

三维半导体器件漂移扩散模型的并行有限元方法研究^{*1)}

王芹 马召灿 白石阳 张林波²⁾ 卢本卓²⁾
(LSEC, 国家数学与交叉科学中心, 中国科学院数学与系统科学研究院,
计算数学与科学工程计算研究所, 北京 100190;
中国科学院大学数学科学学院, 北京 100049)

李鸿亮²⁾
(四川师范大学数学科学学院, 成都 610066)

摘要

本文设计了一种新的三维自适应迎风稳定化有限元方法 (SUPG-IP), 并对比研究了多种半导体器件模拟的并行有限元方法. 数值模拟结果表明: 稳定化有限元方法适用于大偏压以及高掺杂器件模拟; 而经典的 Zlamal 有限元方法更适用于计算半导体器件的电学响应曲线. 我们基于三维并行自适应有限元平台 PHG 开发了半导体器件漂移扩散模型求解器 DevSim, 并对几种典型的半导体器件进行了模拟测试. 计算结果与商业软件 Sentaurus 吻合较好, 验证了算法的有效性. 我们对 PN 结进行了超大规模网格并行模拟测试, 网格达 8 亿单元并使用 2048 进程计算, 展示了算法良好的并行可扩展性.

关键词: 半导体器件; 漂移扩散模型; 有限元方法; 并行数值模拟; DevSim
MR (2010) 主题分类: 62P35

RESEARCH ON PARALLEL FINITE ELEMENT METHODS FOR THE DRIFT-DIFFUSION MODEL IN SEMICONDUCTOR DEVICE SIMULATIONS

Wang Qin Ma Zhaocan Bai Shiyang Zhang Linbo Lu Benzhuo
(LSEC, NCMIS, Institute of Computational Mathematics and Scientific/Engineering Computing,
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China;
School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China)

Li Hongliang
(Department of Mathematics, Sichuan Normal University, Chengdu 610066, China)

Abstract

In this paper, we propose a novel 3D adaptive upwind stabilized finite element method (SUPG-IP), and perform a study to compare several parallel finite element methods for semiconductor device simulations. Numerical results show that the stabilized finite element methods are applicable to problems with large biases and high doping concentrations, while

^{*} 2020年4月2日收到.

¹⁾ 基金项目: 科学挑战专题 (TZ2016003), 国家重点研发计划 (2016YFB0201304), NSFC(11771435).

²⁾ 通讯作者: 李鸿亮, lihongliang@mtrc.ac.cn; 张林波, zlb@lsec.cc.ac.cn; 卢本卓, bzlu@lsec.cc.ac.cn.

the classic Zlamal finite element method is more suitable for calculation of electrical characteristics of devices. Based on the three-dimensional parallel adaptive finite element toolbox PHG, a Drift-Diffusion model solver, DevSim, for semiconductor device simulations is developed. It is applied to simulating several typical semiconductor devices. The results are consistent with those of the commercial software Sentaurus, which verifies the effectiveness of the algorithms. A large-scale numerical experiment of PN diode, with 800 million elements and 2048 MPI processes, was carried out, demonstrating good parallel scalability of the algorithms.

Keywords: semiconductor device; drift-diffusion model; finite element methods; parallel simulation; DevSim

2010 Mathematics Subject Classification: 62P35

1. 引言

半导体器件是利用半导体材料(如硅、锗等)的特殊电学特性来完成特定功能的电子器件. 随着微电子技术的快速发展, 对半导体器件的技术要求也越来越高. 通过实验的方法研制新的半导体器件需要经过一个“设计 - 试制 - 测试 - 修改设计 - 再试制”的过程, 不断重复直至达到产品的要求, 效率低下且代价十分昂贵. 因此, 各种半导体工艺模拟及器件模拟(TCAD)软件应运而生. 在新器件的研制过程中, 借助 TCAD 软件预测它的电学特性, 可以大大缩短新器件的设计制造周期, 降低制造成本. 然而现阶段国内的半导体数值分析工作绝大多数还是依靠国外的商业软件, 目前 TCAD 商业软件的两大巨头分别是 Synopsys 公司的 Sentaurus^[1]和 Silvaco 公司的 Atlas^[2], 迄今为止国内还没有有一款真正通用的半导体器件模拟软件.

在半导体器件中, 电流传导靠的是电子和空穴的移动, 因此电子和空穴被统称为载流子. 我们可以通过建立载流子的输运模型来模拟半导体器件的电学特性. 最早被使用的载流子输运模型是 VanRoosbroeck 于 1950 年提出的漂移扩散模型(Drift-Diffusion Model)^[3], 该模型包括一个 Poisson 方程和两个载流子连续性方程. 其中, Poisson 方程确定了电势的分布, 而连续性方程则描述了在电场、浓度梯度等作用下的载流子浓度分布. 通过求解漂移扩散模型, 我们可以得到半导体器件在不同偏压下的载流子输运行为, 从而得到“电流 - 电压”特性曲线. 在漂移扩散模型的基础上, 还有一些更为复杂的输运模型, 如流体动力学模型. 它包括了一个 Poisson 方程, 两个连续性方程, 两个动量守恒方程和两个能量守恒方程. 流体动力学模型能够捕捉到所谓的“velocity overshoot”现象, 即非线性双曲型方程中的激波现象. 但是流体动力学模型的求解十分困难, 目前学术界提出的算法均不是十分完善. 此外, 在实际生产和应用中, 器件的稳定性还会受到制造工艺、环境等各种因素的影响, 如生产过程中使用的化学试剂、高温器材和人体沾污或设备使用过程中来自周围环境的电离辐射等, 具体可参考本刊的另一篇文章^[4], 这里我们不再赘述.

漂移扩散模型作为最早被用于半导体器件模拟的模型, 迄今已有几十年历史. 该模型适用于微米级别的器件模拟, 一直占据着半导体器件模拟的主导地位. 最初, 人们通过解析的方法对该模型进行分析, 这种方法对于一维的分析行之有效, 但对于二维或三维的分析很难进行. 1964 年, Gummel^[5]首次使用数值方法代替解析方法, 开启了半导体数值模拟的时代. 由于有限差分法易于实施的特性和有限体积法流守恒的性质, 这两种方法在求解漂移扩散模型中被广泛应用^[6], 其中以 Scharfetter-Gummel 格式^[7]最为著名. 有限元方法^[8]在许多领域取得了很大的成功, 但在半导体方程中却比较少用. 有限元方法有很多内在的优点, 它善于处理复杂的几何形状和边界条件, 适于构造高精度的格式以及适于进行自适应计算等. 此外, 网

格生成技术和软件的发展^[9-12]也促使有限元方法成为求解半导体方程的有力工具.

然而,一般的有限元方法仍然会存在一些问题.当器件中的电场强度较强或载流子浓度变化比较剧烈时,漂移扩散模型中的载流子连续性方程可以看做典型的对流占优的对流扩散方程.众所周知,此类方程的真解存在边界层或内部层,一般的有限元方法很容易产生非物理的数值震荡.为克服数值震荡,常用的方法是在数值格式中使用迎风技术^[13]或指数拟合技术^[14].目前学术界有很多针对对流扩散问题的稳定化有限元方法,其中比较成功的方法之一就是 Streamline-upwind/Petrov-Galerkin (SUPG) 有限元^[15].另外还有一类稳定化方法,即由 Douglas 和 Dupont 提出的边稳定化方法^[16],该方法通过添加单元边界法向导数的跳量罚项来提高有限元解的光滑性,抑制数值震荡.基于上述两种稳定化方法,本文设计了自适应迎风稳定化有限元方法 (SUPG-IP).除此之外,我们可以基于漂移扩散方程的特有属性,设计专门针对它的有限元离散格式.漂移扩散模型中,对流运动是由电场 $E = -\nabla\psi$ (ψ 为静电势) 驱动的,载流子浓度在空间中呈指数分布.因此,指数拟合技巧在构造连续性方程的离散格式时十分常用.例如,在 Scharfetter-Gummel 格式的推导中,就假设了电流在网格单元的边上为常数,得到了载流子浓度包含指数项的解析表达式,再基于这个表达式进行求导运算,进而得到了电流更精确的数值离散格式.1983年,Zlamal^[17]将指数拟合技巧运用到了针对漂移扩散模型的有限元离散格式的设计中.他使用分片线性基函数对连续性方程进行有限元离散,并在计算变分问题中的双线性型时,利用电流密度在四面体单元内为常数的假设得到了积分点处更精确的电流密度值.本文对以上几种有限元方法在漂移扩散模型中的应用和发展进行了详细介绍,并通过几种典型的半导体器件的数值模拟实验来验证这些方法的有效性.基于三维并行自适应有限元软件平台 Parallel Hierarchical Grid (PHG)^[18],我们实现了以上算法,并研制了三维半导体器件并行有限元求解器 DevSim.同时,DevSim 也嵌入到了本小组开发的半导体器件及辐照损伤效应仿真的应用软件系统 3Ddevice 中,详情可参考 [19].

本文剩余部分安排如下.在第2节中,我们详细介绍求解漂移扩散模型的常用技巧和几种有限元方法的数值离散格式.第3节展示几种典型半导体器件的数值实验结果,验证这些方法的实用性.第4节对本文的工作进行总结.

2. 求解漂移扩散模型的几种有限元方法

记求解区域 $\Omega \subset \mathbb{R}^3$,经典的漂移扩散模型^[3]包括一个 Poisson 方程和两个载流子连续性方程:

$$\begin{aligned} \text{Poisson 方程: } & -\nabla \cdot (\epsilon \nabla \psi) = q(p - n + N_D - N_A) & \text{in } \Omega, \\ \text{空穴连续性方程: } & \frac{\partial p}{\partial t} = -\nabla \cdot \mathbf{J}_p - R_p(\psi, p, n) & \text{in } \Omega, \\ \text{电子连续性方程: } & \frac{\partial n}{\partial t} = -\nabla \cdot \mathbf{J}_n - R_n(\psi, p, n) & \text{in } \Omega. \end{aligned}$$

其中,

$$\begin{aligned} \mathbf{J}_p &= -D_p \nabla p + \mu_p p \left(-\nabla \psi + \frac{k_B T}{q} \nabla \log n_{ie} \right), \\ \mathbf{J}_n &= -D_n \nabla n - \mu_n n \left(-\nabla \psi - \frac{k_B T}{q} \nabla \log n_{ie} \right). \end{aligned}$$

待求解的变量是静电势 ψ 、空穴浓度 p 和电子浓度 n , \mathbf{J}_p 和 \mathbf{J}_n 分别是空穴电流密度和电子电流密度. Poisson 方程中, ϵ 是介电常数, q 是单位电荷, N_A 是受主杂质俘获电子后留在晶格

上的带负电的离子浓度, N_D 是施主杂质贡献一个电子后留在晶格上的带正电的离子浓度. 连续性方程中, D_p 和 D_n 分别是空穴和电子的扩散系数, μ_p 和 μ_n 分别是空穴和电子的迁移率, R_p 和 R_n 分别是空穴和电子的复合产生项, k_B 是 Boltzmann 常数, T 是温度, n_{ie} 是半导体材料的本征浓度. 若半导体材料为同质结且不考虑能带变窄效应, 则 n_{ie} 为常数, $\nabla \log n_{ie} = 0$, 且 Einstein 关系

$$D_p = \frac{k_B T}{q} \mu_p, \quad D_n = \frac{k_B T}{q} \mu_n$$

成立. 我们引入变量 $u = \frac{q\psi}{k_B T}$ 对静电势 ψ 进行无量纲化, 则方程组变为 (这里主要考虑稳态的情况):

$$-\nabla \cdot (\epsilon \nabla u) = \frac{q^2}{k_B T} (p - n + N_D - N_A), \quad \text{in } \Omega, \quad (2.1)$$

$$\nabla \cdot \mathbf{J}_p = -R_p(\psi, p, n), \quad \text{in } \Omega, \quad (2.2)$$

$$\nabla \cdot \mathbf{J}_n = -R_n(\psi, p, n), \quad \text{in } \Omega. \quad (2.3)$$

其中

$$\mathbf{J}_p = -D_p(\nabla p + p \nabla u),$$

$$\mathbf{J}_n = -D_n(\nabla n - n \nabla u).$$

在边界上, 有

$$u|_{\partial\Omega_{D,\psi}} = u_D, \quad (2.4)$$

$$p|_{\partial\Omega_{D,c}} = p_D, \quad (2.5)$$

$$n|_{\partial\Omega_{D,c}} = n_D, \quad (2.6)$$

$$\nabla u \cdot \mathbf{n}|_{\partial\Omega_{N,\psi}} = 0, \quad (2.7)$$

$$\mathbf{J}_p \cdot \mathbf{n}|_{\partial\Omega_{N,c}} = \mathbf{J}_n \cdot \mathbf{n}|_{\partial\Omega_{N,c}} = 0. \quad (2.8)$$

其中 \mathbf{n} 为边界单位外法向, $\partial\Omega_{D,\psi}$, $\partial\Omega_{D,c}$ 和 $\partial\Omega_{N,\psi}$, $\partial\Omega_{N,c}$ 分别是电势和载流子浓度的 Dirichlet 型和 Neumann 型边界, 且有 $\partial\Omega = \partial\Omega_{D,\psi} \cup \partial\Omega_{N,\psi} = \partial\Omega_{D,c} \cup \partial\Omega_{N,c}$.

首先, 我们引入物理分析上和计算上常用的变量 ϕ_p 和 ϕ_n , 它们称为准费米势, 与电势和载流子浓度有如下关系:

$$p = n_{ie} e^{\phi_p - u}, \quad n = n_{ie} e^{u - \phi_n}.$$

准费米势在后续欧姆接触边界的处理中将会使用. 其次, 我们引入数学分析上常用的 Slotboom 变量 Φ_p 和 Φ_n 如下:

$$\Phi_p = p e^u, \quad \Phi_n = n e^{-u}.$$

变量替换后, 可得 \mathbf{J}_p 和 \mathbf{J}_n 为:

$$\mathbf{J}_p = -D_p(\nabla p + p \nabla u) = -D_p p \nabla \phi_p = -D_p e^{-u} \nabla \Phi_p,$$

$$\mathbf{J}_n = -D_n(\nabla n - n \nabla u) = D_n n \nabla \phi_n = -D_n e^u \nabla \Phi_n.$$

在 2.3 节中, 我们将基于 Slotboom 变量引入求解漂移扩散模型的 Zlamal 有限元方法.

2.1. 标准有限元方法

在这一小节, 我们将详细介绍漂移扩散模型的标准有限元方法的数值离散格式. 1964 年, Gummel 首次使用解耦迭代算法^[5] 对一维稳态双极性晶体管的电特性进行分析, 由此开启了半导体数值模拟的时代. Gummel 迭代算法的核心思想是将方程组解耦后分别求解, 通过迭代循环使得单个方程的解逐渐逼近整体方程组的解, 其主要步骤为, 在每个迭代步中依次求解各个单独的方程, 并将之前求得的解作为已知量代入即将求解的方程中.

解耦后, Poisson 方程和两个连续性方程分别单独求解. 首先考虑 Poisson 方程的数值离散格式. 定义 $H^1(\Omega)$ 的子空间 $H_u^1(\Omega) = \{v \in H^1(\Omega) : v|_{\partial\Omega_{D,\psi}} = 0\}$, Poisson 方程 (2.1) 的弱形式为: 求 $u \in H^1(\Omega)$ 满足边界条件 (2.4) 使得

$$\int_{\Omega} \epsilon \nabla u \cdot \nabla v d\Omega = \frac{q^2}{k_B T} \int_{\Omega} (p - n + N_D - N_A) v d\Omega, \quad \forall v \in H_u^1(\Omega).$$

对于连续性方程, 方程中的复合项 $R_p(\psi, p, n)$ 和 $R_n(\psi, p, n)$ 通常是关于 p 和 n 的非线性函数, 需要对其进行一定的线性化处理, 可以采用显式或半隐式的方法. 为叙述简洁, 这里我们采用显式方法, 即在求解连续性方程时设

$$R_p(\psi, p, n) = R_p(\psi^*, p^*, n^*), \quad R_n(\psi, p, n) = R_n(\psi^*, p^*, n^*),$$

其中 (ψ^*, p^*, n^*) 是上一迭代步的计算结果. 于是要求解的方程为

$$\begin{aligned} -\nabla \cdot D_p(\nabla p + p \nabla u) &= -R_p(\psi^*, p^*, n^*), & \text{in } \Omega, \\ -\nabla \cdot D_n(\nabla n - n \nabla u) &= -R_n(\psi^*, p^*, n^*), & \text{in } \Omega. \end{aligned}$$

定义 $H^1(\Omega)$ 的子空间 $H_c^1(\Omega) = \{w \in H^1(\Omega) : w|_{\partial\Omega_{D,c}} = 0\}$, 则空穴连续性方程的弱形式为: 求 $p \in H^1(\Omega)$ 满足边界条件 (2.5) 使得

$$A_p(p, w) := \int_{\Omega} D_p(\nabla p + p \nabla u) \cdot \nabla w d\Omega = - \int_{\Omega} R_p w d\Omega, \quad \forall w \in H_c^1(\Omega), \quad (2.9)$$

以及电子连续性方程的弱形式为: 求 $n \in H^1(\Omega)$ 满足边界条件 (2.6) 使得

$$A_n(n, w) := \int_{\Omega} D_n(\nabla n - n \nabla u) \cdot \nabla w d\Omega = - \int_{\Omega} R_n w d\Omega, \quad \forall w \in H_c^1(\Omega). \quad (2.10)$$

设 \mathcal{T}_h 是区域 Ω 的协调四面体网格剖分, P_m 是三维 m 阶多项式, 定义 $H^1(\Omega)$ 的有限元子空间 $H_h^1(\Omega)$ 如下:

$$H_h^1(\Omega) := \{f_h \in C(\bar{\Omega}) : f_h|_T \in P_m, \forall T \in \mathcal{T}_h\}.$$

定义 H_u^1 和 H_c^1 的子空间 V_h 和 W_h 为有限元测试函数空间,

$$\begin{aligned} V_h &:= \{v_h \in C(\bar{\Omega}) : v_h|_{\partial\Omega_{D,\psi}} = 0, v_h|_T \in P_m, \forall T \in \mathcal{T}_h\}, \\ W_h &:= \{w_h \in C(\bar{\Omega}) : w_h|_{\partial\Omega_{D,c}} = 0, w_h|_T \in P_m, \forall T \in \mathcal{T}_h\}. \end{aligned}$$

定义有限元函数 $u_{D,h}, p_{D,h}, n_{D,h} \in H_h^1$, 使得在有限元空间 H_h^1 中的所有插值节点 x_h 上满足

$$\begin{aligned} u_{D,h}(x_h) &= \begin{cases} u_D(x_h), & x_h \in \partial\Omega_{D,\psi} \\ 0, & x_h \notin \partial\Omega_{D,\psi} \end{cases} \\ p_{D,h}(x_h) &= \begin{cases} p_D(x_h), & x_h \in \partial\Omega_{D,c} \\ 0, & x_h \notin \partial\Omega_{D,c} \end{cases} \end{aligned}$$

$$n_{D,h}(x_h) = \begin{cases} n_D(x_h), & x_h \in \partial\Omega_{D,c} \\ 0, & x_h \notin \partial\Omega_{D,c} \end{cases}$$

则 Poisson 方程的离散弱形式为: 求 $u_h \in u_{D,h} + V_h$ 使得

$$\int_{\Omega} \epsilon \nabla u_h \cdot \nabla v_h d\Omega = \frac{q^2}{k_B T} \int_{\Omega} (p_h - n_h + N_D - N_A) v_h d\Omega, \quad \forall v_h \in V_h. \quad (2.11)$$

其中 $p_h \in p_{D,h} + W_h$ 和 $n_h \in n_{D,h} + W_h$ 作为已知量处理. 空穴连续性方程的离散弱形式为: 求 $p_h \in p_{D,h} + W_h$ 使得

$$A_p(p_h, w_h) := \int_{\Omega} D_p(\nabla p_h + p_h \nabla u_h) \cdot \nabla w_h d\Omega = - \int_{\Omega} R_p w_h d\Omega, \quad \forall w_h \in W_h,$$

以及电子连续性方程的离散弱形式为: 求 $n_h \in n_{D,h} + W_h$ 使得

$$A_n(n_h, w_h) := \int_{\Omega} D_n(\nabla n_h - n_h \nabla u_h) \cdot \nabla w_h d\Omega = - \int_{\Omega} R_n w_h d\Omega, \quad \forall w_h \in W_h.$$

其中 u_h 也作为已知量处理.

设 $\{v_j\}_{j=1}^N$ 是有限元空间 V_h 的一组基, $\{w_j\}_{j=1}^{\tilde{N}}$ 是 W_h 的一组基, 则对于任意的 $u_h \in u_{D,h} + V_h$, 其可以表示为如下形式:

$$u_h = u_{D,h} + \sum_{j=1}^N u_j v_j$$

将 u_h 代入离散弱形式 (2.11), 并取 $v_h \in \{v_j\}_{j=1}^N$ 可得:

$$\begin{aligned} \sum_{j=1}^N u_j \int_{\Omega} \epsilon \nabla v_j \cdot \nabla v_i d\Omega &= \frac{q^2}{k_B T} \int_{\Omega} (p_h - n_h + N_D - N_A) v_i d\Omega \\ &\quad - \int_{\Omega} \epsilon \nabla u_{D,h} \cdot \nabla v_i d\Omega, \quad i = 1, \dots, N. \end{aligned}$$

于是, Poisson 方程的离散格式可以写成矩阵形式 $MU = F$, 其中 $M = [M_{ij}] \in \mathbb{R}^{N \times N}$ 且

$$M_{ij} = \int_{\Omega} \epsilon \nabla v_j \cdot \nabla v_i d\Omega,$$

$F = [F_i] \in \mathbb{R}^N$ 且

$$F_i = \frac{q^2}{k_B T} \int_{\Omega} (p_h - n_h + N_D - N_A) v_i d\Omega - \int_{\Omega} \epsilon \nabla u_{D,h} \cdot \nabla v_i d\Omega.$$

其中当 v_i 在包含边界 $\partial\Omega_{D,\psi}$ 的单元内取值非零时, 体积分 $\int_{\Omega} \epsilon \nabla u_{D,h} \cdot \nabla v_i d\Omega$ 非零.

类似地, 我们可以将任意的 $p_h \in p_{D,h} + W_h$ 和 $n_h \in n_{D,h} + W_h$ 表示为:

$$\begin{aligned} p_h &= p_{D,h} + \sum_{j=1}^{\tilde{N}} p_j w_j, \\ n_h &= n_{D,h} + \sum_{j=1}^{\tilde{N}} n_j w_j, \end{aligned}$$

分别代入空穴连续性方程和电子连续性方程的离散弱形式可得:

$$\begin{aligned} & \sum_{j=1}^{\tilde{N}} p_j \int_{\Omega} D_p(\nabla w_j \cdot \nabla w_i + w_j \nabla u_h \cdot \nabla w_i) d\Omega \\ &= - \int_{\Omega} R_p w_i - \int_{\Omega} D_p(\nabla p_{D,h} \cdot \nabla w_i + p_{D,h} \nabla u_h \cdot \nabla w_i) d\Omega, \quad i = 1, \dots, \tilde{N}, \\ & \sum_{j=1}^{\tilde{N}} n_j \int_{\Omega} D_n(\nabla w_j \cdot \nabla w_i - w_j \nabla u_h \cdot \nabla w_i) d\Omega \\ &= - \int_{\Omega} R_n w_i - \int_{\Omega} D_n(\nabla n_{D,h} \cdot \nabla w_i - n_{D,h} \nabla u_h \cdot \nabla w_i) d\Omega, \quad i = 1, \dots, \tilde{N}. \end{aligned}$$

相应的矩阵形式为 $MU = F$, $M = [M_{ij}] \in \mathbb{R}^{\tilde{N} \times \tilde{N}}$, $F = [F_i] \in \mathbb{R}^{\tilde{N}}$. 对于空穴连续性方程,

$$\begin{aligned} M_{ij} &= \int_{\Omega} D_p(\nabla w_j \cdot \nabla w_i + w_j \nabla u_h \cdot \nabla w_i) d\Omega, \\ F_i &= - \int_{\Omega} R_p w_i - \int_{\Omega} D_p(\nabla p_{D,h} \cdot \nabla w_i + p_{D,h} \nabla u_h \cdot \nabla w_i) d\Omega. \end{aligned}$$

对于电子连续性方程,

$$\begin{aligned} M_{ij} &= \int_{\Omega} D_n(\nabla w_j \cdot \nabla w_i - w_j \nabla u_h \cdot \nabla w_i) d\Omega, \\ F_i &= - \int_{\Omega} R_n w_i - \int_{\Omega} D_n(\nabla n_{D,h} \cdot \nabla w_i - n_{D,h} \nabla u_h \cdot \nabla w_i) d\Omega. \end{aligned}$$

其中, 当 w_i 在包含边界 $\partial\Omega_{D,c}$ 的单元内取值非零时, 体积分 $\int_{\Omega} D_p(\nabla p_{D,h} \cdot \nabla w_i + p_{D,h} \nabla u_h \cdot \nabla w_i) d\Omega$ 和 $\int_{\Omega} D_n(\nabla n_{D,h} \cdot \nabla w_i - n_{D,h} \nabla u_h \cdot \nabla w_i) d\Omega$ 非零.

2.2. 稳定化有限元方法

在漂移扩散模型中, 随着电极上的电压和掺杂浓度的不断提高, 方程中对流项的强度逐渐增强, 漂移扩散方程逐渐变为对流占优, 此类方程的真解本身存在边界层或内部层, 因此标准有限元方法很容易产生非物理的数值震荡, 从而导致收敛到的解不够精确甚至难以收敛. 为了克服对流占优的问题, 我们引入了具有迎风属性的 SUPG 方法. 另外, 我们还探讨了一类面稳定化方法, 通过添加单元边界法向导数的跳量罚项来提高有限元解的光滑性, 抑制界面和边界上的数值震荡. 下面我们将详细介绍这几种稳定化方法的数值离散格式.

2.2.1. Streamline-upwind/Petrov-Galerkin(SUPG) 方法

在这一小节中, 我们将介绍 SUPG 稳定化方法^[15]的数值格式. 由于稳定化是作用在连续性方程上的, 我们以空穴连续性方程 (2.2) 为例,

$$-\nabla \cdot D_p(\nabla p + p \nabla u) = -R_p.$$

定义 Peclet 数 Pe 作为对流强度的指示子:

$$Pe_T = \frac{\|\mathbf{b}_p\|_2 h_T}{2D_p},$$

其中 h_T 为单元 T 的直径, $\mathbf{b}_p = -D_p \nabla u$ 是相应的速度场. Pe_T 表示 Pe 在单元 T 上的值. 通常, 当 Pe 大于 1 时, 方程变为对流占优, 此时稳定化是必要的^[15]. SUPG 方法在连续性方程

的弱形式 (2.9) 上添加稳定化项, 该稳定化项及相应的右端项定义如下:

$$S_p(p, w) := \sum_T \int_T (-\nabla \cdot D_p(\nabla p + p\nabla u)) \cdot w_{\text{supg}} dT,$$

$$L_p(w) := \sum_T \int_T -R_p w_{\text{supg}} dT,$$

其中 $w_{\text{supg}} = \sigma_T \mathbf{b}_p \cdot \nabla w$. 稳定化参数 σ_T 定义如下:

$$\sigma_T = \frac{h_T}{2\|\mathbf{b}_p\|_2} \xi(Pe_T), \quad \xi(Pe_T) = \begin{cases} Pe_T/3, & 0 \leq Pe_T \leq 3, \\ 1, & Pe_T > 3. \end{cases}$$

则空穴连续性方程的 SUPG 稳定化弱形式为: 求 $p \in H^1(\Omega)$ 且满足边界条件 (2.5) 使得

$$A_p(p, w) + S_p(p, w) = \int_{\Omega} -R_p w d\Omega + L_p(w), \quad \forall w \in H_c^1(\Omega).$$

类似地可得电子连续性方程的 SUPG 稳定化项为:

$$S_n(n, w) := \sum_T \int_T (-\nabla \cdot D_n(\nabla n - n\nabla u)) \cdot w_{\text{supg}} dT,$$

$$L_n(w) := \sum_T \int_T -R_n w_{\text{supg}} dT.$$

其中, $w_{\text{supg}} = \sigma_T \mathbf{b}_n \cdot \nabla w$, $\mathbf{b}_n = -D_n \nabla u$. 稳定化参数 σ_T 定义为:

$$\sigma_T = \frac{h_T}{2\|\mathbf{b}_n\|_2} \xi(Pe_T)$$

则电子连续性方程的 SUPG 稳定化弱形式为: 求 $n \in H^1(\Omega)$ 且满足边界条件 (2.6) 使得

$$A_n(n, w) + S_n(n, w) = \int_{\Omega} -R_n w d\Omega + L_n(w), \quad \forall w \in H_c^1(\Omega).$$

2.2.2. Interior penalty(IP) 方法

IP 方法是一类通过添加单元边界法向导数的跳量罚项来提高有限元解的光滑性的稳定化方法. 该方法最初由 Douglas 和 Dupont 提出^[16], 用于求解椭圆和抛物方程. 后来, Burman^[20] 等人在双曲极限和原始先验误差估计下证明了它的稳定性. 设 $\mathcal{F}_h^{\text{int}}$ 表示内部单元边界的集合, 仍以空穴连续性方程 (2.2) 为例, IP 方法在其弱形式上添加了如下的跳量罚项:

$$B_p(p, w) := \sum_{F \in \mathcal{F}_h^{\text{int}}} \alpha h_F^2 \int_F [\nabla p \cdot \mathbf{n}] [\nabla w \cdot \mathbf{n}] dF.$$

其中 \mathbf{n} 是面 F 的单位外法向, h_F 是 F 的直径, α 是一个正的常数. 则空穴连续性方程的 IP 稳定化弱形式为: 求 $p \in H^1(\Omega)$ 且满足边界条件 (2.5) 使得

$$A_p(p, w) + B_p(p, w) = \int_{\Omega} -R_p w d\Omega, \quad \forall w \in H_c^1(\Omega).$$

类似地, 对于电子连续性方程有:

$$B_n(n, w) := \sum_{F \in \mathcal{F}_h^{\text{int}}} \alpha h_F^2 \int_F [\nabla n \cdot \mathbf{n}] [\nabla w \cdot \mathbf{n}] dF.$$

则电子连续性方程的 IP 稳定化弱形式为: 求 $n \in H^1(\Omega)$ 且满足边界条件 (2.6) 使得

$$A_n(n, w) + B_n(n, w) = \int_{\Omega} -R_n w d\Omega, \quad \forall w \in H_c^1(\Omega).$$

2.2.3. SUPG-IP 方法

为了结合 SUPG 方法的迎风属性和 IP 方法数值解的光滑性, 我们提出一种新的将体积分与面积分相结合的稳定化方法, 称为 SUPG-IP 方法. 仍以空穴连续性方法为例, 该方法在其弱形式上添加如下稳定化项:

$$C_p(p, w) = \sum_T \int_T (-\nabla \cdot D_p(\nabla p + p\nabla u)) \cdot w_{\text{supg}} dT + \sum_{F \in \mathcal{F}_h^{\text{int}}} \zeta \tau_F^{\text{int}} \int_F [\nabla p \cdot \mathbf{n}] [\nabla w \cdot \mathbf{n}] dF.$$

相应的右端项为:

$$L_p(w) = \sum_T \int_T -R_p w_{\text{supg}} dT.$$

其中, ζ 是一个正的常数. 稳定化参数 τ_F^{int} 定义如下^[21]:

$$\tau_F^{\text{int}} = \frac{h_F^3 \|\mathbf{b}_p\|_{L^2(F)}^2}{\|\mathbf{b}_p\|_{L^2(F)} h_F + D_p}.$$

则空穴连续性方程的 SUPG-IP 稳定化弱形式为: 求 $p \in H^1(\Omega)$ 且满足边界条件 (2.5) 使得

$$A_p(p, w) + C_p(p, w) = \int_{\Omega} -R_p w d\Omega + L_p(w), \quad \forall w \in H_c^1(\Omega).$$

类似地, 对于电子连续性方程有:

$$C_n(n, w) := \sum_T \int_T (-\nabla \cdot D_n(\nabla n - n\nabla u)) \cdot w_{\text{supg}} dT + \sum_{F \in \mathcal{F}_h^{\text{int}}} \zeta \tau_F^{\text{int}} \int_F [\nabla n \cdot \mathbf{n}] [\nabla w \cdot \mathbf{n}] dF,$$

$$L_n(w) := \sum_T \int_T -R_n w_{\text{supg}} dT.$$

其中,

$$\tau_F^{\text{int}} = \frac{h_F^3 \|\mathbf{b}_n\|_{L^2(F)}^2}{\|\mathbf{b}_n\|_{L^2(F)} h_F + D_n},$$

则电子连续性方程的 SUPG-IP 稳定化弱形式为: 求 $n \in H^1(\Omega)$ 且满足边界条件 (2.6) 使得

$$A_n(n, w) + C_n(n, w) = \int_{\Omega} -R_n w d\Omega + L_n(w), \quad \forall w \in H_c^1(\Omega).$$

2.3. Zlamal 有限元方法

上一节中介绍的几种稳定化方法主要基于 SUPG 方法, 通过使用迎风技术来克服数值震荡. 除此之外, 我们还可以基于漂移扩散模型的特有属性, 使用指数拟合技术设计专门针对它的有限元离散格式. 在本节中, 我们将重点介绍基于指数拟合思想的 Zlamal 有限元方法^[17], 该方法使用分片线性基函数对漂移扩散模型进行有限元离散.

采用 Slotboom 变量来表述连续性方程:

$$\text{空穴连续性方程: } \nabla \cdot \mathbf{J}_p = -R_p,$$

$$\text{电子连续性方程: } \nabla \cdot \mathbf{J}_n = -R_n.$$

其中, $\mathbf{J}_p = -D_p e^{-u} \nabla \Phi_p$, $\mathbf{J}_n = -D_n e^u \nabla \Phi_n$. 这里, 我们考虑如下的齐次边界条件,

$$\begin{aligned} \Phi_p|_{\partial\Omega_{D,c}} &= 0, \quad \Phi_n|_{\partial\Omega_{D,c}} = 0, \\ (D_p e^{-u} \nabla \Phi_p) \cdot \mathbf{n}|_{\partial\Omega_{N,c}} &= 0, \quad (D_n e^u \nabla \Phi_n) \cdot \mathbf{n}|_{\partial\Omega_{N,c}} = 0. \end{aligned}$$

记 $H_c^1(\Omega) = \{w \in H^1(\Omega) : w|_{\partial\Omega_{D,c}} = 0\}$, 则连续性方程的弱形式为: 求 $\Phi_p \in H_c^1(\Omega)$ 和 $\Phi_n \in H_c^1(\Omega)$ 使得

$$\begin{aligned} A_p(\Phi_p, w) &:= - \int_{\Omega} \mathbf{J}_p \cdot \nabla w d\Omega = - \int_{\Omega} R_p w d\Omega, \quad \forall w \in H_c^1(\Omega), \\ A_n(\Phi_n, w) &:= - \int_{\Omega} \mathbf{J}_n \cdot \nabla w d\Omega = - \int_{\Omega} R_n w d\Omega, \quad \forall w \in H_c^1(\Omega). \end{aligned}$$

设 \mathcal{T}_h 是区域 Ω 的协调四面体网格剖分, $\{x^i\}_{i=1}^{\tilde{N}}$ 为网格上的顶点集合. 假设电流密度 \mathbf{J}_p 和 \mathbf{J}_n 在单元 $T \in \mathcal{T}_h$ 内为常数, 且设 \mathbf{J}_p 和 \mathbf{J}_n 在单元 T 上的逼近分别为常向量 \mathbf{j}_p 和 \mathbf{j}_n , 即

$$\mathbf{j}_p = -D_p e^{-u} \nabla \Phi_p, \quad \mathbf{j}_n = -D_n e^u \nabla \Phi_n.$$

设单元 T 的顶点为 x^1, x^2, x^3, x^4 , \hat{T} 为参考单元, 其顶点为 $(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)$. 定义如下 \hat{T} 到 T 的映射,

$$x = x(\xi) = x^1 \phi_1(\xi) + x^2 \phi_2(\xi) + x^3 \phi_3(\xi) + x^4 \phi_4(\xi), \quad \xi = (\xi_1, \xi_2, \xi_3) \in \hat{T}, \quad (2.12)$$

其中,

$$\phi_1(\xi) = 1 - \xi_1 - \xi_2 - \xi_3, \quad \phi_2(\xi) = \xi_1, \quad \phi_3(\xi) = \xi_2, \quad \phi_4(\xi) = \xi_3.$$

记 $\hat{u}(\xi) = u(x(\xi))$, $\hat{\Phi}_p(\xi) = \Phi_p(x(\xi))$, $\hat{\Phi}_n(\xi) = \Phi_n(x(\xi))$, 则 $\nabla \hat{\Phi}_p = J^T \nabla \Phi_p$, $\nabla \hat{\Phi}_n = J^T \nabla \Phi_n$, 其中 J 是从 \hat{T} 到 T 的映射的 Jacobian 矩阵,

$$J = \frac{\partial(x_1, x_2, x_3)}{\partial(\xi_1, \xi_2, \xi_3)}.$$

于是,

$$e^{\hat{u}} J^T \mathbf{j}_p = -D_p \nabla \hat{\Phi}_p, \quad e^{-\hat{u}} J^T \mathbf{j}_n = -D_n \nabla \hat{\Phi}_n. \quad (2.13)$$

记 $\Phi_{p1}, \Phi_{p2}, \Phi_{p3}, \Phi_{p4}$ 为 Φ_p 在顶点 x^1, x^2, x^3, x^4 处的值, $\Phi_{n1}, \Phi_{n2}, \Phi_{n3}, \Phi_{n4}$ 为 Φ_n 在顶点 x^1, x^2, x^3, x^4 处的值. 对式 (2.13) 的第 j ($j = 1, 2, 3$) 个分量沿着 ξ_j 轴的 $[0, 1]$ 区间进行积分, 有

$$\begin{aligned} -D_p(\Phi_{p,j+1} - \Phi_{p1}) &= \int_0^1 e^{\hat{u}_j} d\xi_j (J^T \mathbf{j}_p)_j, \\ -D_n(\Phi_{n,j+1} - \Phi_{n1}) &= \int_0^1 e^{-\hat{u}_j} d\xi_j (J^T \mathbf{j}_n)_j. \end{aligned}$$

其中, $\hat{u}_1 = \hat{u}(\xi_1, 0, 0)$, $\hat{u}_2 = \hat{u}(0, \xi_2, 0)$, $\hat{u}_3 = \hat{u}(0, 0, \xi_3)$, $(J^T \mathbf{j}_p)_j$ 和 $(J^T \mathbf{j}_n)_j$ 分别表示 $J^T \mathbf{j}_p$ 和 $J^T \mathbf{j}_n$ 的第 j 个分量. 由于求解 Poisson 方程时使用的是线性元, 电场强度 $\mathbf{E} = -\nabla \psi$ 在单元内是常向量, 那么 ∇u 也是常向量. 记 u_1, u_2, u_3, u_4 为 u 在顶点 x^1, x^2, x^3, x^4 处的值,

$$\nabla u = \begin{pmatrix} u_2 - u_1 \\ u_3 - u_1 \\ u_4 - u_1 \end{pmatrix},$$

则上式的积分可以解析地积出,

$$\begin{aligned} \int_0^1 e^{\hat{u}_j} d\xi_j &= \frac{1}{u_{j+1} - u_1} \int_0^1 \frac{d\hat{u}_j}{d\xi_j} e^{\hat{u}_j} d\xi_j = \frac{e^{u_{j+1}} - e^{u_1}}{u_{j+1} - u_1} \\ \int_0^1 e^{-\hat{u}_j} d\xi_j &= \frac{1}{u_{j+1} - u_1} \int_0^1 \frac{d\hat{u}_j}{d\xi_j} e^{-\hat{u}_j} d\xi_j = -\frac{e^{-u_{j+1}} - e^{-u_1}}{u_{j+1} - u_1} \end{aligned}$$

于是我们有

$$\begin{aligned} \mathbf{j}_p &= -D_p e^{-u_1} (J^T)^{-1} \tilde{B}_T \nabla \hat{\Phi}_p = -D_p e^{-u_1} (J^T)^{-1} \tilde{B}_T J^T \nabla \Phi_p, \\ \mathbf{j}_n &= -D_n e^{u_1} (J^T)^{-1} B_T \nabla \hat{\Phi}_n = -D_n e^{u_1} (J^T)^{-1} B_T J^T \nabla \Phi_n. \end{aligned}$$

其中, $B_T = \text{diag}(B(u_1 - u_2), B(u_1 - u_3), B(u_1 - u_4))$, $\tilde{B}_T = \text{diag}(B(u_2 - u_1), B(u_3 - u_1), B(u_4 - u_1))$, $B(\xi) = \xi(e^\xi - 1)^{-1}$.

定义有限元函数空间

$$W_h := \{w_h \in C(\bar{\Omega}) : w_h|_{\partial\Omega_{D,c}} = 0, w_h|_T \in P_1, \forall T \in \mathcal{T}_h\},$$

其中 P_1 为三维一阶多项式, 即 W_h 是片分线性多项式空间. 设 $\{w_j\}_{j=1}^{\tilde{N}}$ 是 W_h 的一组基, w_k 是对应于顶点 x^k 的基函数, 即 $w_k(x^j) = \delta_{jk}$, $j, k = 1, \dots, \tilde{N}$. 定义映射 (2.12) 使得顶点 x^k 映射到参考单元 \hat{T} 的顶点 $(0, 0, 0)$, 令 S_k 是所有包含顶点 x^k 的单元集合, 对任意的 $w_h = \sum_{k=1}^{\tilde{N}} \tilde{w}_k w_k(x) \in W_h$, 定义

$$\begin{aligned} A_{ph}(\Phi_p, w_h) &:= \sum_{k=1}^{\tilde{N}} \tilde{w}_k A_{ph}(\Phi_p, w_k), \\ A_{nh}(\Phi_n, w_h) &:= \sum_{k=1}^{\tilde{N}} \tilde{w}_k A_{nh}(\Phi_n, w_k). \end{aligned}$$

其中,

$$\begin{aligned} A_{ph}(\Phi_p, w_k) &:= \sum_{T \in S_k} - \int_T \mathbf{j}_p \cdot \nabla w_k dT = \sum_{T \in S_k} D_p e^{-u_k} \int_T (J^T)^{-1} \tilde{B}_T J^T \nabla \Phi_p \cdot \nabla w_k dT, \\ A_{nh}(\Phi_n, w_k) &:= \sum_{T \in S_k} - \int_T \mathbf{j}_n \cdot \nabla w_k dT = \sum_{T \in S_k} D_n e^{u_k} \int_T (J^T)^{-1} B_T J^T \nabla \Phi_n \cdot \nabla w_k dT. \end{aligned}$$

则 A_{ph} 和 A_{nh} 是关于 Φ_p, w_h 和 Φ_n, w_h 的对称双线性型. 我们定义连续性方程的有限元离散弱形式如下: 求 $\Phi_{ph} \in W_h$ 和 $\Phi_{nh} \in W_h$, 使得

$$\begin{aligned} A_{ph}(\Phi_{ph}, w_h) &= - \int_{\Omega} R_p w_h d\Omega, \quad \forall w_h \in W_h, \\ A_{nh}(\Phi_{nh}, w_h) &= - \int_{\Omega} R_n w_h d\Omega, \quad \forall w_h \in W_h. \end{aligned}$$

在一个四面体剖分中, 如果所有四面体的面夹角都小于 $\pi/2$, 称这个剖分为锐角剖分. 当 \mathcal{T}_h 是锐角剖分网格时, 上述问题所得到的刚度矩阵是 M - 矩阵, 该性质保证了 Φ_{ph} 和 Φ_{nh} 非负.

SUPG 方法采用了迎风技术, 通过引入人工扩散项来克服由对流占优引起的数值震荡, 根据单元上电势的变化调节人工扩散项的大小; 而 SUPG-IP 方法在 SUPG 迎风属性的基础上, 添加了单元边界法向导数的跳量罚项来提高有限元解的光滑性, 进一步增强了算法抑制数值震荡的能力. Zlamal 有限元方法则是基于指数拟合技巧, 通过假设电流密度在单元内为常数, 得到积分点处更精确的电流密度值, 是一种自适应的指数型迎风格式. 在下一节中, 我们将通过数值实验对这几种有限元方法进行比较.

3. 数值实验

基于三维并行自适应有限元软件平台 Parallel Hierarchical Grid (PHG) [18], 我们编写了上述几种有限元方法的程序, 并研制了相应的三维半导体器件并行有限元求解器 DevSim. 该求

解器使用 C 语言开发, 采用 MPI 通讯技术, 支持大规模并行. 用户只需导入半导体器件的网格文件、记录相应边界条件的边界标记文件和包含载流子迁移率等物理信息的粒子信息文件, 就可以调用上述任一方法对器件的电学特性进行模拟. 根据用户的选择, DevSim 可以在计算结束后输出含有电势和载流子浓度分布的 VTK 文件以及各电极上的电流等信息. 目前, DevSim 已经嵌入到了本小组开发的半导体器件及辐照损伤效应仿真的应用软件系统 3Ddevice 中, 详情可参考 [19].

在接下来的数值实验中, 我们将首先对这几种有限元方法进行对比研究, 然后对几种典型的半导体器件进行模拟, 并将 DevSim 的计算结果与商业软件 Sentaurus 进行比较, 最后对算法的并行性能进行测试. 所有的数值实验都是在“科学与工程计算国家重点实验室” LSSC-IV 集群上进行的. LSSC-IV 集群基于联想深腾 8810 系统构建, 包含超算和大数据计算两部分. 计算集群主体部分总共拥有 14688 个处理器核, 理论峰值性能为 1081 TFlops, 实测 LINPACK 性能 703 TFlops.

下面我们先介绍一下实际半导体器件数值模拟中常用到的几种边界条件 [22]. 在实际半导体器件模拟中, 常用到以下四种边界条件:

1. **欧姆接触:** 欧姆接触是 Dirichlet 型边界, 半导体材料上引出的金属电极都作为此类边界处理. 在边界上, 多子和少子的准费米势与作用在电极上的偏压相等, 即 $\phi_p = \phi_n = V_{app}$. 电势和载流子浓度在边界上的取值如下:

$$\begin{aligned} p &= \frac{1}{2} \left(N_A - N_D + \sqrt{(N_A - N_D)^2 + 4n_{ie}^2} \right), \\ n &= \frac{1}{2} \left(N_D - N_A + \sqrt{(N_D - N_A)^2 + 4n_{ie}^2} \right), \\ \psi &= \phi_p - \frac{k_B T}{q} \ln \left(\frac{p}{n_{ie}} \right) = \phi_n + \frac{k_B T}{q} \ln \left(\frac{n}{n_{ie}} \right). \end{aligned}$$

2. **氧化物 - 半导体界面:** 氧化物是绝缘层, 内部没有电流流过, 因此电流在界面上的外法向分量为零, 即

$$\mathbf{J}_p \cdot \mathbf{n} = \mathbf{J}_n \cdot \mathbf{n} = 0.$$

其中, \mathbf{n} 是界面上指向半导体材料外部的单位法向量. 电势则满足以下条件:

$$\epsilon_s \frac{\partial \psi}{\partial \mathbf{n}} - \epsilon_o \frac{\partial \psi}{\partial \mathbf{n}} = \sigma.$$

其中, ϵ_s 和 ϵ_o 分别为半导体和氧化物的介电常数, σ 为界面电荷密度.

3. **栅电极:** 栅电极由氧化层上引出, 电势在其上也作为 Dirichlet 型边界处理, 取值如下:

$$\psi = V_{app} - V_{ms}.$$

其中, V_{ms} 为金属与半导体的接触电势差. 由于氧化层内不存在载流子, 连续性方程在其上没有定义.

4. **自由边界:** 金属电极以外的边界均为自由边界, 电势和载流子电流满足如下条件:

$$\frac{\partial \psi}{\partial \mathbf{n}} = 0 \quad \text{以及} \quad \mathbf{J}_p \cdot \mathbf{n} = \mathbf{J}_n \cdot \mathbf{n} = 0$$

3.1. 几种有限元方法的比较

我们将通过对 PN 结的数值计算, 从升压范围、掺杂浓度和电流稳定性三个方面对上述几种有限元方法进行比较. 关于 PN 结的工作原理, 我们将在下一小节中详细介绍. 实验中使用的 PN 结的版图和纵剖面结构图如图 1 所示, P 区嵌入 N 型衬底中. P 区顶部与 N 区底部分别通过欧姆接触引入电极.

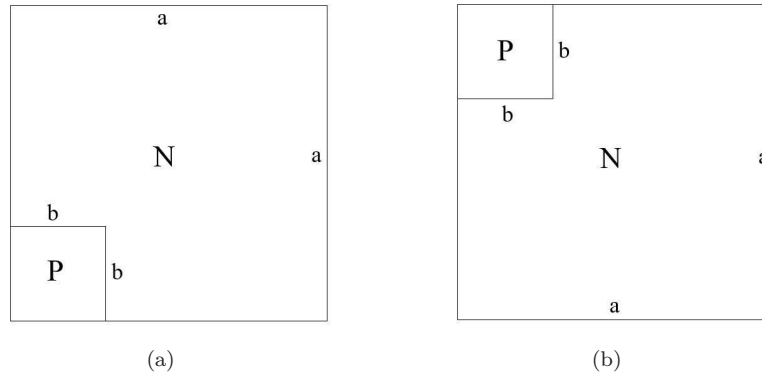


图 1 PN 结结构图: (a) 横切面版图; (b) 纵剖面结构示意图.

为了更好地测试几种有限元方法的稳定性, 我们在数值实验中采用一次性升压的方式, 即将电极上的偏压升到指定的伏数. 在第一组实验中, 我们测试不同偏压下的收敛情况. 取 $a = 0.1 \mu\text{m}$, $b = 0.03 \mu\text{m}$, P 区和 N 区的掺杂浓度均为 $5.0 \times 10^{18} \text{ cm}^{-3}$, 分别一次性将电极上的偏压升至 0.5 V、1 V、1.5 V 和 2 V. 在第二组实验中, 我们测试不同掺杂浓度下的收敛情况. 取 $a = 1 \mu\text{m}$, $b = 0.3 \mu\text{m}$, 偏压固定为 1 V, 仍采用一次性升压的方式. P 区和 N 区的掺杂浓度相同, 分别取为 $1.0 \times 10^{17} \text{ cm}^{-3}$ 、 $5.0 \times 10^{17} \text{ cm}^{-3}$ 和 $1.0 \times 10^{18} \text{ cm}^{-3}$. 最后, 我们观察两组实验中三种有限元方法计算出的电流. 根据流守恒定律, 阳极和阴极上的电流大小应该相等. 实验的结果如表 1 所示. "✓" 表示收敛或电流守恒, "×" 表示发散或电流不守恒.

表 1 三种有限元方法在不同条件下的收敛情况及电流稳定性

	偏压 (V)				掺杂浓度 (cm^{-3})			电流是否守恒
	0.5	1	1.5	2	1.0×10^{17}	5.0×10^{17}	1.0×10^{18}	
SUPG	✓	✓	✓	✓	✓	✓	✓	×
SUPG-IP	✓	✓	✓	✓	✓	✓	✓	×
Zlamal	✓	✓	✓	×	✓	✓	×	✓

数值实验表明, 两类稳定化有限元方法均能有效提高有限元方法的稳定性. SUPG 系列方法的优势在于其在较大的偏压或较高的掺杂浓度下更为稳健, 但由于人工扩散项的引入, 计算得到的数值解比真实解要光滑, 且数值结果会受到算法中稳定化参数选取的影响; Zlamal 有限元方法计算得到的电流更为稳定, 但 Zlamal 方法对网格要求更高, 一般要求为锐角剖分网格. 在锐角网格的条件下, Zlamal 有限元可以保证计算所得的浓度非负. 在接下来的数值实

验中,我们将模拟几种典型的半导体器件的电学特性,并与商业软件 Sentaurus 的结果进行比较.考虑到电流的稳定性,我们将主要采用 Zlamal 有限元方法,同时使用连续升压的方式进一步增强 Zlamal 方法的稳定性.

3.2. 典型器件模拟

在本小节中,我们将模拟几种典型的半导体器件^[23]的“电流-电压”特性曲线,并对比 DevSim 与商业软件 Sentaurus 的模拟结果. Sentaurus 是目前国际上领先的半导体器件模拟的商业软件,具备网格生成、器件模拟、后处理等多个功能,其中器件模拟模块主要采用的是有限体积法.

3.2.1. PN 结

我们模拟了一个宽 $20\mu\text{m}$ 、高 $10\mu\text{m}$ 、厚 $10\mu\text{m}$ 的 PN 结器件. P 区宽 $6\mu\text{m}$ 、深 $3\mu\text{m}$ 、厚 $3\mu\text{m}$, 嵌入 N 型衬底中. P 区顶端接阳极 (Anode), 电极为边长 $4\mu\text{m}$ 的正方形, N 区底部接阴极 (Cathode). P 区和 N 区的掺杂浓度均为 10^{17}cm^{-3} . 由对称性,我们选择 PN 结的四分之一区域进行计算. 其中,阳极和阴极作为欧姆接触边界处理,其余均为自由边界. 计算网格如图 2 所示,网格由 Sentaurus 生成,共包含 25066 个单元.

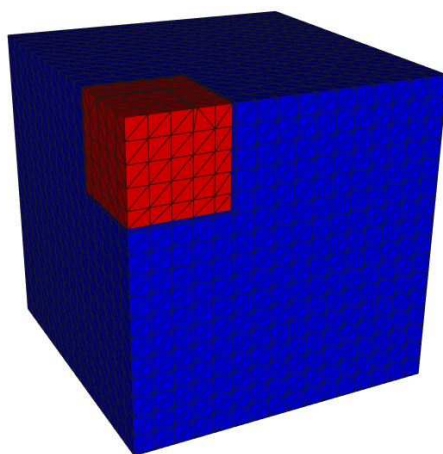


图 2 PN 结四面体网格图

PN 结是最典型的半导体器件,把一块 P 型半导体和一块 N 型半导体结合在一起,在两者的交界面处就形成了所谓的 PN 结. 由于它们之间存在载流子浓度梯度,导致空穴从 P 区扩散到 N 区、电子从 N 区扩散到 P 区,分别在 P 区和 N 区留下了带负电的电离受主和带正电的电离施主,形成空间电荷区. 这些电荷产生了从 N 区指向 P 区的电场,称为内建电场. 在内建电场的作用下,载流子做漂移运动,其运动方向与扩散运动相反. 在没有外加电压时,随着扩散运动的进行,内建电场逐渐增强,最终载流子的扩散运动和漂移运动将达到动态平衡,此时 PN 结内净电流为零.

当 PN 结施加正向偏压,即 P 区接电源正极, N 区接电源负极,则正向偏压将产生与内建电场方向相反的电场,削弱了内建电场的强度,扩散运动超过漂移运动,扩散电流占主导地位. 此时电子将源源不断地从 N 区扩散到 P 区,空穴从 P 区扩散到 N 区,成为非平衡载流子,

这种现象称为 PN 结的正向注入. 由于 N 区的电子和 P 区的空穴均为多子, 因此在正向偏压下, PN 结内的电流随电压增长很快. 我们分别用 DevSim 中的 Zlamal 有限元方法和商业软件 Sentaurus 对上述 PN 结进行了模拟计算, 图 3 为直角坐标和对数坐标下 PN 结的“电流 - 电压”特性曲线. 首先, 图中结果表明 Zlamal 有限元计算出的结果和 Sentaurus 的模拟结果非常贴近. 然后, 由“电流 - 电压”特性曲线可以看到 PN 结正向导通的特性, 启动电压约为 0.7 V, 启动前的电流与电压呈指数关系, 启动后电流与电压呈线性关系.

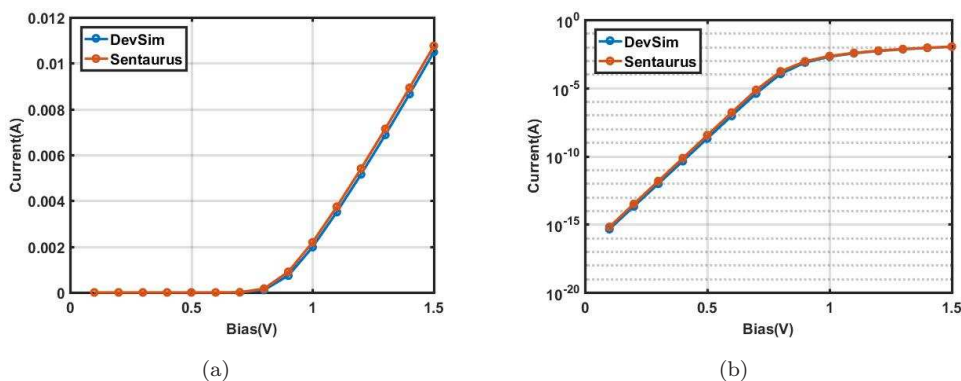


图 3 PN 结“电流 - 电压”特性曲线: (a) 直角坐标; (b) 对数坐标

3.2.2. MOSFET

MOSFET 是一种半导体场效应器件. 由于它工艺较简单, 易实现集成化, 因此在大规模集成电路中得到了广泛的应用. MOSFET 的版图和纵剖面结构图如图 4 所示, 其中版图里的阴影部分为金属电极. 图中为 N 型 MOSFET 器件, 在 P 型衬底上, 有两个 N^+ 型扩散区, 分别叫源 (Source) 和漏 (Drain). 源和漏之间的 P 型硅上有一层二氧化硅层, 叫栅氧化层. 氧化层上有一层金属, 也可用多晶硅代替, 称为栅 (Gate). 当栅极外加电压为 0 时, 源区和漏区被中间的 P 型硅隔开, 相当于两个背靠背的 PN 结, 即使在源和漏之间加一定的电压, 也没有明显的电流. 但当栅极加适当的正电压 $V_G > 0$ 时, 在 P 型硅的表面有可能形成反型层 (即 N 型层). 反型层将源区和漏区沟通, 形成 N 型沟道. 此时再在源和漏之间加一定的电压, 就会有

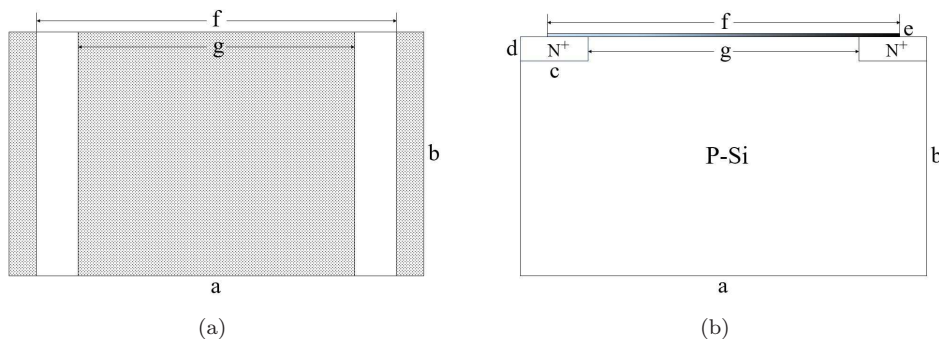


图 4 MOSFET 结构图: (a) 横切面版图; (b) 纵剖面结构图.

明显的电流流过. 当漏极上所加电压 V_D 较小时, 漏极电流 I_D 与漏极电压成正比, 称为线性区. 随着漏极电压的不断增大, 漏极电流和漏极电压的关系与线性关系的偏离也越来越大. 当漏极电压达到某一阈值时, 沟道在漏极附近被夹断, 漏极电流不再随电压增加, 称为饱和区.

数值模拟的器件尺寸如下: $a = 3 \mu\text{m}$, $b = 2 \mu\text{m}$, $c = 0.5 \mu\text{m}$, $d = 0.2 \mu\text{m}$, $f = 2.4 \mu\text{m}$, 氧化层厚度为 $e = 0.025 \mu\text{m}$, 沟道长度为 $g = 2 \mu\text{m}$. 源区和漏区掺杂浓度均为 10^{18}cm^{-3} , 衬底掺杂浓度为 10^{16}cm^{-3} . 源极和漏极作为欧姆接触边界处理, P 型衬底和氧化层之间的界面则采用氧化物 - 半导体界面的条件, 栅极上的电势采用栅电极边界条件, 其余均为自由边界. 计算网格如图 5 所示, 网格由 Sentaurus 生成, 共包含 163961 个单元.

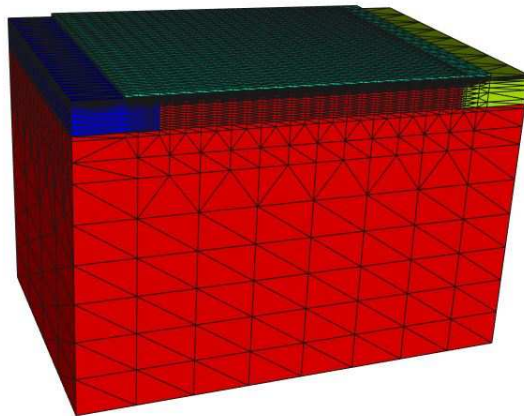


图 5 MOSFET 四面体网格图

我们固定漏极上的电压为 $V_D = 1 \text{V}$, 通过计算漏极电流 I_D 随栅极电压 V_G 的变化可以推导出器件的开启阈值电压. 如图 6(a) 所示, 当栅极电压 $V_G \leq 1 \text{V}$, 由于源区和漏区被中间的 P 型硅隔开, 此时几乎没有电流流过. 当 V_G 大于 1V 后, 漏极开始有明显的电流流过, 且当 V_G 不断增大, I_D 也随之增大. 我们固定栅极电压为 $V_G = 4 \text{V}$, 观察漏极电流 I_D 随漏极电压 V_D 的变化, 图 6(b) 为计算出的漏极“电流 - 电压”特性曲线. 随着漏极电压的增大, 漏极电流也由线性增长逐渐趋于饱和, 与理论结果是吻合的. 图 7 中展示了 V_G 分别为 0.5V 和 4V 时 MOSFET 的一个纵截面上的电子浓度分布, 可以看到 $V_G = 4 \text{V}$ 时在 P 型硅的表面形成了反型层, 将源区和漏区沟通, 形成了 N 型沟道.

3.2.3. BJT 数值模拟结果

BJT (Bipolar Junction Transistor) 常称为双极性结型晶体管, 简称晶体管或三极管. 三极管是由两个相邻的 PN 结组成, 按类型分为两种: NPN 型和 PNP 型, 两种类型的 BJT 具有几乎等同的特性. 下面以 PNP 型晶体管为例, 介绍 BJT 的工作原理. 图 8 为 PNP 型晶体管的版图和纵剖面结构图. 它的基本结构是两个 P 型层中间夹着一个很薄的 N 型层, 这个 N 型层叫基区, 两个 P 型层分别叫发射区和集电区, 形成两个相邻的 PN 结, 分别叫发射结和集电结. 在三个区域分别引出三根电极引线, 称为发射极 (Emitter)、基极 (Base) 和集电极 (Collector).

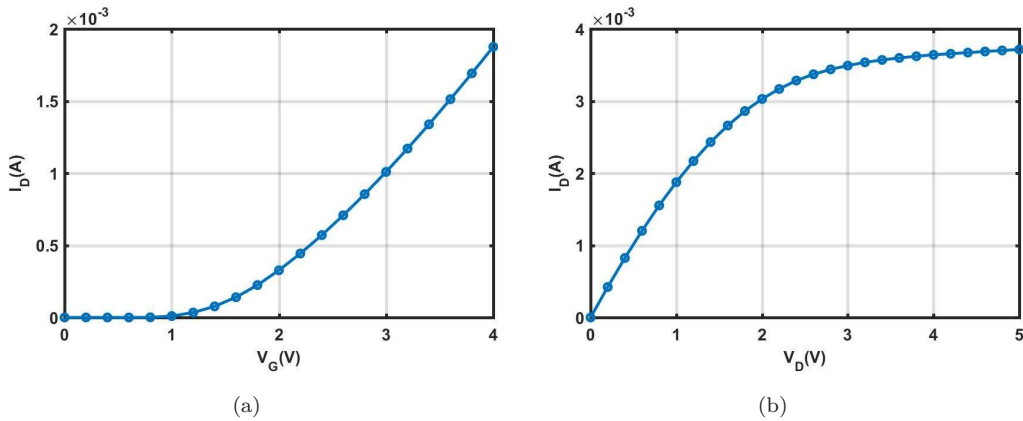


图 6 N-MOSFET 电学响应特性曲线: (a) 源漏偏压固定为 $V_D = 1$ V, 漏极电流随栅极电压变化曲线 $I_D - V_G$; (b) 栅压固定为 4 V, 漏极电流随源漏偏压变化曲线 $I_D - V_D$.

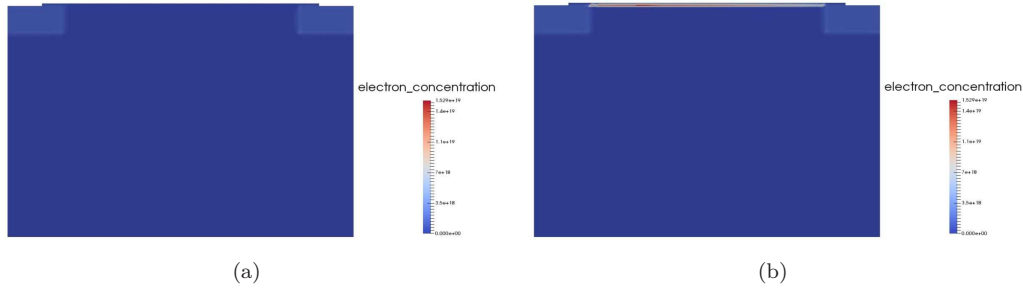


图 7 N-MOSFET 纵截面电子浓度分布: (a) $V_G = 0.5$ V; (b) $V_G = 4$ V

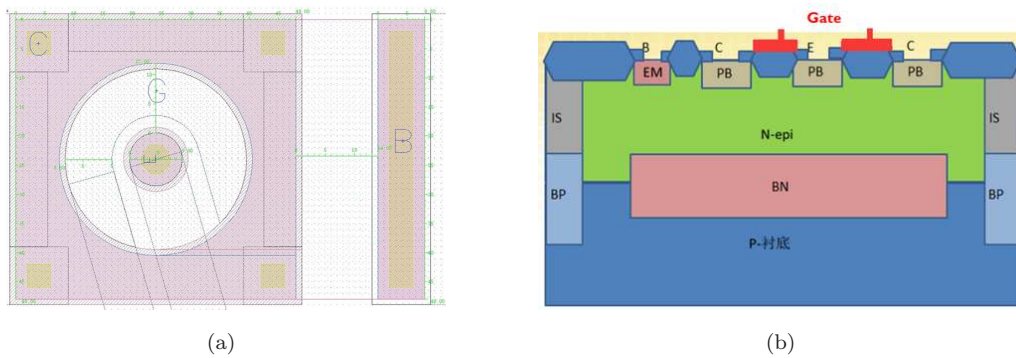


图 8 BJT 晶体管结构图: (a) 横切面版图; (b) 纵剖面结构图.

通常发射极接地, 在基极和发射极之间加一较小的电压 $V_{BE} (< 0)$, 使发射结处于正向. 在集电极和发射极之间加一较大的电压 $V_{CE} (< 0)$, 由于集电极的电势比基极低, 此时集电结处于反向. 当电路接通时, 由于发射结处于正向, 发射区的空穴注入基区, 而基区的电子也会注入发射区. 在扩散过程中, 空穴会与基区中的电子复合, 形成基极电流 I_B . 为了减少复合, 常把基区做得很薄, 且使基区掺杂的浓度较低, 因而注入到基区的空穴大部分会被集电结的抽

取作用拉到集电极. 此时集电结虽然处于反向, 却流过很大的电流. 改变基极电流 I_B 的大小, 会发现集电极电流 I_C 将按照 I_B 的很大的倍数变化, 这就是三极管的电流放大作用.

我们模拟该横向 PNP (LPNP) 晶体管的输入特性曲线. 输入特性是指当集电极和发射极之间的电压 V_{CE} 为常数时, 基极电流 I_B 与加在基极和发射极之间的电压 V_{BE} 之间的关系曲线, 用函数关系表示为

$$I_B = f(V_{BE}) \Big|_{V_{CE}=\text{const}}$$

发射区的掺杂浓度为 $5.0 \times 10^{15} \text{ cm}^{-3}$, 集电区的掺杂浓度为 $1.0 \times 10^{15} \text{ cm}^{-3}$, 基区的掺杂浓度为 $5.0 \times 10^{13} \text{ cm}^{-3}$. 发射极、基极和集电极作为欧姆接触边界处理, 其余为自由边界. 计算网格如图 9 所示, 网格由 Sentaurus 生成, 共包含 74140 个单元.

LPNP 晶体管的输入特性曲线如图 10 所示, 当 V_{CE} 反向增大时, 由于集电结处于反向, 此时集电结的反向抽取能力增强, 使得从发射区注入基区的空穴更多地流向集电区, 因此对于相同的 V_{BE} , 基极电流 I_B 相应地减小, 故而特性曲线右移.

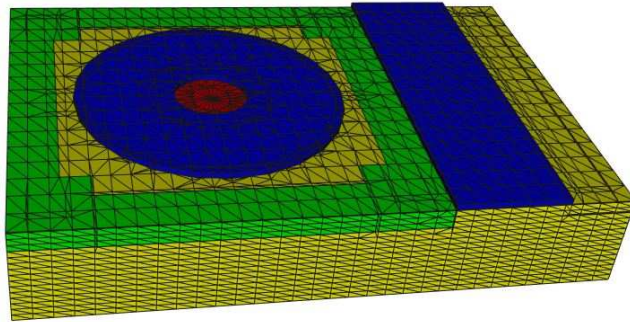


图 9 LPNP 晶体管四面体网格图

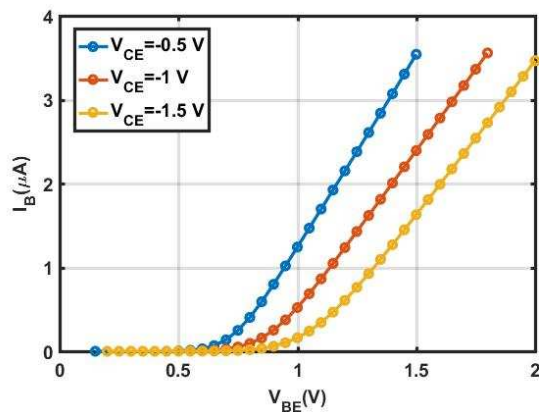


图 10 LPNP 晶体管输入特性曲线

3.3. 大规模并行测试

PHG 软件平台中提供了并行的三维四面体网格的管理, 支持大规模并行并且具有良好的可扩展性. 这一小节我们将以 PN 结为例, 对算法的并行性能进行测试. 本次模拟采用的 PN 结结构与 3.2 小节相同, P 区和 N 区的掺杂浓度均为 10^{16} cm^{-3} . 在 PHG 中对 PN 结的四面体网格进行一致加密, 加密后的网格共包含 821,362,688 个网格单元, 137,701,745 个顶点.

设置偏压为 0.05 V, 我们测试不同规模的网格在不同进程数下的 Gummel 迭代步数和求解一次 Gummel 迭代平均所用的时间. 数值结果表明, 不同规模的网格在不同进程数下的 Gummel 迭代步数非常稳定, 均为 3 步, 求解一次 Gummel 迭代平均所用的时间及并行效率如表 2 所示. 其中, 第一列和第二列的并行效率以 32 进程的运行时间为基准; 第三列中, 由于进程数为 32、64 或 128 时计算内存不够, 并行效率以 256 进程的运行时间为基准. 从表中可以看到, 在第一列中, 2048 进程的运行时间反而比 1024 进程的要长, 这是因为问题的计算规模不够大, 而 2048 进程的通讯复杂度增加, 此时扩大并行规模所减少的计算时间无法覆盖增加的通讯开销, 导致运行时间增加; 相比之下, 第二、三列问题规模扩大, 使用 2048 进程的运行时间相比于 1024 进程也相应地在减少. 尤其是第三列的运行时间, 与 1024 进程时相比减少了将近一半, 展示出算法良好的并行可扩展性.

表 2 PN 结数值模拟并行可扩展性测试

单元数 进程数	12,833,792		102,670,336		821,362,688	
	时间 (s)	并行效率	时间 (s)	并行效率	时间 (s)	并行效率
32	212.0	—	2046.8	—	—	—
64	103.8	102.2%	1085.3	94.3%	—	—
128	56.7	93.4%	495.5	103.3%	—	—
256	32.9	80.6%	256.8	99.6%	2248.5	—
512	21.7	61.1%	133.3	96.0%	1118.8	100.5%
1024	18.9	35.1%	81.4	78.5%	606.8	92.6%
2048	23.6	14.0%	67.5	47.4%	334.5	84.0%

4. 总 结

本文基于漂移扩散模型, 采用有限元离散, 并引入了几种稳定化的技巧, 对几种典型的半导体器件的电学特性进行了三维数值模拟. 计算结果与商业软件 Sentaurus 一致, 验证了算法的有效性. 对 PN 结的大规模数值模拟实验展示了算法良好的并行可扩展性. 未来的研究工作包括基于漂移扩散模型对更复杂器件的模拟、发展更加高效稳定的数值方法以及对其他计算模型如流体力学模型、量子修正模型等的数值求解.

参 考 文 献

- [1] <https://www.synopsys.com/silicon/tcad/device-simulation/sentaurus-device.html>. Accessed March 7, 2020.
- [2] https://www.silvaco.com/products/tcad/device_simulation/atlas/atlas.html. Accessed March 7, 2020.

- [3] Vanroosbroeck W. Theory of the flow of electrons and holes in germanium and other semiconductors [J]. *Bell System Technical Journal*, 1950, 29(4): 560–607.
- [4] 马召灿, 许竞劫, 卢本卓, 李鸿亮. 半导体器件辐照损伤效应模拟的数值算法及应用 [J]. *数值计算与计算机应用*, 2020, 41(2): 105–120.
- [5] Gummel H. Self-consistent iterative scheme for one-dimensional steady state transistor calculations [J]. *IEEE Transactions on Electron Devices*, 1964, 11(10): 455–465.
- [6] 成杰, 张林波. 一种可扩展的三维半导体器件并行数值模拟算法 [J]. *计算物理*, 2012, 29(3): 439–448.
- [7] Scharfetter D, Gummel H. Large-signal analysis of a silicon read diode oscillator [J]. *IEEE Transactions on Electron Devices*, 1969, 16(1): 64–77.
- [8] 许竞劫, 谢妍, 卢本卓. 一个基于 PHG 平台的并行有限元生物分子模拟解法器 [J]. *数值计算与计算机应用*, 2016, 37(1): 67–82.
- [9] Chen M, Lu B. TSMesh: A robust method for molecular surface mesh generation using a trace technique [J]. *Journal of Chemical Theory and Computation*, 2010, 7(1): 203–212.
- [10] Liu T, Bai S, Tu B, Chen M, Lu B. Membrane-channel protein System mesh construction for finite element simulations [J]. *Molecular Based Mathematical Biology*, 2015, 3(1): 128–139.
- [11] Liu T, Chen M, Lu B. Efficient and qualified mesh generation for Gaussian molecular surface using adaptive partition and piecewise polynomial approximation [J]. *SIAM Journal on Scientific Computing*, 2018, 40(2): B507–B527.
- [12] Gui S, Khan D, Wang Q, Yan D, Lu B. Frontiers in biomolecular mesh generation and molecular visualization systems [J]. *Visual Computing for Industry, Biomedicine, and Art*, 2018, 1:7.
- [13] Bank R, Burgler J, Fichtner W, Smith R. Some upwinding techniques for finite-element approximations of convection-diffusion equations [J]. *Numerische Mathematik*, 1990, 58(2): 185–202.
- [14] Miller J, Schilders W, Wang S. Application of finite element methods to the simulation of semiconductor devices [J]. *Reports on Progress in Physics*, 1999, 62(3): 277–353.
- [15] Brooks A N, Hughes T J R. Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations [J]. *Computer Methods in Applied Mechanics and Engineering*, 1982, 32(1-3): 199–259.
- [16] Douglas J, Dupont T. Interior Penalty Procedures for Elliptic and Parabolic Galerkin Methods[M], pages 207–216. Springer Berlin Heidelberg, Berlin, Heidelberg, 1976.
- [17] Zlamal M. Finite element solution of the fundamental equations of semiconductor devices. I [J]. *Mathematics of Computation*, 1986, 46(173): 27–43.
- [18] Zhang L. A Parallel Algorithm for Adaptive Local Refinement of Tetrahedral Meshes Using Bisection [J]. *Numerical Mathematics-Theory Methods and Applications*, Feb 2009, 2(1): 65–89.
- [19] 黄成梓, 白石阳, 王芹, 马召灿, 张倩茹, 刘田田, 桂升, 卢本卓, 陈旻昕, 李鸿亮. 3Ddevice: 半导体器件及其辐照损伤效应仿真软件系统 [J]. *数值计算与计算机应用*, 2020, 41(2): 121–142.
- [20] Burman E, Hansbo P. Edge stabilization for Galerkin approximations of convection-diffusion-reaction problems [J]. *Computer Methods in Applied Mechanics and Engineering*, 2004, 193(15–16): 1437–1453.
- [21] Duan H, Wei Y. A New Edge Stabilization Method for the Convection-Dominated Diffusion-Convection Equations[M], pages 48–60. Springer International Publishing, Cham, 2018.
- [22] 成杰. 基于 PHG 平台的半导体器件模拟研究与结构力学有限元软件研制 [D]. PhD thesis, 中国科学院研究生院, 2012.
- [23] 黄昆, 韩汝琦. 半导体物理基础 [M]. 科学出版社, 2010.