 1 所示的 MOSFET), 则采取简化方法<sup>[2]</sup>, 设氧化层中的电位分布在垂直表面方向近似为线性。这样, 在 Si-SiO<sub>2</sub> 界面 S 上, 电位移连续条件就可近似为:

$$\epsilon_{ox}(V_G - V_{Si-SiO_2})/t_{ox}(y, z) = \epsilon_{Si} \left. \frac{\partial \psi}{\partial x} \right|_S$$

其中  $V_G$  为栅电极电位,  $V_{Si-SiO_2}$  为界面电位,  $t_{ox}$  为氧化层厚度。此外, 各电极接触面为强加边界条件, 而其它边界均为第二类齐次边界条件。

若 Si-SiO<sub>2</sub> 界面的形状较为复杂, 则需要计算氧化层区结点。这时, 我们用区域延拓法处理。

器件的衬底为均匀掺杂, 源漏区杂质浓度, 根据不同的工艺条件, 可选用高斯分布或余误差分布, 或者直接输入工艺模拟结果, 并可计算沟道离子注入。

TDMOS-I 中有两个供用户选择的迁移率模型。模型一仅考虑速度饱和效应的影响<sup>[3]</sup>, 模型二还考虑了杂质散射和载流子散射等效应的影响<sup>[2]</sup>, 用户也可以使用自己的迁移率模型。

为了扩大适用范围, TDMOS-I 允许用户根据需要在三个方程 (1a), (1b) 和 (1c) 中选用任意几个方程 [但 (1a) 必须选用]。例如, 若不考虑电流时可仅用 (1a); 若需要同时考虑多子及少子电流时, 则需同时解三个方程。

### 三、计算方法及改进

用数值方法解非线性微分方程组(4)和(5)的计算量是非常可观的。往往由于计算时间过长,而使程序不能实用。根据以往的报道,在中、小型计算机上进行二维分析,每个电压点的计算时间约5—10分钟<sup>[1]</sup>。在三维分析中,离散化结点多达数千,如果直接沿用二维分析的计算方法,计算量将以 $N_z^3$ 的比例增长( $N_z$ 是第三维的网格层数),这是根本不能实用的。因此,找到合适的计算方法并加以改进是计算方案成败的关键。

#### 1. 基本计算方法

我们选用不等间距的七点差分格式进行微分方程组(4)和(5)的离散化。同时,边界条件的导数项均以一阶均差近似。它的优点是离散化方程组的 Jacobi 矩阵的稀疏性远好于有限元素法。特别是在三维分析中,前者每行有7个非零元素,而后者则多达21个(若采用三棱柱单元)。这就大大节约了解线性方程组的计算量。

离散化方程组为:

$$[A] \text{ 方程组: } A\phi - [\Delta V] \cdot ([\epsilon^*] \cdot \Phi_n - \Phi_p \exp(-\phi) - N_d) = 0. \quad (6)$$

$$[B] \text{ 方程组: } B(\phi) \cdot \Phi_n + D(\phi) = 0. \quad (7)$$

其中  $A$  和  $B$  是  $N$  阶对称、对角方阵 ( $N$  是未知结点数),  $\phi$ 、 $D$ 、 $\Phi_n$ 、 $\exp(-\phi)$  和  $N_d$  是  $N$  维列向量,  $[\epsilon^*]$  和  $[\Delta V]$  是  $N$  阶对角方阵 ( $\Delta V$  是离散化单元的体积元素)。在电流连续方程的离散化中,采用了 Gummel-Scharfetter 方法以保证其稳定性<sup>[3]</sup>。

离散化方程组  $[A]$  和  $[B]$  相互耦合,形成复杂的非线性方程组。我们采取 Gummel 的方法<sup>[4]</sup>交替地迭代求解这组方程,即:在解  $[A]$  方程组时把  $\Phi_n$  看作已知向量,  $\phi$  看作未知向量;而解  $[B]$  方程组时,把  $\phi$  看作已知向量,  $\Phi_n$  看作未知向量,如此反复迭代,直至收敛。

在 Gummel 迭代中,  $[A]$  方程组关于  $\phi$  是非线性的。我们采用 Newton-Raphson 法迭代求解。首先,定义剩余向量:

$$F(\phi) = A\phi - [\Delta V] \cdot ([\epsilon^*] \cdot \Phi_n - \Phi_p \cdot \exp(-\phi) - N_d). \quad (8)$$

其 Jacobi 矩阵:  $H(\phi) = \partial F / \partial \phi$ 。

则迭代公式为:

$$H(\phi_k) \cdot \Delta\phi_k = -F(\phi_k). \quad (9)$$

其中  $\Delta\phi_k = \phi_{k+1} - \phi_k$ ,  $\phi_k$  是第  $k$  次迭代中的  $\phi$  近似值。反复解方程(9),直至收敛,就可得到  $[A]$  方程组的解。

$N-R$  法的优点是收敛快,它具有二阶局部收敛速度。但若初值不好,它又很容易发散。为此,我们一方面尽量提供较好的初值,另一方面也采取了搜索步长的控制措施来保证剩余平方和单调下降,收到了很好的效果。

求解线性方程组占总计算量的相当大部分,其算法优劣对程序质量影响极大。近十几年来,由于稀疏矩阵预排序等技术的进展,在二维分析中,直接解法已可以同迭代算法相竞争。但在三维分析中,迭代算法由于其内存省,计算量小的优点,仍具有很大的优越性。其缺点是方程组系数矩阵性质对收敛速度影响很大。另外,在许多新发展的迭代算

法中,往往需要用迭代加速因子,而迭代因子的合适选择又相当困难。

在我们的问题中,待解线性方程组(7)和(9)的系数矩阵  $B$  和  $H$  都是对称- $M$  矩阵<sup>[5]</sup>。因此,我们选用了 ICCG (不完全 Cholesky 分解十共轭梯度法)<sup>[5,6]</sup>解线性方程组。它不用任何迭代因子,同时具有很高的收敛速度。其基本步骤是,先在某个选定的元素位置集合  $\Omega$  内对系数矩阵进行不完全 Cholesky 分解 (IC), 然后用所得的矩阵改善系数矩阵的条件,再用共轭梯度法 (CG) 迭代求解。对于稀疏线性方程组,由于可以选择位置集合  $\Omega$ ,使其很好地反映系数矩阵的稀疏性质,就可以花费很少的 IC 分解计算量而大大改善系数矩阵的性质,从而获得很高的收敛速度。

## 2. N-R 迭代次数的合理控制

如上所述,在 Gummel 迭代中,要用 N-R 法线性化  $[A]$  方程组并迭代求解。对于 N-R 迭代次数的控制有两种常用的方法。方法一是一直迭代到足够精确地解出  $[A]$  方程组,这使每次解  $[A]$  方程组较费时间。方法二是对  $[A]$  方程组只进行很少几次(例如一次) N-R 迭代,就去解  $[B]$  方程组。这样虽然减少了解  $[A]$  方程组的计算量,但由于  $[A]$  方程组中的解误差很大,容易影响 Gummel 外迭代的收敛速度。因此,这两种方法都不能令人满意。通过理论分析和实算,我们发现,在 Gummel 迭代初始时,使用方法二会严重影响 Gummel 外迭代的收敛速度。但在接近收敛时,方法二与方法一对 Gummel 迭代来说几乎没有什么差别。由此出发,我们采用了分段控制的方法:在开始时,让 N-R 迭代次数多一些,使  $[A]$  方程组的解满足一定的精度要求,而当剩余平方和  $R_s = (F, F)$  小于某一标准时(实算中取 10),让 N-R 只迭代一次。

实算证明,这种分段控制是比较合理的。它比前两种方法平均减少计算量达 40—60%。表 1 是三种方法对一个典型算例的比较。

表 1 三种 N-R 次数控制方法的比较

	方法一	方法二	分段控制
解方程组(9)次数	22	11	11
解 $[B]$ 方程组次数	4	11	4
CPU 时间	5M11S	5M31S	3M06S

## 3. ICCG 法中选用合适的集合

我们试验了元素集合  $\Omega$  中分别含有 7, 13 和 23 对角行的三种 ICCG 法,并分别记为 ICCG(7), ICCG(13) 和 ICCG(23)。试算结果见表 2。

从表 2 可见,ICCG 法收敛速度很快,一般在十几次之内迭代即可收敛。而且其收敛性质对方程阶数不敏感( $N$  从 266 增加到 3458 时,ICCG(7)的迭代次数仅从 7—8 增长到 8—10 次),其计算量约正比于  $N^{1.2}$ (作为对比,直接解法计算量正比于  $N^2 \sim N^3$ ),因而特别适合于三维分析。另外,从表 2 中还可看到, $\Omega$  中元素越多,CG 收敛就越快,但每次迭代的计算量也随之增加。综合起来考虑,我们采用了计算量最省的 ICCG(7)。

[注]: 如果矩阵  $A$  满足 i) 非奇异, ii)  $A^{-1} \geq 0$ , iii) 当  $i \neq j$  时,  $a_{ij} \leq 0$  则称为  $M$  矩阵。

表 2 三种 ICCG 法收敛性质比较

	迭代次数		每次 IC 分解 乘除法次数	每次 CG 迭代 乘除法次数	每次解方程平均总乘除法次数	
	二维	三维			二维	三维
ICCG(7)	7—8	8—10	6N	19N	~150N	~170N
ICCG(13)	5—7	7—9	21N	25N	~170N	~220N
ICCG(23)	5—6	6—8	65N	35N	~260N	~310N

注: N 为未知结点总数, 二维  $N = 266$ , 三维  $N = 3458$ .

#### 4. 电位 $\phi$ 的初值选取

当 MOS 器件在一定的栅压下开启之后, 沟道电流急剧增加, [A], [B] 方程组之间耦合加强, 这使得 Gummel 迭代的收敛速度大大下降, 计算时间急剧增加<sup>[1]</sup>. 为了解决这个问题, 我们采用了以二维分析结果为三维分析提供初值的方法。首先, 在沟道中心对称面  $z = 0$  上进行二维分析, 然后令栅区下方各结点, 电位初值等于中心对称面上相应结点的二维分析结果, 对其它结点的电位初值也做相应调整。这样, 由于对称面上没有电流穿过, 二维分析同三维分析结果相同, 此面上的二维分析结果就给所有沟道区结点提供了相当好的初值, 一般可使初始的剩余平方和  $R_s$  下降 4~5 个数量级。尤其是在开启条件下, 由于电流主要集中在沟道区, 而沟道区具有较准确的电位初值, 就可以大大减少 Gummel 迭代的次数, 从而在一定程度上解决了大电流条件下计算量急剧增加的问题。表 3 是电流开启条件下的一个典型算例: 一般说来, 二维提供初值方法比通常的分区给初值方法节省计算量 40—70%. 在各种电压条件下, 三维分析基本上可在 6—7 次 Gummel 迭代以内, 沟道电流收敛。

表 3 两种初值方法的比较

	解方程组 (9) 次数		解 [B] 方程组次数		CPU 时间
	二维	三维	二维	三维	
分区提供初值	0	73	0	19	52M31S
二维提供初值	53	19	15	7	18M29S

采用以上改进措施后, 在二维分析中, 每个电压需计算约 15—20 秒, 在三维分析中, 每个电压点需计算 4—7 分钟。

### 四、计算实例

例 1: 图 2 是 TDMOS-I 对一个 MOSFET 的二维分析结果与 [8] 中实验曲线的比较, 两者基本相符。这验证了本程序的可靠性。器件的主要参数为: 有效沟道长度  $L = 1.2 \mu\text{m}$ , 横向与纵向结深  $x_L = x_J = 1.85 \mu\text{m}$ , 衬底受主浓度  $N_A = 6.5 \times 10^{15}/\text{cm}^3$ ,  $\text{N}^+$  扩散区表面浓度  $C_0 = 1.0 \times 10^{19}/\text{cm}^3$ , 栅氧厚度  $t_{ox} = 700 \text{\AA}$ .

例 2: 图 3 是 MOSFET 阈电压  $V_T$  的短沟与窄沟效应的模拟。

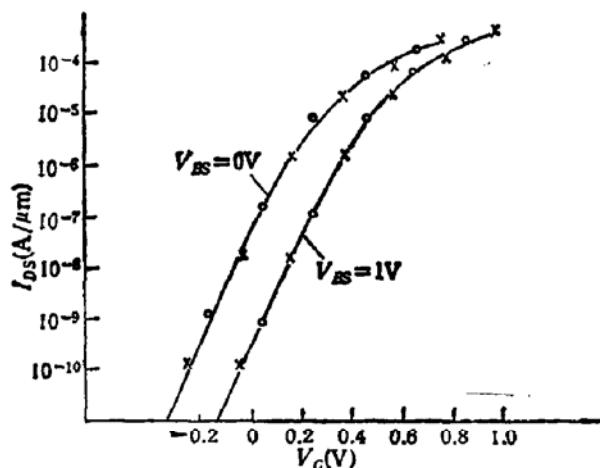


图 2 MOSFET 二维分析结果同实验曲线[8]比较  
 ——实验曲线 ○[6]模拟结果  
 ×本程序模拟结果

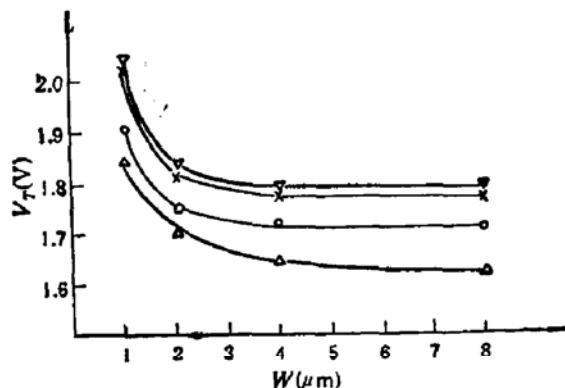


图 3 MOSFET 阈电压  $V_T$  的短沟与窄沟效应的模拟  
 ▽—— $L = 8\mu\text{m}$  ×—— $L = 4\mu\text{m}$   
 ○—— $L = 2\mu\text{m}$  △—— $L = 1\mu\text{m}$

器件结构见图 1。其主要参数为:  $N_A = 10^{16}/\text{cm}^3$ ,  $C_0 = 10^{19}/\text{cm}^3$ ,  $x_j = 0.5\mu\text{m}$ ,  $x_t = 0.4\mu\text{m}$ ,  $t_{ox} = 500\text{ \AA}$ , 场氧厚度  $t_F = 1.0\mu\text{m}$ , 过渡氧化层斜率为 1:1。

由图 3 中可见: 当沟道长度  $L$  或沟道宽度  $W$  小于 4 微米时, 开启电压  $V_T$  就会受到短沟或窄沟效应的影响。随着  $L$  的缩短,  $V_T$  将下降, 而随着  $W$  的缩短,  $V_T$  将会上升。当  $L$ 、 $W$  按同样比例压缩时, 窄沟效应会超过短沟效应, 而使  $V_T$  上升。这些初步结果显示了三维分析是分析窄沟及短沟效应的有力工具。

## 五、结 论

本文叙述了 MOS 器件三维数值模拟程序 TD MOS-I 的基本算法和主要算法改进。初步试算表明: 本程序的计算方案是可行的, 它具有计算量省通用性强, 使用方便的特点。它可以成为研究各种 MOS 器件微观机理和宏观特性的有力工具。

## 参 考 文 献

- [1] 秦树, 硕士学位论文, 清华大学微电子所, 1982.
- [2] Mamoru Kurata, Numerical Analysis for Semiconductor Devices, D. C. Heath and Company, U. S. A., (1982).
- [3] W. Fichtner, D. J. Rose and R. E. Bank, *IEEE Trans. on Electr. Devices*, 30, 1018 (1983).
- [4] H. K. Gummel, *IEEE Trans. on Electr. Devices*, 11, 455(1964).
- [5] J. A. Meijerink and H. A. van der Vorst, *Math. of Compu.*, 31, 148(1977).
- [6] T. A. Manteuffel, *Math. of Compu.*, 34, 473 (1980).
- [7] E. M. Buturla *Solid-State Electronics*, 23, 331, (1980).
- [8] P. E. Cottrell, Proceeding of the Inter. Electr. Devices Meeting, P. 51, Dec. 1975.

## Three Dimensional Numerical Simulation of MOS Devices

Chen Datong, Wu Qiming, Wang Zeyi and Li Zhijian

(Institute of Microelectronics and Department of Computer, Qinghua University)

### Abstract

Based on the finite difference solution of respective equations, a computer analysis program, TDMOS-1, for 2 and 3-dimensional simulation of characterization of MOS devices has been developed. A self consistent solution of Poisson's equation and current continuity equation by using Gummel's approach is carried out. Newton-Raphson algorithm for linearization of discreted Poisson's equation and ICCG method for resolving linear equations are utilized. By reasonably controlling the Newton iteration times, suitably choosing the incomplete cholesky decomposition matrix and by taking the 2-D analysis results as the initial values for 3-D analysis, the computational time is remarkably reduced. On a VAX11/750 mainframe, for each potential point, about 20 seconds for 2-D analysis and 5 minutes for 3-D analysis are needed.