

场效应晶体管：用电场开关电流

第 43 届决赛短训讲义组

2026 年 10 月

目录

1 导读	2
2 名词与符号	6
3 前置模块 I：能带、电子与空穴	6
4 前置模块 II：载流子统计	9
5 前置模块 III：漂移、扩散与半导体静电学	13
6 MOS 电容：栅电场如何弯曲能带	16
7 MOSFET 的电流	25
8 亚阈值区与玻尔兹曼极限	34
9 CMOS 与器件缩小	37
10 决赛风格例题	43
11 实验与数据处理：从电学测量提取 MOSFET 的参数	52
12 小结：推导链	56
A 计算细节与插值模型	60
B 泊松方程第一积分的推广与弱反型电荷	62
C 自然长度：改进的抛物线近似与精确本征方程	63
D 最小二乘直线拟合	64

1 导读

一块手机芯片上有一百多亿个晶体管，每一个都是一只“电控开关”：在第三个电极（栅极）上加一个电压，就能让另外两个电极（源极与漏极）之间的电流导通或关断。今天几乎所有芯片里的晶体管都是金属-氧化物-半导体场效应晶体管（MOSFET）。它的工作原理用一句话就能说清：把平行板电容器的一块极板换成半导体，给电容器充电，半导体表面就感应出一层可以导电的电荷，于是表面电导受栅压控制。可是要把这句话变成可以计算的物理，需要回答一连串问题：半导体里的电荷是什么？电场能伸进半导体多深？为什么栅压超过某个“阈值”之后表面突然出现大量电子？这些电子在源漏电压下怎样流动？为什么电流在漏压增大到一定程度后不再增加？为什么晶体管不能被“完全关断”，而这又如何限制了芯片的功耗？

本讲沿着“栅极电场 → 表面能带弯曲 → 反型沟道 → 电流”这条主线回答这些问题。我们先用普通物理（静电学、玻尔兹曼因子、漂移与扩散）补齐半导体物理最少够用的部分，然后由一维泊松方程推出 MOS 电容的全部静电学，再推出 MOSFET 的电流-电压关系、亚阈值区的玻尔兹曼极限，最后讨论 CMOS 电路与器件缩小。

背景：一个早到的想法：Lilienfeld 的专利

1925 年 10 月，物理学家 J. E. Lilienfeld 在加拿大申请了用电场控制固体中电流的专利，1926 年 10 月 8 日又在美国申请，美国专利 1,745,175 号于 1930 年 1 月 28 日获批 [1, 2]。图 1 是这件专利的附图：玻璃基片上镀两块金属电极，其上覆盖一层铜硫化物薄膜；两电极之间的玻璃中嵌入一片铝箔作为控制电极，铝箔的上缘与玻璃表面齐平，与薄膜构成单向导电的接触。专利设想：在铝箔上加电压，其静电作用改变薄膜的导电性 [1]。1934 年 3 月 2 日，当时在英国剑桥工作的 O. Heil 在德国首次申请了类似的专利，1935 年 3 月 4 日在英国申请（据专利原件；CHM 作 3 月 5 日），完整说明书于同年 12 月 6 日被接受（英国专利 439,457 号）[2, 3]。但没有证据表明他们真正做出了能工作的器件 [2]。

背景：失败的场效应实验、表面态与晶体管的诞生

1945 年 4 月，贝尔实验室的 W. Shockley 构想了一种场效应放大器 [4]：把一层硅或锗薄膜做成平行板电容器的一块极板，电容器充电后感应出的电荷如果是可移动的载流子，薄膜的电导就会明显改变 [5]。然而用蒸镀薄膜做的实验都只得到可以忽略的效应。J. Bardeen 后来在诺贝尔演讲中回忆，在有的实验中，理论预期的效应比实验的探测下限还大一千多倍；分析表明，其中约 50–100 倍的差距来自薄膜中电子迁移率远低于体材料，其余部分归因于表面态的屏蔽 [6]。Bardeen 在 1947 年的论文中提出：半导体表面存在束缚电子的局域能级（表面态），感应电荷的大部分被这些能级“俘获”而不能移动 [7]。

为了研究表面态，Bardeen 与 W. Brattain 用电解液在硅、锗表面施加强电场，观察表面反型层的电导受电场调制。这一系列实验最终导向了 1947 年 12 月 16 日首次成功的点接触晶体管（12 月 23 日向实验室管理层演示）[4, 6, 8]。1948 年，Shockley 与 G. L. Pearson 首次发表了正面的场效应实验结果 [9]。1952 年 Shockley 提出了用 pn 结控制沟道的“单极型场效应晶体管”理论（即结型场效应管 JFET）[10]。1956 年诺贝尔物理学奖授予 Shockley、Bardeen 与 Brattain[5]。本讲第 6.7 节会定量说明表面态为什么能让场效应“消失”。

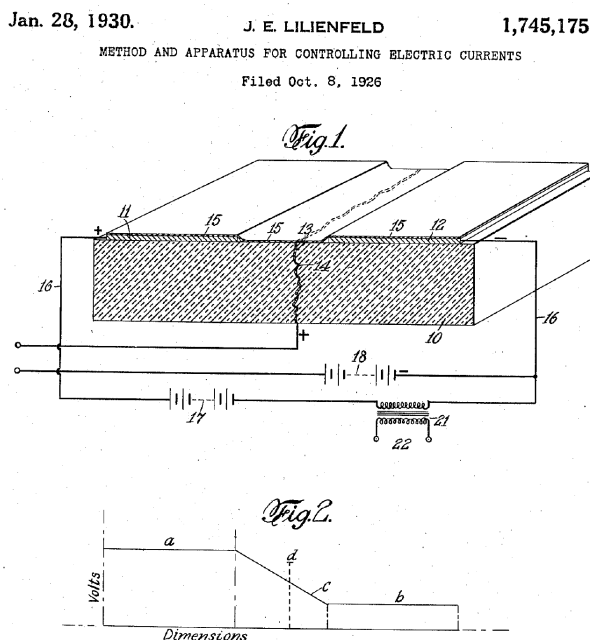


图 1 Lilienfeld 美国专利 1,745,175 号 (1926 年申请, 1930 年授权) 的附图 1 与图 2。图 1 中 10 为玻璃基片, 11、12 为薄膜两端的金属电极, 13 为嵌在玻璃裂缝 14 中的铝箔控制电极, 15 为铜硫化物薄膜; 图 2 示意沿薄膜的电压分布。(图源: J. E. Lilienfeld, US Patent 1,745,175 (1930), Figs. 1-2, 经裁剪; 原件 <https://patents.google.com/patent/US1745175A/en>; 美国专利的文本与附图不受版权限制 (37 CFR 1.71), 公有领域)

背景：热氧化硅与第一只 MOSFET

转机来自界面。贝尔实验室的 M. M. Atalla 等人发现, 在硅表面热生长一层二氧化硅可以使表面电学性质稳定下来 (1959 年发表) [11]: Si/SiO₂ 界面上的电子态远少于裸露的硅表面。1959 年, Atalla 与 D. Kahng 做出了第一只成功的绝缘栅场效应晶体管, 1960 年 6 月在匹兹堡的固态器件研究会议上报告 [12, 13]。图 2 是 Kahng 1960 年 5 月 31 日申请、1963 年 8 月 27 日授权的美国专利 3,102,230 号的附图 [14]: n 型硅片上扩散两个 p 型区 (13、14), 其间的表面覆盖约 100 nm (1000 埃) 厚的热氧化层 (19), 氧化层上的铝电极 (21) 就是栅。这其实是一只 p 沟道 MOSFET。1963 年, 仙童公司的 F. Wanlass 与 C. T. Sah 在国际固态电路会议上报告了把 n 沟道与 p 沟道 MOSFET 互补连接的“纳瓦逻辑”, 即今天的 CMOS, 静态时几乎不耗电 (Wanlass 的专利 3,356,858 号于 1963 年 6 月 18 日申请, 1967 年授权) [15, 16]。

背景：缩小：从摩尔到 Dennard, 再到 FinFET 与全环绕栅

1965 年, G. Moore 在《Electronics》杂志上指出, 集成电路上的元件数大约每年翻一番, 并预计 1975 年可达 65000 个; 1975 年他把速度修正为约每两年翻一番 [17-19]。1974 年, IBM 的 R. H. Dennard 等人给出了 MOSFET 的等比例缩小规则: 所有尺寸、电压缩小为原来的 $1/\kappa$, 掺杂浓度增大为 κ 倍, 则器件的电场分布不变, 速度提高、功耗降低, 而单位面积功耗不变 [20]。他们在同一篇论文中已经指出: 亚阈值区的特性“不能按比例缩小” [20]。三十年后这句话成了现实: 约 2004 年起, 电源电压不能再随尺寸降低, 处理器时钟频率与单核性能的增长基本停滞 [21, 22]。此后器件的改进靠的是新材料 (2007 年量产的高介电常数栅介质与金属栅

Aug. 27, 1963

DAWON KAHNG

3,102,230

ELECTRIC FIELD CONTROLLED SEMICONDUCTOR DEVICE

Filed May 31, 1960

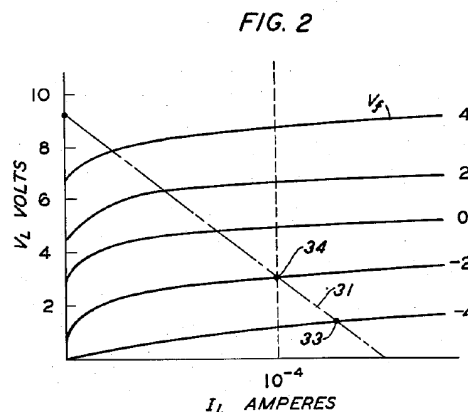
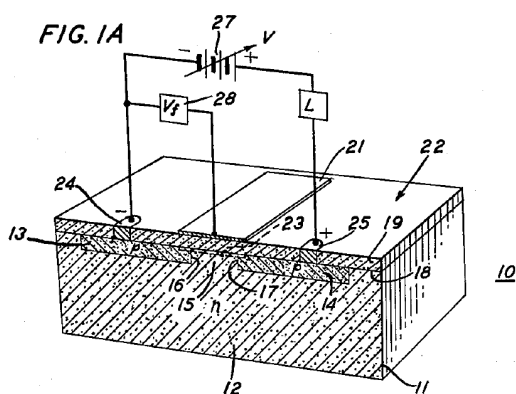


图2 Kahng 美国专利 3,102,230 号 (1960 年申请, 1963 年授权) 的附图 1A 与图 2。图 1A: n 型硅片 (12) 中的两个 p 型区 (13、14) 作源与漏, 热氧化层 (19) 上的电极 (21) 为栅, 电池 27 经负载 L 加在触点 24、25 之间, 使 13 与 n 型衬底之间的 pn 结 (第 7.1 节) 反偏 (13 一侧为漏), 28 为栅压源 V_f 。图 2: 不同 V_f (-4 到 +4 V) 下的典型特性, 纵轴 V_L 是器件两端的电压, 横轴是电流 I_L ; 虚线 31 是负载线。注意 V_f 接在栅与触点 24 (漏) 之间, 固定的是栅-漏电压, 而不是今天习惯的栅-源电压。专利原文指出, 每条曲线都有一段电压对电流不敏感, 因而器件可作可调的稳压器。用第 7 节的结果可以理解这一点 (以 n 沟道写法, p 沟道各量取绝对值)。只要 V_{GD} 不足以使漏端反型 ($V_{GD} < V_T$), 漏端就处于夹断, 器件饱和, $I_D = \frac{k}{2}(V_{GS} - V_T)^2$; 又 $V_{GS} = V_{GD} + V_{DS}$, 于是 $V_{DS} = V_T - V_{GD} + \sqrt{2I_D/k}$ 。图中 $V_f = 4, 2, 0, -2$ V 四条曲线在 $I_L = 0$ 处的截距约为 6.7、4.4、2.9、0.8 V, 随 V_f 大体等间隔移动, 正是 $V_T - V_{GD}$, 由此可读出 $|V_T| \approx 3$ V; $V_f = -4$ V 的曲线从原点出发, 对应 V_{GD} 已越过阈值、漏端也已反型的线性区。饱和段电压随电流缓慢上升, 斜率 $dV_{DS}/dI_D = 1/g_m = 1/\sqrt{2kI_D}$; 饱和只是前提 (电流只由 V_{GS} 决定), 平缓来自电流对 V_{GS} 的平方律。不要把图中的水平段误读为电流饱和。 (图源: D. Kahng, US Patent 3,102,230 (1963), Figs. 1A, 2, 经裁剪拼接; 原件 <https://patents.google.com/patent/US3102230A/en>; 美国专利的文本与附图不受版权限制 (37 CFR 1.71), 公有领域)

[23]) 和新几何：2011 年英特尔宣布 22 nm 工艺采用三面包围沟道的“三栅”(FinFET)晶体管，2011 年底起量产 [24]；2022 年 6 月三星宣布开始量产采用全环绕栅(GAA)纳米片晶体管的 3 nm 工艺 [25]。本讲第 8、9 节会推出这些变化背后的物理：亚阈值摆幅的玻尔兹曼极限与“自然长度”。

学习目标 学完本讲，你应当能够：

1. 说明能带、电子与空穴、有效质量、费米能级的含义；由玻尔兹曼因子推出费米-狄拉克分布，由态密度推出 $n = N_c e^{-(E_c - E_F)/k_B T}$ 与质量作用定律 $np = n_i^2$ ；由零电流条件推出爱因斯坦关系；
2. 画出 MOS 结构的能带图，解一维泊松方程得到表面电荷与表面势的精确关系，理解积累、耗尽、反型三个区，推出耗尽层宽度、强反型判据 $\psi_s = 2\phi_F$ 与阈值电压公式，并知道它们的误差有多大；
3. 由渐变沟道近似推出 MOSFET 线性区与饱和区的电流公式，说明夹断后电流为何不再增加，处理体效应、沟道长度调制与速度饱和；
4. 推出亚阈值摆幅 $S = (k_B T/q) \ln 10 (1 + C_d/C_{ox}) \geq 60 \text{ mV/dec}$ ，并用它解释“关断漏电与电源电压不能同时降低”；
5. 分析 CMOS 反相器的传输特性、噪声容限与动态功耗 $\alpha CV^2 f$ ；推出 Dennard 缩小规则及其失效原因，用自然长度说明 FinFET 与全环绕栅为何改善栅控；
6. 处理 MOSFET 转移特性、亚阈值特性与输出特性的实测数据：提取阈值电压、迁移率、串联电阻、亚阈值摆幅、界面陷阱密度与沟道长度调制系数。

前置知识：本讲需要的普通物理与数学

- 电磁学：高斯定理、电势、平行板电容器、电介质与电位移 $\mathbf{D} = \epsilon \mathbf{E}$ 、导体静电平衡；
- 热学：玻尔兹曼因子 $e^{-E/k_B T}$ 、熵 $S = k_B \ln \Omega$ 、温度的统计定义；
- 近代物理：德布罗意关系 $p = \hbar k$ 、 $E = \hbar \omega$ 、一维势阱中的能级、群速度、隧穿概率 $\propto e^{-2\kappa d}$ ；
- 数学：泰勒展开、定积分、一阶与二阶常微分方程、对数坐标、最小二乘直线拟合。

本讲新引入的概念（能带、空穴、有效质量、费米-狄拉克分布、化学势与费米能级、态密度、掺杂、迁移率、扩散系数、准费米能级、德拜长度、功函数、耗尽近似、表面势、阈值电压、界面陷阱、渐变沟道近似、速度饱和、亚阈值摆幅、噪声容限、自然长度等）都在首次使用前给出定义与推导。

建议学时 共 2 学时（约 90 分钟）：半导体物理前置 25 分钟（以课前自学为主，课上只串讲结论与推导思路）；MOS 电容 25 分钟；MOSFET 的电流与亚阈值区 25 分钟；CMOS 与缩小 15 分钟。例题、实验题与附录供课后阅读与练习。

推导路线

1. 能带与空穴 $\Rightarrow$ (玻尔兹曼因子) 费米-狄拉克分布 $\Rightarrow$ (态密度) n 、 p 与 $np = n_i^2 \Rightarrow$ (掺杂) 费米势 ϕ_F ;
2. 漂移 + 扩散 $\Rightarrow$ (平衡态零电流) 爱因斯坦关系 $\Rightarrow$ 准费米能级; 高斯定理 $\Rightarrow$ 泊松方程;
3. MOS 电容: 栅压分配 $V_G = V_{FB} + \psi_s - Q_s/C_{ox} \Rightarrow$ 泊松方程第一积分 $Q_s(\psi_s) \Rightarrow$ 耗尽近似、强反型、阈值电压、 $C-V$;
4. MOSFET: $I_D = -W\mu Q_n dV/dy \Rightarrow$ (渐变沟道近似) 平方律、夹断、体效应 $\Rightarrow$ 速度饱和; 弱反型 $\Rightarrow$ 扩散电流与亚阈值摆幅;
5. CMOS 反相器 $\Rightarrow$ 动态功耗; 泊松方程的标度不变性 $\Rightarrow$ Dennard 缩小 $\Rightarrow$ 其失效; 二维泊松方程 $\Rightarrow$ 自然长度 $\Rightarrow$ FinFET、GAA。

2 名词与符号

以下字母在个别地方另作他用, 用到时会再说明: S (亚阈值摆幅; 第 4.1 节中为熵); m (体因子; m^* 、 m_0 、 m_c 、 m_v 、 m_h 、 m_{ox} 为各种质量); κ (缩小因子; κ_t 为隧穿衰减常数); Δ (强反型时 $\psi_s - 2\phi_F$; ΔL 为夹断区长度); k (波矢; 第 9 节中 k_n 、 k_p 为器件的增益因子); τ (散射时间; 陷阱响应时间); θ (迁移率随栅压退化的系数); β (实验题中的 $\mu_0 C_{ox} W/L$); f (费米-狄拉克分布 $f(E)$; 时钟频率)。

两点约定:

- **符号的正方向:** x 轴从 Si/SiO₂ 界面指向衬底内部; 静电势 ψ 以衬底体内为零点。 $\psi > 0$ 表示表面电势比体内高, 电子在表面的势能 $-q\psi$ 比体内低, 能带“向下弯”。除第 7 节讨论 p 沟道器件的少数地方, 全讲以 p 型衬底上的 n 沟道器件为例。
- **单位:** 半导体文献习惯用 cm 为长度单位 (浓度用 cm^{-3} , 迁移率用 $\text{cm}^2/(\text{Vs})$)。本讲给数值时沿用这一习惯, 代入公式计算时请统一换成国际单位制 ($1 \text{ cm}^{-3} = 10^6 \text{ m}^{-3}$)。

3 前置模块 I: 能带、电子与空穴

3.1 从原子能级到能带

先回忆一个熟悉的例子: 两个相同的单摆用弹簧弱耦合, 原来相同的频率分裂成两个简正频率 (同相与反相)。量子力学里两个相同的势阱也是这样: 一个电子在单个势阱中有能级 E_0 ; 两个势阱靠近时, 电子可以隧穿到另一个阱, 能级分裂成两个, 一个略低 (波函数在两阱中同号), 一个略高 (异号), 分裂的大小由隧穿的强弱决定。 N 个相同的原子排成晶体, 每个原子能级就分裂成 N 个能级。一块 1 立方厘米的硅有约 5×10^{22} 个原子, 这 N 个能级挤在几电子伏的范围内, 间隔小到可以看成连续, 称为能带。不同原子能级形成的能带之间可能留下没有能级的能量区间, 称为禁带。

表 1 本讲符号表 (材料常数取 300 K 硅的数值 [26])

符号	含义	单位或数值
q	元电荷 (电子电荷为 $-q$)	1.602×10^{-19} C
$U_T = k_B T / q$	热电压	300 K 时 25.85 mV
E_c, E_v, E_g	导带底、价带顶能量, 禁带宽度	Si: $E_g = 1.12$ eV
E_F, E_i	费米能级, 本征能级 (约在禁带中央)	eV
E_{Fn}, E_{Fp}	电子、空穴的准费米能级	eV
m^*	有效质量	kg
N_c, N_v	导带、价带有效态密度	Si: $3.2, 1.8 \times 10^{19}$ cm $^{-3}$
n, p	电子、空穴浓度	cm $^{-3}$
n_i	本征载流子浓度	Si: 约 1.0×10^{10} cm $^{-3}$
N_A, N_D	受主、施主浓度	cm $^{-3}$
$\phi_F = U_T \ln(N_A/n_i)$	费米势 (p 型衬底)	V
μ, μ_{eff}	迁移率, 反型层有效迁移率	cm ^{2} /(V s)
D	扩散系数	cm ^{2} /s
$\mathcal{E}$	电场强度 (为与能量 E 区分, 用花体)	V/m
$\psi(x), \psi_s$	半导体中的静电势 (以体内为零), 表面势	V
$\epsilon_s, \epsilon_{\text{ox}}$	硅、二氧化硅的介电常数	$11.7\epsilon_0, 3.9\epsilon_0$
L_D	德拜长度	m
$t_{\text{ox}}, C_{\text{ox}} = \epsilon_{\text{ox}}/t_{\text{ox}}$	栅氧化层厚度, 单位面积栅氧电容	m, F/m ^{2}
$\phi_m, \chi, \phi_{\text{ms}}$	金属功函数, 电子亲和能, 功函数差	V (或 eV)
V_G, V_{FB}, V_T	栅压, 平带电压, 阈值电压	V
Q_s, Q_d, Q_n	半导体中单位面积总电荷、耗尽层电荷、反型层电荷	C/m ^{2}
$W_d, C_d = \epsilon_s/W_d$	耗尽层宽度, 耗尽层电容	m, F/m ^{2}
$D_{\text{it}}, C_{\text{it}}$	界面陷阱密度, 界面陷阱电容	cm $^{-2}$ eV $^{-1}$, F/m ^{2}
$\gamma = \sqrt{2\epsilon_s q N_A} / C_{\text{ox}}$	体效应系数	V $^{1/2}$
$m = 1 + C_d/C_{\text{ox}}$	体因子 (亚阈值区还要加 $C_{\text{it}}/C_{\text{ox}}$)	1
$V_{\text{GS}}, V_{\text{DS}}, V_{\text{SB}}$	栅-源、漏-源、源-衬底电压	V
W, L	沟道宽度、长度	m
$V(y)$	沟道中电子准费米势 (沟道电势), $V(0) = 0, V(L) = V_{\text{DS}}$	V
I_D	漏极电流 (由漏流入为正)	A
$v_{\text{sat}}, \mathcal{E}_{\text{sat}} = v_{\text{sat}}/\mu$	饱和漂移速度, 临界场	Si 电子 $v_{\text{sat}} \approx 10^7$ cm/s
λ	沟道长度调制系数	V $^{-1}$
g_m, r_o	跨导, 输出电阻	S, Ω
S	亚阈值摆幅	mV/dec
α, f, C_L	翻转活动因子, 时钟频率, 负载电容	1, Hz, F
κ	等比例缩小因子	1
ℓ	自然长度 (短沟道效应的特征长度)	m

定义: 价带、导带、禁带宽度

在绝对零度下被电子完全填满的最高能带叫价带 (其顶部能量记为 E_v), 其上方第一个空能带叫导带 (底部能量记为 E_c), 二者之差 $E_g = E_c - E_v$ 叫禁带宽度。硅在 300 K 时 $E_g = 1.12$ eV[26], 二氧化硅约为 9 eV。

为什么满带不导电 能带中的每个状态由波矢 k 标记 (相当于电子的德布罗意波矢), 群速度 $v(k)$ 满足 $v(-k) = -v(k)$ 。满带中 k 与 $-k$ 的状态都被占据, 所有电子的速度之和为零, 加上电场也

是如此（电场只是让所有电子的 k 一起平移，满带平移后仍是满带）。所以只有部分填充的能带才能导电。金属的最高能带只填了一部分；绝缘体与半导体在 $T = 0$ 时价带全满、导带全空，都不导电。二者的区别只在 E_g 的大小：室温下 $k_B T \approx 0.026 \text{ eV}$ ，热运动能把电子从价带激发到导带的概率 $\propto e^{-E_g/2k_B T}$ （第 4.4 节），对硅约为 10^{-10} ，对 SiO_2 小到 10^{-76} 。

3.2 空穴

价带中少了一个电子会怎样？设价带只有波矢为 k_0 的一个状态空着。满带的总电流为零，所以

$$\begin{aligned} \mathbf{J} &= \sum_{\text{占据}} (-q) \mathbf{v}(k) \\ &= \underbrace{\sum_{\text{全部}} (-q) \mathbf{v}(k)}_{=0} - (-q) \mathbf{v}(k_0) \\ &= +q \mathbf{v}(k_0). \end{aligned} \quad (3.1)$$

整个价带的电流等于一个带正电荷 $+q$ 、以空状态的速度运动的粒子。这个“粒子”叫空穴。下一小节会说明，在外力作用下空穴的运动也像一个具有正质量的正电荷，所以在描述电流时可以把近满的价带完全用少量空穴代替。半导体中的电流因而由两种载流子承担：导带中的电子（浓度 n ）与价带中的空穴（浓度 p ）。

3.3 有效质量

定义：有效质量

在导带底附近，电子能量随波矢的关系可以展开为

$$E(k) \approx E_c + \frac{\hbar^2 k^2}{2m^*}, \quad (3.2)$$

形式与自由电子 $E = \hbar^2 k^2 / 2m$ 相同，但其中的 m^* 由晶体决定，称为有效质量。它刻画晶格周期势对电子运动的全部影响。

推导：有效质量的力学含义

第一步，电子是一个波包，其速度是群速度。由 $E = \hbar\omega$ ， $v = \frac{d\omega}{dk} = \frac{1}{\hbar} \frac{dE}{dk}$ ，代入式 (3.2) 得 $v = \hbar k / m^*$ 。

第二步，外力（例如电场力 $F = -q\mathcal{E}$ ）对电子做功的功率 Fv 等于电子能量的变化率： $Fv = \frac{dE}{dt} = \frac{dE}{dk} \frac{dk}{dt} = \hbar v \frac{dk}{dt}$ ，所以

$$\hbar \frac{dk}{dt} = F. \quad (3.3)$$

晶格的周期力已经包含在 $E(k)$ 中，这里的 F 只是外力。

第三步，对 $v = \hbar k / m^*$ 求导并用式 (3.3)： $\frac{dv}{dt} = \frac{\hbar}{m^*} \frac{dk}{dt} = \frac{F}{m^*}$ ，即 $m^* \frac{dv}{dt} = F$ 。电子在外力下的运动与质量为 m^* 的经典粒子一样。

价带顶附近 $E(k) \approx E_v - \hbar^2 k^2 / 2m_h$ ($m_h > 0$)，价带顶的电子有效质量为负。空状态的 k_0 也按式 (3.3) 随外力变化（它就是那个“缺席的电子”的状态），由第三步，其速度变化率为

$dv/dt = -F/m_h$, 其中 $F = -q\mathcal{E}$ 是电子受的力; 于是 $m_h dv/dt = +q\mathcal{E}$ 。这正是电荷 $+q$ 、质量 m_h 的粒子在电场中的运动方程: 空穴是带正电、正质量的准粒子。

硅的导带底有 6 个等价的“能谷”, 每个能谷中有效质量依方向而不同 (纵向 $0.98m_0$, 横向 $0.19m_0$, m_0 为自由电子质量) [26]。计算载流子浓度时这些细节可以合并成一个“态密度有效质量”: 导带 $m_c \approx 1.18m_0$, 价带 $m_v \approx 0.81m_0$ (300 K) [26]。本讲只需要知道 m^* 是一个与 m_0 同数量级的参数。

4 前置模块 II: 载流子统计

4.1 费米-狄拉克分布与费米能级

电子是费米子, 每个量子态 (含自旋) 最多容纳一个电子 (泡利不相容原理)。在温度 T 下, 一个能量为 E 的量子态被占据的概率是多少?

推导: 由玻尔兹曼因子推出费米-狄拉克分布

把这—个量子态当作“系统”, 晶体的其余部分 (含与之相连的导线、电源) 当作“热库”。系统只有两种状态: 空 (电子数 $n = 0$, 能量 0) 或占据 ($n = 1$, 能量 E)。总能量 E_{tot} 与总电子数 N_{tot} 守恒。

第一步, 系统处于某一状态的概率正比于此时热库可取的微观状态数 $\Omega_R(E_{\text{tot}} - nE, N_{\text{tot}} - n)$ (等概率原理)。

第二步, 热库很大, nE 与 n 相对于 E_{tot} 、 N_{tot} 很小, 把 $\ln \Omega_R$ 展开到一阶:

$$\begin{aligned} \ln \Omega_R(E_{\text{tot}} - nE, N_{\text{tot}} - n) \\ \approx \ln \Omega_R(E_{\text{tot}}, N_{\text{tot}}) - nE \frac{\partial \ln \Omega_R}{\partial E} - n \frac{\partial \ln \Omega_R}{\partial N}. \end{aligned}$$

由温度的统计定义 $\frac{\partial \ln \Omega_R}{\partial E} = \frac{1}{k_B T}$; 再定义化学势 μ 使 $\frac{\partial \ln \Omega_R}{\partial N} = -\frac{\mu}{k_B T}$ 。

第三步, 取指数, 得到占据与空两种状态的概率之比 $P(1)/P(0) = e^{-(E-\mu)/k_B T}$ 。由 $P(0) + P(1) = 1$, 占据概率为

$$f(E) = \frac{1}{e^{(E-\mu)/k_B T} + 1}. \quad (4.1)$$

这就是费米-狄拉克分布。在固体物理中化学势 μ 通常称为费米能级, 记为 E_F : 它是占据概率恰为 $\frac{1}{2}$ 的能量。

μ 的物理意义可以从第二步读出: $-\mu/k_B T$ 是热库每失去一个电子时 $\ln \Omega$ 的变化, 所以 μ 是“在保持熵不变的条件下向系统加入一个电子所需的能量”。它有一个与温度完全平行的性质:

要点: 平衡时费米能级处处相等

两个可以交换电子的物体 1、2 达到平衡时, 总熵 $\ln \Omega_1 + \ln \Omega_2$ 对电子转移 δN (从 2 到 1) 取极大: $\delta(\ln \Omega_1 + \ln \Omega_2) = \left(-\frac{\mu_1}{k_B T} + \frac{\mu_2}{k_B T}\right) \delta N = 0$, 所以 $\mu_1 = \mu_2$ 。若 $\mu_1 < \mu_2$, 电子自发地从 2 流向 1。因此: 处于热平衡的一个系统, 无论由多少种材料拼接而成, 费米能级处处相同; 外加电压 V 使两端的费米能级相差 qV (电压表测的正是费米能级之差除以 q)。能带图的全

部规则都建立在这一点上。

当 $E - E_F \gg k_B T$ 时, 分母中的 1 可以略去:

$$\begin{aligned} f(E) &\approx e^{-(E-E_F)/k_B T} \\ (E - E_F &\gtrsim 3k_B T), \end{aligned} \quad (4.2)$$

即熟悉的玻尔兹曼因子。 $E - E_F = 3k_B T$ 时近似值偏大 $e^{-3} \approx 5\%$ (图 3(a))。同理, 能量为 E 的状态空着 (有一个空穴) 的概率 $1 - f(E) \approx e^{-(E_F-E)/k_B T}$ ($E_F - E \gtrsim 3k_B T$)。

4.2 态密度

推导: 导带底附近的态密度

考虑边长为 L_0 的立方体晶体, 采用周期性边界条件 (波函数在相对两面取相同值), 允许的波矢为 $\mathbf{k} = \frac{2\pi}{L_0}(n_x, n_y, n_z)$, $n_{x,y,z}$ 为整数。每个允许的 $\mathbf{k}$ 在 k 空间占体积 $(2\pi/L_0)^3$ 。

第一步, 能量低于 E 的状态对应 k 空间中半径 $k = \sqrt{2m^*(E - E_c)}/\hbar$ 的球 (式 (3.2))。计入每个 $\mathbf{k}$ 有两个自旋态, 状态数为

$$\begin{aligned} \mathcal{N}(E) &= 2 \times \frac{\frac{4}{3}\pi k^3}{(2\pi/L_0)^3} \\ &= \frac{L_0^3}{3\pi^2} \left[\frac{2m^*(E - E_c)}{\hbar^2} \right]^{3/2}. \end{aligned}$$

第二步, 对 E 求导并除以体积 L_0^3 , 得到单位体积、单位能量间隔内的状态数 (态密度):

$$\begin{aligned} g(E) &= \frac{1}{2\pi^2} \left(\frac{2m^*}{\hbar^2} \right)^{3/2} \sqrt{E - E_c} \\ (E &> E_c). \end{aligned} \quad (4.3)$$

结果与 L_0 无关, 正如应当的那样。

4.3 导带电子浓度与有效态密度

推导: 非简并半导体的电子浓度

电子浓度等于态密度乘占据概率再对导带积分: $n = \int_{E_c}^{\infty} g(E)f(E) dE$ 。若 $E_c - E_F \gtrsim 3k_B T$, 整个导带都满足式 (4.2)。令 $x = (E - E_c)/k_B T$:

$$\begin{aligned} n &= \frac{1}{2\pi^2} \left(\frac{2m^*k_B T}{\hbar^2} \right)^{3/2} e^{-(E_c-E_F)/k_B T} \\ &\quad \times \int_0^{\infty} x^{1/2} e^{-x} dx. \end{aligned}$$

积分等于 $\sqrt{\pi}/2$ (附录 A.1)。整理得

$$\begin{aligned} n &= N_c e^{-(E_c-E_F)/k_B T}, \\ N_c &\equiv 2 \left(\frac{2\pi m_c k_B T}{h^2} \right)^{3/2}. \end{aligned} \quad (4.4)$$

同理，价带中的空穴浓度

$$p = N_v e^{-(E_F - E_v)/k_B T},$$

$$N_v \equiv 2 \left(\frac{2\pi m_v k_B T}{h^2} \right)^{3/2}. \quad (4.5)$$

式 (4.4) 的含义很直观：导带像是在 E_c 处集中了 N_c 个“有效状态”，每个状态的占据概率是玻尔兹曼因子。 N_c 是能量在导带底以上约 $k_B T$ 范围内的状态数（图 3(b)：导带电子的平均动能为 $\frac{3}{2}k_B T$ ，与经典理想气体相同）。代入数值， $2(2\pi m_0 k_B T/h^2)^{3/2} = 2.51 \times 10^{19} \text{ cm}^{-3}$ (300 K)，乘以 $(1.18)^{3/2}$ 与 $(0.81)^{3/2}$ 得 $N_c = 3.2 \times 10^{19} \text{ cm}^{-3}$ ， $N_v = 1.8 \times 10^{19} \text{ cm}^{-3}$ ，与文献值一致 [26]。

假设：非简并条件

式 (4.4)、(4.5) 要求费米能级离两个带边都至少约 $3k_B T$ ，即 $n \lesssim 0.05 N_c \approx 2 \times 10^{18} \text{ cm}^{-3}$ (p 同理)。本讲的衬底 ($N_A \sim 10^{16} - 10^{18} \text{ cm}^{-3}$) 满足此条件；强反型层表面电子浓度可达 10^{19} cm^{-3} 量级（图 5 情形约 $7 \times 10^{18} \text{ cm}^{-3}$ ），用式 (4.4) 会把电子浓度高估约 7%（费米-狄拉克积分的数值），相应的表面势误差不到 $0.1k_B T/q$ ；重掺杂的源、漏与多晶硅栅是简并的（费米能级进入能带），只用到“ E_F 约在带边”这一结论。

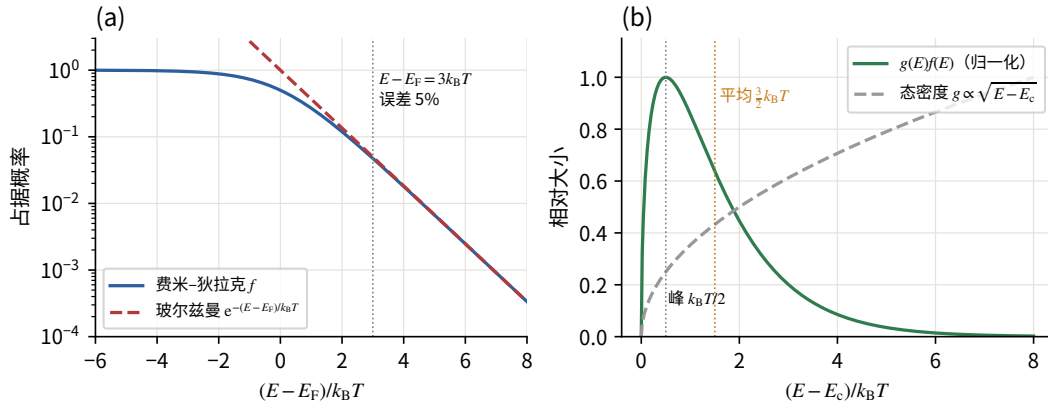


图 3 (a) 费米-狄拉克分布与玻尔兹曼近似（对数坐标）： $E - E_F > 3k_B T$ 时二者相差不到 5%。(b) 非简并情形下导带电子按能量的分布 $g(E)f(E) \propto \sqrt{E - E_c} e^{-(E - E_c)/k_B T}$ ：峰在导带底以上 $k_B T/2$ ，平均动能 $\frac{3}{2}k_B T$ （脚本 `scripts/fig_fd.py` 用数值积分核对）。

4.4 质量作用定律与本征载流子浓度

把式 (4.4) 与式 (4.5) 相乘， E_F 消去：

$$np = N_c N_v e^{-E_g/k_B T} \equiv n_i^2. \quad (4.6)$$

这就是质量作用定律：在热平衡、非简并条件下，电子与空穴浓度之积只取决于材料与温度，与掺杂无关。对于不掺杂的（本征）半导体， $n = p = n_i$ ，

$$n_i = \sqrt{N_c N_v} e^{-E_g/2k_B T}. \quad (4.7)$$

代入硅的数值， $n_i = \sqrt{3.2 \times 1.8 \times 10^{19}} \times e^{-1.12/(2 \times 0.02585)} \text{ cm}^{-3} \approx 0.94 \times 10^{10} \text{ cm}^{-3}$ 。本讲取 $n_i = 1.0 \times 10^{10} \text{ cm}^{-3}$ [26]（较新的精密测量给出 $9.65 \times 10^9 \text{ cm}^{-3}$ [27]，差别只使 ϕ_F 变化约 1 mV）。

n_i 对温度极为敏感: $d \ln n_i / dT = \frac{3}{2T} + \frac{E_g}{2k_B T^2} \approx 0.077 \text{ K}^{-1}$, 温度每升高约 9 K, n_i 就翻一番。这将决定 MOSFET 关断电流的温度依赖 (例题 2)。

本征半导体的费米能级称为本征能级 E_i 。由 $n = p$ 与式 (4.4)、(4.5), $E_i = \frac{1}{2}(E_c + E_v) + \frac{1}{2}k_B T \ln(N_v/N_c)$, 对硅只比禁带中央低 7 meV, 本讲把它取在禁带中央。用 E_i 作参考, 式 (4.4)、(4.5) 可以改写成 (把 $n_i = N_c e^{-(E_c - E_i)/k_B T}$ 代入相除)

$$\begin{aligned} n &= n_i e^{(E_F - E_i)/k_B T}, \\ p &= n_i e^{(E_i - E_F)/k_B T}. \end{aligned} \quad (4.8)$$

费米能级每比本征能级高 $k_B T$, 电子浓度就增大 e 倍、空穴浓度减小 e 倍。这是本讲最常用的一对公式。

4.5 掺杂：施主、受主与费米势

定义：施主与受主

在硅 (4 价) 中掺入 5 价的磷 (P) 或砷 (As), 杂质原子与周围 4 个硅原子成键后多出一个电子; 这个电子只被杂质离子微弱地束缚, 很容易进入导带, 这种杂质叫施主 (浓度 N_D), 掺施主的半导体叫 n 型。掺入 3 价的硼 (B), 杂质原子成键时缺一个电子, 很容易从价带夺取一个电子 (即向价带释放一个空穴), 这种杂质叫受主 (浓度 N_A), 掺受主的半导体叫 p 型。浓度较高的载流子叫多子, 另一种叫少子。

束缚能的估计 多出的电子与带正电的施主离子组成一个“类氢原子”, 但它处在介电常数为 $\epsilon_s = 11.7\epsilon_0$ 的介质中, 质量为有效质量 m^* 。把氢原子结果中的 $\epsilon_0 \rightarrow \epsilon_s$ 、 $m_0 \rightarrow m^*$: 束缚能 $E_d = 13.6 \text{ eV} \times \frac{m^*/m_0}{(\epsilon_s/\epsilon_0)^2}$, 玻尔半径 $a = 0.053 \text{ nm} \times \frac{\epsilon_s/\epsilon_0}{m^*/m_0}$ 。取电导有效质量 $m^* \approx 0.26m_0$, 得 $E_d \approx 0.026 \text{ eV}$, $a \approx 2.4 \text{ nm}$ 。实测磷与硼的电离能都约为 0.045 eV, 砷为 0.054 eV[26], 量级正确。 a 远大于晶格常数 (0.356 nm), 说明把晶体当作连续介质是自洽的。

室温下杂质几乎全部电离 束缚能与 $k_B T$ 相当, 似乎只有一部分施主会电离; 实际上导带有 $N_c \sim 10^{19} \text{ cm}^{-3}$ 个有效状态可去, “熵”大大有利于电离。定量地, 施主能级 (能量 $E_c - E_d$) 可以空着 (电离, 1 种方式) 或被一个电子占据 (自旋向上或向下, 2 种方式; 由于库仑排斥不能放两个电子)。由第 4.1 节的推导第三步, 占据与电离的概率之比为 $2e^{-(E_c - E_d - E_F)/k_B T}$; 再用式 (4.4) 消去 E_F :

$$\frac{N_D^0}{N_D^+} = 2 \frac{n}{N_c} e^{E_d/k_B T}. \quad (4.9)$$

对 $n = N_D = 10^{17} \text{ cm}^{-3}$ 的磷掺杂硅, $2 \times (10^{17}/3.2 \times 10^{19}) \times e^{0.045/0.02585} = 0.036$: 只有约 3.5% 的施主未电离。本讲一律假定杂质完全电离。

推导：掺杂半导体的载流子浓度与费米势

完全电离时, 半导体内部的电中性条件为 $p + N_D = n + N_A$ 。与 $np = n_i^2$ 联立, 对 p 型 ($N_D = 0$) 解二次方程:

$$p_0 = \frac{N_A}{2} + \sqrt{\frac{N_A^2}{4} + n_i^2} \approx N_A,$$

$$n_0 = \frac{n_i^2}{p_0} \approx \frac{n_i^2}{N_A} \quad (N_A \gg n_i).$$

下标 0 表示衬底体内的平衡值。由式 (4.8), $p_0 = n_i e^{(E_i - E_F)/k_B T}$, 于是体内费米能级在本征能级以下

$$E_i - E_F = q\phi_F,$$

$$\phi_F \equiv U_T \ln \frac{N_A}{n_i}. \quad (4.10)$$

ϕ_F 叫费米势。对 n 型衬底, 同样的定义给出负值 $\phi_F = -U_T \ln(N_D/n_i)$ 。

例题：数值：本讲的“标准器件”

本讲反复使用一个参数组: p 型衬底 $N_A = 10^{17} \text{ cm}^{-3}$, 栅氧化层 $t_{\text{ox}} = 10 \text{ nm}$, n^+ 多晶硅栅, 300 K。此时 $\phi_F = 0.02585 \times \ln(10^7) = 0.417 \text{ V}$, $p_0 = 10^{17} \text{ cm}^{-3}$, $n_0 = 10^3 \text{ cm}^{-3}$ 。少子浓度比多子小 14 个数量级——但下一节将看到, 只要把表面的能带向下弯 $2\phi_F \approx 0.83 \text{ V}$, 表面的电子浓度就会升到 10^{17} cm^{-3} 。

5 前置模块 III: 漂移、扩散与半导体静电学

5.1 漂移与迁移率

推导：德鲁德模型

导带电子在电场 $\mathcal{E}$ 中加速, 又不断被晶格振动 (声子) 与杂质散射。把散射的平均效果写成一个“阻尼力” $-m^*v/\tau$ (τ 为平均散射时间), 电子的平均速度满足

$$m^* \frac{dv}{dt} = -q\mathcal{E} - \frac{m^*v}{\tau}.$$

稳态时 $v = -\frac{q\tau}{m^*}\mathcal{E} \equiv -\mu_n\mathcal{E}$, $\mu_n = q\tau/m^*$ 叫电子的迁移率 (单位速度/单位电场)。电流密度为电荷密度乘平均速度: $J_n = (-q)nv = qn\mu_n\mathcal{E}$ 。空穴同理, $J_p = qp\mu_p\mathcal{E}$ 。

低掺杂硅中 $\mu_n \approx 1400 \text{ cm}^2/(\text{V s})$, $\mu_p \approx 450 \text{ cm}^2/(\text{V s})$ [26]。由 $\mu_n = q\tau/m^*$ (取 $m^* = 0.26m_0$), $\tau \approx 0.2 \text{ ps}$; 乘以热运动速度 $2.3 \times 10^5 \text{ m/s}$ [26], 平均自由程约 50 nm。MOSFET 中的电子被栅电场压在离界面约 1–2 nm 的薄层内, 额外受到界面粗糙度的散射, 反型层的有效迁移率 μ_{eff} 只有体材料的 1/5 到 1/2, 且随栅压增大而下降 [28]。电场很强时 ($\geq 10^4 \text{ V/cm}$) 电子从电场获得能量的速度超过散射带走能量的速度, 漂移速度趋于饱和值 $v_{\text{sat}} \approx 1 \times 10^7 \text{ cm/s}$ [26], 第 7.9 节再讨论。

5.2 扩散

推导：由一维随机行走推出菲克定律

设粒子每隔时间 τ_s 随机向左或向右跳一步，步长为 l ，浓度 $n(x)$ 随位置缓慢变化。考察穿过平面 x 的净通量（单位时间、单位面积穿过的粒子数）：位于 $x - l/2$ 附近一个步长内的粒子有一半向右跳过平面，位于 $x + l/2$ 附近的有一半向左跳过。单位时间的净通量为

$$\begin{aligned}\Phi &= \frac{1}{2} \frac{l}{\tau_s} [n(x - \frac{l}{2}) - n(x + \frac{l}{2})] \\ &\approx -\frac{l^2}{2\tau_s} \frac{dn}{dx} \equiv -D \frac{dn}{dx}.\end{aligned}$$

这就是菲克定律， $D = l^2/2\tau_s$ 叫扩散系数。粒子从浓处流向稀处，并不需要任何力——只是因为浓处跳出的粒子更多。

电子带电 $-q$ ，扩散电流密度 $J_n = (-q)\Phi = qD_n dn/dx$ ；空穴为 $J_p = -qD_p dp/dx$ 。于是电子与空穴的总电流密度为

$$\begin{aligned}J_n &= qn\mu_n\mathcal{E} + qD_n \frac{dn}{dx}, \\ J_p &= qp\mu_p\mathcal{E} - qD_p \frac{dp}{dx}.\end{aligned}\tag{5.1}$$

这叫漂移-扩散模型。

5.3 爱因斯坦关系

推导：由平衡态零电流推出 $D = \mu k_B T/q$

考虑处于热平衡、但内部有静电势 $\psi(x)$ 的半导体（例如掺杂不均匀）。电子在势场中的势能为 $-q\psi(x)$ ，由玻尔兹曼分布， $n(x) = n_{\text{ref}} e^{q\psi(x)/k_B T}$ ，于是

$$\frac{dn}{dx} = \frac{q}{k_B T} n \frac{d\psi}{dx} = -\frac{q}{k_B T} n \mathcal{E}.$$

热平衡时各处电流为零（否则能量会不断耗散）： $0 = J_n = qn\mu_n\mathcal{E} + qD_n \left(-\frac{q}{k_B T} n \mathcal{E}\right)$ 。对任意 $\mathcal{E}$ 成立，必须

$$\begin{aligned}\frac{D_n}{\mu_n} &= \frac{k_B T}{q} = U_T, \\ \text{同理 } \frac{D_p}{\mu_p} &= U_T.\end{aligned}\tag{5.2}$$

这是爱因斯坦关系：扩散（无规热运动）与迁移率（对外力的响应）由同一个散射过程决定，二者之比是热能 $k_B T$ 。硅中 $D_n = 1400 \times 0.02585 \approx 36 \text{ cm}^2/\text{s}$ ，与实测一致 [26]。

5.4 准费米能级：漂移与扩散合为一项

定义：准费米能级

有电流时半导体不处于平衡，费米能级没有定义。但电子在导带内部的散射极快 (~ 0.1 ps)，导带电子自身在每一点都近似达到局部平衡。仿照式 (4.8)，用局部的电子浓度定义电子准费米能级 $E_{Fn}(x)$ ：

$$\begin{aligned} n(x) &= n_i e^{[E_{Fn}(x) - E_i(x)]/k_B T}, \\ p(x) &= n_i e^{[E_i(x) - E_{Fp}(x)]/k_B T}. \end{aligned} \quad (5.3)$$

平衡时 $E_{Fn} = E_{Fp} = E_F$ 。

推导： $J_n = \mu_n n dE_{Fn}/dx$

第一步，静电势 ψ 使所有能级一起移动： $E_i(x) = \text{常数} - q\psi(x)$ (下一小节)。对式 (5.3) 第一式取对数再求导：

$$\frac{1}{n} \frac{dn}{dx} = \frac{1}{k_B T} \left(\frac{dE_{Fn}}{dx} + q \frac{d\psi}{dx} \right).$$

第二步，代入式 (5.1) 并用爱因斯坦关系 $qD_n = \mu_n k_B T$ ：

$$\begin{aligned} J_n &= -qn\mu_n \frac{d\psi}{dx} \\ &\quad + \mu_n k_B T \cdot \frac{n}{k_B T} \left(\frac{dE_{Fn}}{dx} + q \frac{d\psi}{dx} \right) \\ &= \mu_n n \frac{dE_{Fn}}{dx}. \end{aligned}$$

漂移项与扩散项中含 $d\psi/dx$ 的部分正好相消。空穴同理： $J_p = \mu_p p dE_{Fp}/dx$ 。

要点：准费米能级的作用

- 平衡 ($J = 0$) $\Leftrightarrow$ 准费米能级平直；有电流 $\Leftrightarrow$ 准费米能级倾斜，倾斜率与“载流子浓度 $\times$ 迁移率”成反比。
- 漂移与扩散不必分开讨论：电流只由准费米能级的梯度驱动。MOSFET 中，强反型时电流以漂移为主、亚阈值时以扩散为主，但两种情形都由同一个公式 (7.1) 描述 (第 7 节)。
- 把电子准费米能级写成 $E_{Fn} = -qV$ ， V 就是“电子所感受到的电压”：接在半导体上的金属电极测到的正是 V 。

5.5 能带图、泊松方程与德拜屏蔽

能带图就是电势图 若半导体中有静电势 $\psi(x)$ ，一个电子在 x 处的额外势能为 $-q\psi(x)$ 。这对导带底、价带顶、本征能级都一样，所以

$$\begin{aligned} E_c(x) &= E_{c0} - q\psi(x), \\ E_v(x) &= E_{v0} - q\psi(x), \\ E_i(x) &= E_{i0} - q\psi(x), \end{aligned} \quad (5.4)$$

下标 0 表示体内 ($\psi = 0$)。“能带弯曲”就是静电势的空间变化：电势高处能带低。平衡时 E_F 平直，代入式 (4.8) 与式 (4.10)：

$$\begin{aligned} n(x) &= n_0 e^{\psi(x)/U_T}, \\ p(x) &= p_0 e^{-\psi(x)/U_T}. \end{aligned} \quad (5.5)$$

这就是电子与空穴各自在势能 $\mp q\psi$ 中的玻尔兹曼分布。

推导：一维泊松方程

设一切只随 x 变化。在 x 与 $x + dx$ 之间取截面积为 A 的薄片，对它用电介质中的高斯定理（电位移 $\mathcal{D} = \varepsilon_s \mathcal{E}$ 的通量等于自由电荷）： $[\mathcal{D}(x + dx) - \mathcal{D}(x)]A = \rho A dx$ ，即 $\varepsilon_s d\mathcal{E}/dx = \rho$ 。再用 $\mathcal{E} = -d\psi/dx$ ：

$$\begin{aligned} \frac{d^2\psi}{dx^2} &= -\frac{\rho}{\varepsilon_s}, \\ \rho &= q(p - n + N_D - N_A). \end{aligned} \quad (5.6)$$

这里的“自由电荷”是载流子与电离杂质；硅原子的极化已经包含在介电常数 $\varepsilon_s = 11.7\varepsilon_0$ 中。对一个跨过两种介质界面的薄片用同样的高斯定理，得到界面条件：若界面上没有自由电荷， $\varepsilon_{ox}\mathcal{E}_{ox} = \varepsilon_s\mathcal{E}_s$ （电位移法向分量连续）。

推导：德拜屏蔽

在 p 型体内 ($\rho = 0$ 时 $p_0 - n_0 = N_A$) 加一个小的电势扰动 $|\psi| \ll U_T$ 。由式 (5.5) 展开到一阶： $p - n - N_A \approx p_0(1 - \psi/U_T) - n_0(1 + \psi/U_T) - N_A = -(p_0 + n_0)\psi/U_T$ 。代入泊松方程：

$$\begin{aligned} \frac{d^2\psi}{dx^2} &= \frac{\psi}{L_D^2}, \\ L_D &\equiv \sqrt{\frac{\varepsilon_s U_T}{q(p_0 + n_0)}}. \end{aligned} \quad (5.7)$$

在 $x > 0$ 中有界的解是 $\psi = \psi(0)e^{-x/L_D}$ ：外加电场被载流子的重新分布屏蔽，只能伸进约一个德拜长度 L_D 。

$N_A = 10^{17} \text{ cm}^{-3}$ 时 $L_D = 12.9 \text{ nm}$ ； $N_A = 10^{15} \text{ cm}^{-3}$ 时为 129 nm 。金属中电子浓度约 10^{22} cm^{-3} ，屏蔽长度不到 0.1 nm^1 ，外加电场完全进不去——这说明了场效应器件为什么必须用半导体而不能用金属薄膜：在金属中电场只能改变表面一个原子层内的电荷，与整个薄膜的电子总数相比微不足道。半导体载流子少，电场能伸进几十纳米，同时“感应电荷”本身就足以与原有载流子相比。

6 MOS 电容：栅电场如何弯曲能带

6.1 结构与理想化假设

MOS 电容由栅（金属或重掺杂多晶硅；多晶硅是由许多微小晶粒组成的硅，重掺杂后像金属一样导电，耐高温工艺）、栅氧化层（ SiO_2 ，厚 t_{ox} ）和半导体衬底组成，衬底背面接地。它就是一

¹金属中的电子是简并的，屏蔽长度应当用托马斯-费米公式计算，但数量级仍是一个原子间距。

个平行板电容器，只是下极板换成了半导体。

假设：理想 MOS 结构

- 一维：栅的面积远大于 t_{ox} 与半导体中电场的穿透深度，一切只随 x 变化；
- 氧化层是理想绝缘体，没有电流穿过（第 9.4 节讨论极薄氧化层的隧穿）；
- 半导体均匀掺杂、杂质完全电离、非简并；半导体中处于热平衡（没有电流），费米能级平直；
- 先假定氧化层中和界面上没有电荷，第 6.2、6.7 节再计入固定电荷与界面陷阱。

6.2 功函数与平带电压

定义：真空能级、功函数、电子亲和能

真空能级 E_{vac} 是一个电子静止于材料表面外附近时的能量。金属的功函数 $q\phi_{\text{m}} = E_{\text{vac}} - E_{\text{F}}$ ；半导体的电子亲和能 $q\chi = E_{\text{vac}} - E_{\text{c}}$ （硅 $\chi = 4.05 \text{ V}$ [26]），其功函数 $q\phi_{\text{s}} = E_{\text{vac}} - E_{\text{F}}$ 随掺杂而变。对 p 型衬底， $E_{\text{c}} - E_{\text{F}} = E_{\text{g}}/2 + q\phi_{\text{F}}$ ，所以 $\phi_{\text{s}} = \chi + E_{\text{g}}/2q + \phi_{\text{F}}$ 。记 $\phi_{\text{ms}} \equiv \phi_{\text{m}} - \phi_{\text{s}}$ （功函数差）。

推导：栅与衬底之间的静电势差

真空能级就是 “ $-q \times$ 静电势”（加一个常数），所以栅与衬底体内的静电势差 $\Phi_{\text{g}} - \Phi_{\text{bulk}} = -(E_{\text{vac,g}} - E_{\text{vac,s}})/q$ 。用功函数的定义把真空能级换成费米能级：

$$\begin{aligned} & E_{\text{vac,g}} - E_{\text{vac,s}} \\ &= (E_{\text{F,g}} + q\phi_{\text{m}}) - (E_{\text{F,s}} + q\phi_{\text{s}}) \\ &= (E_{\text{F,g}} - E_{\text{F,s}}) + q\phi_{\text{ms}}. \end{aligned}$$

外加栅压 V_{G} （栅相对衬底）使两边费米能级相差 $E_{\text{F,g}} - E_{\text{F,s}} = -qV_{\text{G}}$ （第 4.1 节要点）。于是

$$\Phi_{\text{g}} - \Phi_{\text{bulk}} = V_{\text{G}} - \phi_{\text{ms}} = V_{\text{ox}} + \psi_{\text{s}}, \quad (6.1)$$

最后一步说明：这个静电势差一部分降在氧化层上（ V_{ox} ），一部分降在半导体中（表面势 ψ_{s} ，即表面相对体内的电势）。

当 $V_{\text{G}} = \phi_{\text{ms}}$ 时，各处都没有电场，能带是平的，这个栅压叫平带电压。即使 $V_{\text{G}} = 0$ ，只要两种材料的功函数不同，用导线连接后电子就会从功函数小的一侧流向大的一侧，直到费米能级对齐，此时电容器已经带电、能带已经弯曲。

若 Si/SiO₂ 界面附近有单位面积为 Q_{f} 的固定正电荷（热氧化工艺的残余，现代工艺中约 $10^{10} \text{ q cm}^{-2}$ [27]），平带时半导体中没有电荷，由高斯定理，栅上的电荷 $Q_{\text{G}} = -Q_{\text{f}}$ ，氧化层上的电压 $V_{\text{ox}} = Q_{\text{G}}/C_{\text{ox}}$ 。所以

$$V_{\text{FB}} = \phi_{\text{ms}} - \frac{Q_{\text{f}}}{C_{\text{ox}}}, \quad C_{\text{ox}} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}}}. \quad (6.2)$$

多晶硅栅：重掺杂 n^+ 多晶硅的费米能级约在导带底， $\phi_{\text{m}} \approx \chi$ ，于是 $\phi_{\text{ms}} \approx -(E_{\text{g}}/2q + \phi_{\text{F}})$ ； p^+

多晶硅的 E_F 约在价带顶, $\phi_m \approx \chi + E_g/q$, $\phi_{ms} \approx E_g/2q - \phi_F$ 。标准器件 (n^+ 栅): $V_{FB} = -(0.56 + 0.417) = -0.98 \text{ V}$, $C_{ox} = 3.9 \times 8.854 \times 10^{-12} \text{ F/m}/10 \text{ nm} = 3.45 \times 10^{-3} \text{ F/m}^2$ 。

6.3 栅压的分配与三个工作区

设半导体中单位面积的总电荷为 Q_s 。整个结构电中性, 栅上的电荷为 $Q_G = -Q_s$ (固定电荷已归入 V_{FB})。氧化层中没有电荷, 其中电场均匀, 由高斯定理 $\mathcal{E}_{ox} = Q_G/\varepsilon_{ox}$, $V_{ox} = \mathcal{E}_{ox}t_{ox} = -Q_s/C_{ox}$ 。代入式 (6.1):

$$V_G = V_{FB} + \psi_s - \frac{Q_s(\psi_s)}{C_{ox}} \quad (6.3)$$

这是 MOS 静电学的总方程: 栅压一部分降在半导体表面 (ψ_s), 一部分降在氧化层上 ($-Q_s/C_{ox}$)。剩下的任务是求出半导体的“电荷-电压关系” $Q_s(\psi_s)$ 。在此之前, 先定性看三种情形 (图 4):

- **积累** ($V_G < V_{FB}$, $\psi_s < 0$): 栅带负电, 表面能带向上弯, 由式 (5.5), 表面空穴浓度 $p_0 e^{|\psi_s|/U_T}$ 指数增大, 多子在表面“积累”, $Q_s > 0$ 。
- **耗尽** ($0 < \psi_s < \phi_F$, 即 V_G 介于 V_{FB} 与使 $\psi_s = \phi_F$ 的栅压之间): 栅带正电, 空穴被推离表面, 留下带负电的电离受主, 形成没有可动载流子的耗尽层。
- **反型** ($\psi_s > \phi_F$): 表面的 E_i 降到 E_F 以下, 由式 (4.8), 表面电子比空穴多, p 型表面“反转”成 n 型。 $\phi_F < \psi_s < 2\phi_F$ 叫弱反型, $\psi_s > 2\phi_F$ 叫强反型 (判据的由来见第 6.6 节)。反型层中的电子与衬底之间隔着耗尽层, 这层电子就是 MOSFET 的导电沟道。

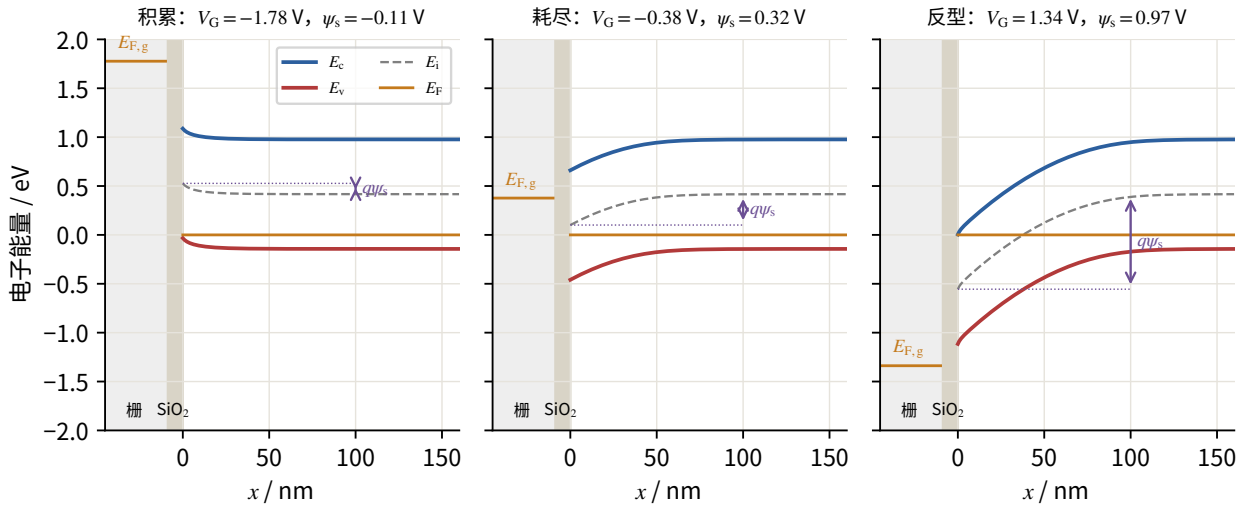


图4 标准器件 ($N_A = 10^{17} \text{ cm}^{-3}$, $t_{ox} = 10 \text{ nm}$, n^+ 多晶硅栅) 在积累、耗尽、反型三种状态下的能带图, 能量零点取衬底费米能级 E_F 。 E_c 、 E_v 、 E_i 由一维泊松方程的数值解 (有限差分 + 牛顿迭代, `scripts/mos1d.py`) 按式 (5.4) 画出; 箭头为 $q\psi_s$ 。左侧灰色区为栅, 其费米能级 $E_{F,g} = E_F - qV_G$; 米色窄条为氧化层 (氧化层的能带远超出图示范围, 未画)。反型时表面的 E_i 已低于 E_F 。

6.4 泊松方程的第一积分：表面电荷的精确式

推导： $Q_s(\psi_s)$

第一步，写出电荷密度。由式 (5.5) 与体内电中性 $p_0 - n_0 - N_A = 0$ ：

$$\rho(\psi) = q \left[p_0 \left(e^{-\psi/U_T} - 1 \right) - n_0 \left(e^{\psi/U_T} - 1 \right) \right].$$

ρ 只通过 ψ 依赖于 x ，这是能求出第一积分的关键。

第二步，泊松方程两边乘以 $2 d\psi/dx$ 。左边 $2\psi'\psi'' = d(\psi'^2)/dx$ ；右边 $-\frac{2}{\varepsilon_s}\rho(\psi)\psi' = -\frac{2}{\varepsilon_s}\frac{d}{dx}\int_0^\psi \rho d\psi'$ 。从深处体内 ($\psi = 0, \psi' = 0$) 积分到 x ：

$$\left(\frac{d\psi}{dx} \right)^2 = -\frac{2}{\varepsilon_s} \int_0^\psi \rho(\psi') d\psi'.$$

第三步，完成积分 ($\int_0^\psi (e^{\mp u/U_T} - 1) du$ 是初等积分)：

$$\begin{aligned} \left(\frac{d\psi}{dx} \right)^2 &= \frac{2qp_0U_T}{\varepsilon_s} F^2(\psi), \\ F^2 &= (e^{-\psi/U_T} + \psi/U_T - 1) \\ &\quad + \frac{n_0}{p_0} (e^{\psi/U_T} - \psi/U_T - 1). \end{aligned} \quad (6.4)$$

第四步，求半导体中的总电荷。取一个柱体，上底在界面内侧，下底在深处体内（那里没有电场）。由高斯定理，柱内电荷等于电位移穿出柱面的通量： $Q_s = -\varepsilon_s \mathcal{E}_s$ ， $\mathcal{E}_s = -\psi'(0)$ 为表面电场（指向 $+x$ 为正）。 $\psi_s > 0$ 时 ψ 向体内减小， $\psi'(0) < 0$ ， $\mathcal{E}_s > 0$ ， $Q_s < 0$ 。于是

$$\begin{aligned} Q_s(\psi_s) &= -\text{sgn}(\psi_s) \\ &\quad \times \sqrt{2\varepsilon_s k_B T p_0} F(\psi_s). \end{aligned} \quad (6.5)$$

式 (6.5) 在第 6.1 节的假设下是精确的。 F^2 中的三类项各有物理意义（另有常数 -1 ）： $e^{-\psi/U_T}$ 来自空穴（积累）， ψ/U_T 来自电离受主（耗尽）， $(n_0/p_0)e^{\psi/U_T}$ 来自电子（反型）。注意 $n_0/p_0 = (n_i/N_A)^2 = e^{-2\phi_F/U_T}$ ，所以电子项可以写成 $e^{(\psi-2\phi_F)/U_T}$ 。三个区的近似（图 6(a)）：

积累 ($-\psi_s \gg U_T$)：

$$Q_s \approx \sqrt{2\varepsilon_s k_B T N_A} e^{-\psi_s/2U_T}, \quad (6.6)$$

耗尽与弱反型 ($U_T \ll \psi_s \lesssim 2\phi_F$)：

$$Q_s \approx -\sqrt{2\varepsilon_s q N_A (\psi_s - U_T)}, \quad (6.7)$$

强反型 ($\psi_s \gtrsim 2\phi_F + 4U_T$)：

$$Q_s \approx -\sqrt{2\varepsilon_s k_B T N_A} e^{(\psi_s - 2\phi_F)/2U_T}. \quad (6.8)$$

积累与强反型时， $|Q_s|$ 每增大 e 倍， ψ_s 只需增加 $2U_T \approx 52 \text{ mV}$ ；耗尽时 $|Q_s| \propto \sqrt{\psi_s}$ ，增长很慢。

6.5 耗尽近似

假设：耗尽近似

在 $0 < x < W_d$ 内可动载流子全部被耗尽， $\rho = -qN_A$ ；在 $x > W_d$ 处半导体电中性， $\rho = 0$ 。适用条件： $U_T \ll \psi_s \lesssim 2\phi_F$ 。理由：由式 (5.5)，电势每升高 U_T ，空穴浓度降到 $1/e$ ；空穴从 p_0 降到可以忽略的过渡区只有约一个德拜长度宽，而 $W_d/L_D = \sqrt{2\psi_s/U_T}$ （由下面的结果）， $\psi_s = 2\phi_F$ 时约为 8。被忽略的效应：过渡区的拖尾（给出式 (6.7) 中的 $-U_T$ 修正）与反型电子。

推导：耗尽层宽度与电荷

在耗尽层内 $\psi'' = qN_A/\epsilon_s$ 。边界条件： $x = W_d$ 处 $\psi = 0$ 、 $\psi' = 0$ （与中性区平滑相接）。积分两次： $\psi(x) = \frac{qN_A}{2\epsilon_s}(W_d - x)^2$ 。令 $x = 0$ 处 $\psi = \psi_s$ ：

$$W_d = \sqrt{\frac{2\epsilon_s\psi_s}{qN_A}}, \quad (6.9)$$

$$Q_d = -qN_A W_d = -\sqrt{2\epsilon_s q N_A \psi_s}.$$

检验：与精确解的比较

- 式 (6.9) 中的 Q_d 与精确式的耗尽区极限 (6.7) 只差 $\psi_s \rightarrow \psi_s - U_T$ ，在 $\psi_s = 2\phi_F$ 时相对误差为 $U_T/(4\phi_F) = 1.6\%$ 。
- 图 5 把耗尽近似的抛物线电势与泊松方程的数值解比较： $\psi(x)$ 几乎重合；电荷密度的数值解在 W_d 处不是台阶而是宽约 $2L_D$ 的过渡区。
- 量纲： $\epsilon_s\psi_s/(qN_A)$ 的单位是 $(F/m) \cdot V/(C/m^3) = m^2$ ，正确。标准器件在 $\psi_s = 2\phi_F = 0.833 V$ 时 $W_d = 104 nm$ ， $Q_d = -1.66 \times 10^{-3} C/m^2$ （每平方厘米约 1.04×10^{12} 个电离受主）。

耗尽层的小信号电容为

$$C_d \equiv -\frac{dQ_d}{d\psi_s} = \sqrt{\frac{\epsilon_s q N_A}{2\psi_s}} = \frac{\epsilon_s}{W_d}, \quad (6.10)$$

恰好是极板间距为 W_d 的平行板电容： ψ_s 变化时，电荷的增减发生在耗尽层边缘。

6.6 强反型与阈值电压

判据 $\psi_s = 2\phi_F$ 的由来 由式 (5.5)，表面电子浓度 $n_s = n_0 e^{\psi_s/U_T}$ 。当 $\psi_s = 2\phi_F$ 时， $n_s = (n_i^2/N_A)(N_A/n_i)^2 = N_A$ ：表面 n 型的程度与体内 p 型的程度相同。这是一个约定，但它标出了一个物理上的转折：式 (6.4) 中电子项 $e^{(\psi-2\phi_F)/U_T}$ 在 $\psi = 2\phi_F$ 时为 1，受主项 $\psi/U_T \approx 32$ ；此后电子项指数增长，在 $\psi_s - 2\phi_F = U_T \ln(\psi_s/U_T) \approx 3.5U_T \approx 0.09 V$ 时超过受主项（脚本 fig_qs.py 数值求得 $3.55U_T$ ）。超过这一点后，表面势只需再增加很少，反型电荷就能大量增加——表面势被“钉”在 $2\phi_F$ 附近。

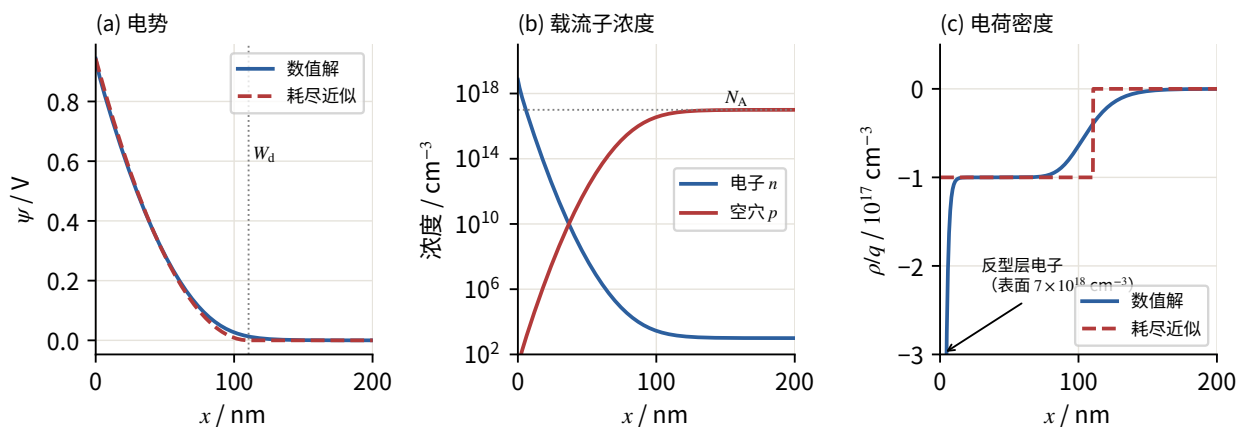


图 5 标准器件在 $V_G = V_T + 0.5 \text{ V}$ ($\psi_s = 0.943 \text{ V}$) 时半导体中的 (a) 电势、(b) 载流子浓度、(c) 电荷密度：一维泊松方程数值解（实线）与耗尽近似（虚线）。数值解的表面电场与第一积分 (6.4) 相差 $< 10^{-5}$ 。(b) 中表面电子浓度约 $7 \times 10^{18} \text{ cm}^{-3}$ ，在约 1 nm 内下降到 $1/e$ ；(c) 中这层反型电子表现为表面处的尖峰。

推导：阈值电压与反型层电荷

在强反型区把半导体电荷分成两部分： $Q_s = Q_d + Q_n$ ， Q_d 为耗尽层的电离受主， Q_n 为反型层电子。

第一步（近似）：强反型时 $\psi_s \approx 2\phi_F$ 不再变化，所以耗尽层也不再加宽， $Q_d \approx -\sqrt{2\epsilon_s q N_A \cdot 2\phi_F}$ 为常数。

第二步，代入栅压方程 (6.3)： $V_G = V_{FB} + 2\phi_F - (Q_d + Q_n)/C_{ox}$ 。阈值电压定义为反型层电荷外推到零 ($Q_n = 0$) 时的栅压：

$$V_T = V_{FB} + 2\phi_F + \frac{\sqrt{2\epsilon_s q N_A \cdot 2\phi_F}}{C_{ox}} \quad (6.11)$$

第三步，两式相减（适用于 $V_G > V_T$ ）：

$$Q_n = -C_{ox}(V_G - V_T) \quad (6.12)$$

阈值以上，栅压每增加 1 V ，表面就多出 $C_{ox} \times 1 \text{ V}$ 的电子——就像普通电容器的下极板一样。

V_T 的三项依次是：消除功函数差与固定电荷所需的栅压、把能带弯曲 $2\phi_F$ 所需的半导体电压降、支撑耗尽层电荷所需的氧化层电压。标准器件： $V_T = -0.977 + 0.833 + \frac{1.663 \times 10^{-3} \text{ C/m}^2}{3.453 \times 10^{-3} \text{ F/m}^2} = 0.338 \text{ V}$ 。

检验：式 (6.12) 有多准？

图 6(c) 把式 (6.12) 与精确解（由式 (6.3)、(6.5) 解出 ψ_s ，再对电子浓度积分求 Q_n ）比较。精确的 Q_n 与 $C_{ox}(V_G - V_T)$ 之比 (scripts/verify_qn.py)：

$V_G - V_T$	0.25 V	0.5 V	1.0 V	1.5 V	2.0 V
标准器件 ($t_{ox} = 10 \text{ nm}$)	0.615	0.742	0.838	0.879	0.903
$t_{ox} = 5 \text{ nm}$	0.562	0.713	0.822	0.868	0.893

也就是说，过驱动 0.5 V 时式 (6.12) 把反型电荷（因而把第 7 节的线性区电流）高估约 35%，1 V 时仍高估约 20%；数字电路正好工作在 $V_G - V_T \approx 0.5\text{--}0.8\text{ V}$ 这一段。差别来自两处：

- **斜率：** ψ_s 并非严格不变。由式 (6.8)， $|Q_n| \approx \sqrt{2\varepsilon_s k_B T N_A} e^{(\psi_s - 2\phi_F)/2U_T}$ ，所以“反型层电容” $C_{\text{inv}} \equiv -dQ_n/d\psi_s = |Q_n|/2U_T$ 有限；它与 C_{ox} 串联，局部斜率 $-dQ_n/dV_G = C_{\text{ox}} C_{\text{inv}} / (C_{\text{ox}} + C_{\text{inv}} + C_d)$ 。 $|Q_n| = C_{\text{ox}} \times 1\text{ V}$ 时 $C_{\text{inv}} \approx 19C_{\text{ox}}$ ，斜率约 $0.94C_{\text{ox}}$ ；越靠近阈值， C_{inv} 越小，斜率越低。
- **截距：**记 $\Delta \equiv \psi_s - 2\phi_F$ 。由上式反解， $\Delta = 2U_T \ln(|Q_n|/\sqrt{2\varepsilon_s k_B T N_A})$ ，对标准器件 $V_G - V_T = 0.5, 1.0, 1.5\text{ V}$ 时分别为 0.110、0.139、0.156 V（数值解）。表面势多出 Δ 时，耗尽层也随之加深，多吸收 $C_d \Delta$ 的电荷，所以栅压方程给出 $-Q_n/C_{\text{ox}} \approx V_G - V_T - m\Delta(V_G)$ ， $m = 1 + C_d/C_{\text{ox}}$ （即第 7.7 节的体因子）。在某一过驱动 V_{ov} 处作切线外推到 $Q_n = 0$ ：因 $d\Delta/dV_G \approx 2U_T/V_{\text{ov}}$ ，到一阶，外推截距

$$V_T^* - V_T \approx m[\Delta(V_{\text{ov}}) - 2U_T]$$

（推导与实验题第 (6) 问相同，`verify_sympy.py` 核对）。所以 V_T^* 依赖拟合区间：用 $V_G - V_T = 1\text{--}2\text{ V}$ 的数据作直线拟合得 $V_T^* - V_T = 0.135\text{ V}$ （斜率 0.967；切点 1.5 V 处 $m(\Delta - 2U_T) = 0.135\text{ V}$ ）；用 $0.3\text{--}0.8\text{ V}$ 的数据只得 0.090 V（斜率 0.910）。本讲图 6(c) 与图 9 用 $V_G = 1.2\text{--}2.0\text{ V}$ ($V_G - V_T = 0.86\text{--}1.66\text{ V}$) 拟合，得 $V_T^* = V_T + 0.128\text{ V} = 0.466\text{ V}$ 。

结论：式 (6.11)、(6.12) 抓住了物理，但阈值上偏 0.1 V 量级， $V_G - V_T \lesssim 0.5\text{ V}$ 时 $|Q_n|$ 被高估 30% 以上。实测器件的阈值都是按某种约定（如在指明区间内的线性外推）提取的，比较阈值时必须说明约定与区间。

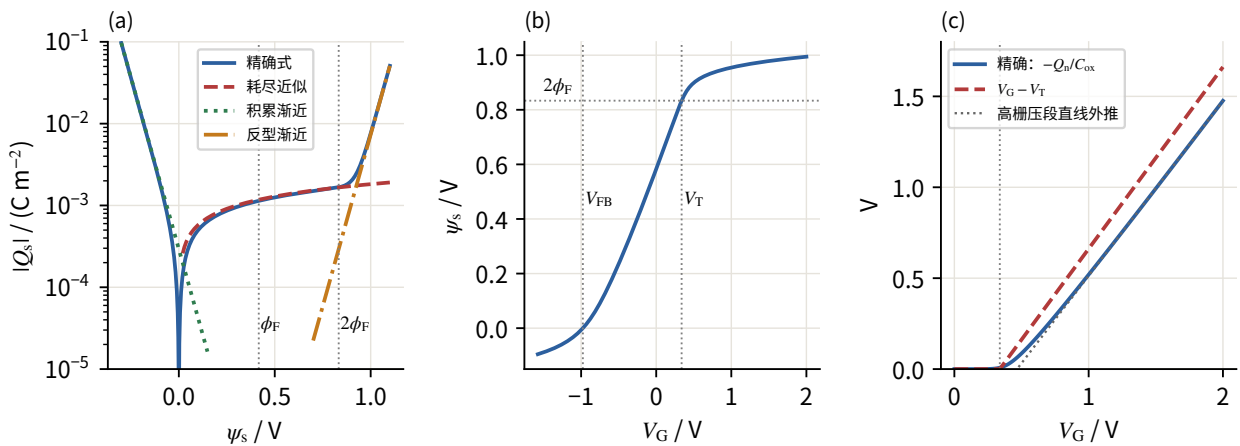


图 6 标准器件的 MOS 静电学。(a) 半导体表面电荷 $|Q_s|$ 与表面势：精确式 (6.5) 与三个区的近似 (6.6)–(6.8)； $\psi_s = 0$ 处 Q_s 变号。(b) 表面势随栅压的变化：越过 V_T 后 ψ_s 几乎被钉住。(c) 反型层电荷：精确解、式 (6.12) 与 $V_G = 1.2\text{--}2.0\text{ V}$ 段的直线外推（截距 $V_T^* = V_T + 0.128\text{ V}$ ；截距依赖拟合区间，见检验框）。

反型层有多薄 表面附近 $\psi \approx \psi_s - \varepsilon_s x$ ，由式 (5.5)，电子浓度按 $e^{-\varepsilon_s x/U_T}$ 衰减，厚度约 U_T/ε_s 。图 5 的情形 $\varepsilon_s \approx 2.9 \times 10^7\text{ V/m}$ ，厚度约 0.9 nm（数值解中降到 $1/e$ 的深度为 1.0 nm）。电子被关在宽约 1 nm 的三角形势阱中，这已接近电子的德布罗意波长，量子化效应会使电子的平均位置离

开界面 1–2 nm，并使阈值电压略有升高 [28]。本讲用经典统计，这一修正不影响主要结论。

p 沟道 在 n 型衬底上，一切符号反过来：栅压要足够负，表面才反型为 p 型（空穴沟道）， $V_T = V_{FB} - 2|\phi_F| - \sqrt{2\epsilon_s q N_D \cdot 2|\phi_F|}/C_{ox} < 0$ ， $Q_p = -C_{ox}(V_G - V_T) > 0$ 。

交互演示：MOS 结构的能带图与电荷分布。拖动栅压 V_G ，观察表面势 ψ_s 、耗尽层宽度与反型层电荷如何变化；可调衬底掺杂 N_A 与氧化层厚度 t_{ox} 。计算用泊松方程第一积分的精确式（数值求 $\psi(x)$ ），读数中给出 V_{FB} 、 V_T 、 $2\phi_F$ 与当前所处的工作区。

网页版中此处为可调参数的交互动画（演示编号 mosfet-band）。

6.7 界面陷阱：表面态为什么能让场效应“消失”

定义：界面陷阱密度

半导体表面的原子少了一侧的邻居，留下未成键的轨道（“悬挂键”），在禁带中形成局域的电子能级，叫表面态；在 Si/SiO₂ 界面上它们叫界面陷阱。设单位面积、单位能量间隔内的界面陷阱数为 D_{it} （常用单位 $\text{cm}^{-2}\text{eV}^{-1}$ ），并设它们与半导体交换电子足够快，占据情况由表面处的 E_F 决定（低于 E_F 的填满，高于的空着）。

推导：界面陷阱电容

表面势增加 $d\psi_s$ ，表面处的能带相对 E_F 下移 $q d\psi_s$ ，于是能量宽度为 $q d\psi_s$ 的一段陷阱从 E_F 之上移到之下而被电子填满，界面电荷变化 $dQ_{it} = -q \cdot D_{it} \cdot q d\psi_s$ 。定义界面陷阱电容

$$C_{it} \equiv -\frac{dQ_{it}}{d\psi_s} = q^2 D_{it}. \quad (6.13)$$

（若 D_{it} 以 $\text{cm}^{-2}\text{eV}^{-1}$ 为单位，则 C_{it} 的数值 $= 1.602 \times 10^{-19} \times D_{it}$ ，单位 F/cm^2 。）界面电荷与半导体电荷一起出现在栅压方程中： $V_G = V_{FB} + \psi_s - (Q_s + Q_{it})/C_{ox}$ 。对 ψ_s 求导：

$$\begin{aligned} \frac{dV_G}{d\psi_s} &= 1 + \frac{C_s + C_{it}}{C_{ox}}, \\ C_s &\equiv -\frac{dQ_s}{d\psi_s}. \end{aligned} \quad (6.14)$$

式 (6.14) 是一个电容分压器： C_{ox} 与 $(C_s + C_{it})$ 串联，栅压的变化按电容的反比分配。栅上新增的电荷 dQ_G 在半导体（ C_s ）与陷阱（ C_{it} ）之间按电容正比分配：只有比例 $C_s/(C_s + C_{it})$ 进入半导体，其余被陷阱俘获、不能导电。

要点：表面态的屏蔽与 Si/SiO₂ 界面的意义

- 裸露的半导体表面每平方厘米约有 10^{15} 个表面原子，若其中百分之几的悬挂键在禁带中提供能级， $D_{it} \sim 10^{13}\text{--}10^{14} \text{ cm}^{-2}\text{eV}^{-1}$ ， $C_{it} \sim 10^{-6}\text{--}10^{-5} \text{ F}/\text{cm}^2$ ，比半导体的电容大一两个数量级：栅上的电荷几乎全被表面态吸收，表面势几乎不动（“费米能级钉扎”），电导几乎不变。这就是 1940 年代场效应实验失败的原因（例题 1 定量讨论）。
- 热生长的 SiO₂ 使界面上的硅原子几乎都成键，再经含氢气氛退火，现代工艺的 $D_{it} \sim 10^{10} \text{ cm}^{-2}\text{eV}^{-1}$ [27, 28]， $C_{it} \approx 1.6 \times 10^{-9} \text{ F}/\text{cm}^2$ ，比 C_{ox} （10 nm 氧化层为 $3.5 \times 10^{-7} \text{ F}/\text{cm}^2$ ）

小两个数量级以上。Atalla 的热氧化工艺 [11] 因此成为 MOSFET 的前提。

- 即使 D_{it} 不大，它也使亚阈值摆幅变差（第 8.3 节）；实验题就用这一点测 D_{it} 。

6.8 C - V 特性

推导：MOS 电容的小信号电容

在直流栅压上叠加一个小的交变电压 δV_G ，测得栅上电荷的变化 δQ_G ， $C \equiv \delta Q_G / \delta V_G$ （单位面积）。由式 (6.3)（先不计界面陷阱）与 $Q_G = -Q_s$ ： $\delta V_G = \delta \psi_s + \delta Q_G / C_{ox}$ ， $\delta Q_G = C_s \delta \psi_s$ 。消去 $\delta \psi_s$ ：

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_s}. \quad (6.15)$$

由式 (6.5) 可以算出 $C_s(\psi_s)$ ，再由式 (6.3) 把横轴换成 V_G ，就得到图 7 的 C - V 曲线。几个特征值：

- **积累**： C_s 随 $|\psi_s|$ 指数增大， $C \rightarrow C_{ox}$ 。用积累区电容可以测氧化层厚度。
- **平带**：把式 (6.4) 在 $\psi \rightarrow 0$ 展开， $F^2 \approx \frac{1}{2}(1 + n_0/p_0)(\psi/U_T)^2$ ，代入式 (6.5) 求得 $C_s(0) = \varepsilon_s/L_D$ (L_D 即式 (5.7))：平带时电荷变化发生在约一个德拜长度内。标准器件 $C_{FB}/C_{ox} = 0.70$ 。
- **耗尽**： $C_s = C_d = \varepsilon_s/W_d$ 随栅压增大而减小。
- **反型**：结果取决于测量频率。若反型层电子能跟上交变电压（低频）， C_s 由反型层贡献，指数增大， C 回到 C_{ox} ；若跟不上（高频），反型电荷保持直流值，交变电荷只能在耗尽层边缘增减， $C_s \approx \varepsilon_s/W_d$ ， W_d 停在强反型时的最大值附近， C 停在最小值 $C_{min} = (1/C_{ox} + W_{d,max}/\varepsilon_s)^{-1}$ 。

为什么低频与高频不同 在一个单独的 MOS 电容里，反型层电子没有现成的来源，只能靠耗尽区中的热产生：价带电子被热激发到导带，产生电子-空穴对，电场把空穴推向体内，把电子拉到表面。耗尽区中单位体积的热产生率约为 n_i/τ_g ， τ_g 叫产生寿命，优质硅中为微秒到毫秒 [27]；单位面积的净产生电流约为 $J_g \approx qn_iW_d/\tau_g$ 。交变电压只要求反型电荷作很小的增减，但增减的速率同样受这一产生（及其逆过程复合）的限制。小信号下，反型层像一个电容 $C_{inv} = |Q_n|/2U_T$ （第 6.6 节检验框），经由产生-复合过程充放电：设反型层电子的准费米能级相对衬底偏离 δV （以电压计），耗尽区的净产生电流约为 $J_g(1 - e^{\delta V/U_T})$ （ $\delta V = 0$ 时产生与复合抵消； $\delta V > 0$ 时 $np > n_i^2$ ，复合占优），小信号下为 $-J_g\delta V/U_T$ ，相当于电导 J_g/U_T 。于是响应时间 $\tau_R \sim C_{inv}U_T/J_g = |Q_n|/2J_g$ 。刚进入强反型时 $|Q_n|$ 与 qN_AW_d 同量级，代入 J_g 得 $\tau_R \sim \frac{1}{2}(N_A/n_i)\tau_g$ ；更仔细的分析给出 $\tau_R \approx 2\tau_gN_A/n_i$ [27]，与粗估同量级。标准器件取 $\tau_g = 1 \mu s$ ， $\tau_R \approx 2 \times 10^7 \times 1 \mu s = 20 s$ 。所以即使在 1 Hz 下反型层也跟不上，常规的 1 MHz 测量给出的是高频曲线；低频曲线要用很慢的线性扫描、测量位移电流的“准静态”方法得到。若栅压扫描得太快，连直流反型层都来不及形成，半导体进入深耗尽， W_d 继续增大， C 降到 C_{min} 以下（图 7 点线）。在 MOSFET 中情形不同：重掺杂的源、漏区就在旁边，电子沿沟道流进流出，所需时间约为渡越时间 $L^2/(\mu_{eff}V)$ （取 $\mu_{eff} = 300 \text{ cm}^2/(\text{Vs})$ ， $L = 1 \mu m$ ， $V = 1 \text{ V}$ 时约 33 ps， $L = 10 \mu m$ 时约 3.3 ns；按实验题器件的 $\mu_0 = 425 \text{ cm}^2/(\text{Vs})$ ， $L = 10 \mu m$ 时约 2.4 ns），所以栅到沟道的电容在常规测量频率下也遵循低频曲线。

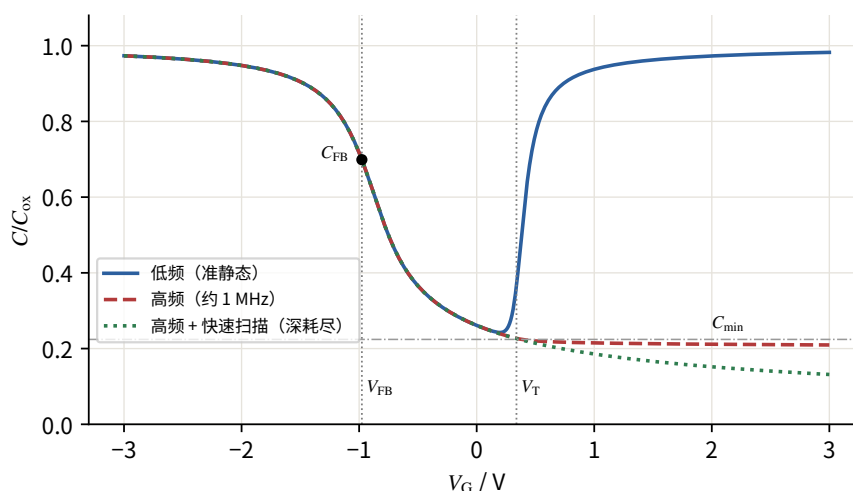


图 7 标准器件的 $C-V$ 曲线 (scripts/fig_cv.py)。低频: $C_s = -dQ_s/d\psi_s$ 由精确式 (6.5) 数值求导; 高频: 直流表面势含反型层, 小信号只由耗尽层边缘的多子响应; 深耗尽: 反型层未形成。 $C_{FB} = 0.70C_{ox}$ 与 ϵ_s/L_D 的串联值一致; 点划线为由耗尽近似在 $\psi_s = 2\phi_F$ 处算得的 $C_{min} = 0.224C_{ox}$, 强反型时表面势比 $2\phi_F$ 再高约 0.15 V, 耗尽层略宽, 高频电容实际停在 $0.21C_{ox}$ 附近。

由高频 $C-V$ 曲线可以测衬底掺杂: C_{ox} 给出 t_{ox} , C_{min} 给出 $W_{d,max} = \epsilon_s(1/C_{min} - 1/C_{ox})$, 再由 $W_{d,max} = \sqrt{2\epsilon_s \cdot 2\phi_F/(qN_A)}$ (ϕ_F 本身依赖 N_A , 迭代求解) 得 N_A 。实验题第 (1) 问就做这件事, 第 (7) 问再讨论上图注中提到的系统误差。

要点: MOS 电容小结

- 总方程 $V_G = V_{FB} + \psi_s - Q_s(\psi_s)/C_{ox}$; $Q_s(\psi_s)$ 由泊松方程第一积分精确给出 (式 (6.5))。
- 耗尽近似: $W_d = \sqrt{2\epsilon_s\psi_s/qN_A}$, $Q_d = -\sqrt{2\epsilon_sqN_A\psi_s}$, $C_d = \epsilon_s/W_d$ 。
- 强反型 $\psi_s \approx 2\phi_F$ (实际高出 0.1–0.15 V); $V_T = V_{FB} + 2\phi_F + \sqrt{4\epsilon_sqN_A\phi_F}/C_{ox}$, $Q_n \approx -C_{ox}(V_G - V_T)$ ($V_G - V_T = 0.5$ V 时把 $|Q_n|$ 高估约 35%, 外推阈值依赖拟合区间)。
- 界面陷阱与半导体并联分走栅电荷; $C-V$ 曲线由串联电容 C_{ox} 、 C_s 决定, 高低频之差来自少子的产生时间。

7 MOSFET 的电流

7.1 预备: pn 结与内建电势

MOSFET 的源、漏与衬底之间是 pn 结, 下面几节反复用到它的三个性质。

前置知识: pn 结、内建电势、正偏与反偏

- **形成:** 把 n 型区与 p 型区接在一起, n 区的电子向 p 区扩散、p 区的空穴向 n 区扩散, 在交界处留下不能移动的电离施主 (+) 与电离受主 (-), 形成没有可动载流子的空间电荷区 (耗尽区)。其中的电场阻止进一步扩散。平衡时费米能级处处相等 (第 4.1 节要点), 所以两侧能带相对错开。

- **内建电势**: 远离交界处, n 区 $E_c - E_F = k_B T \ln(N_c/N_D)$, p 区 $E_F - E_v = k_B T \ln(N_v/N_A)$ (式 (4.4)、(4.5), 设 $n = N_D$ 、 $p = N_A$)。p 区导带底比 n 区高出

$$\begin{aligned} qV_{bi} &= E_g - (E_c - E_F)_n \\ &\quad - (E_F - E_v)_p \\ &= k_B T \ln \frac{N_A N_D}{N_c N_v e^{-E_g/k_B T}} \\ &= k_B T \ln \frac{N_A N_D}{n_i^2}, \end{aligned}$$

最后一步用了式 (4.6)。 V_{bi} 叫内建电势, 等于两侧费米势的大小之和。 n^+ 源漏 ($N_D \sim 10^{20} \text{ cm}^{-3}$, 已简并, 上式只给估计) 与 $N_A = 10^{17} \text{ cm}^{-3}$ 的衬底之间 $V_{bi} \approx 1 \text{ V}$; 与不掺杂的硅之间约 $E_g/2q \approx 0.5 \text{ V}$ (n^+ 区 E_F 在导带底, 本征区在禁带中央)。

- **正偏与反偏**: 在 p 区相对 n 区加电压 V , 势垒变为 $q(V_{bi} - V)$ 。 $V > 0$ (正偏) 时势垒降低, 能越过它的多子按玻尔兹曼因子 e^{V/U_T} 增多, 电流很大; $V < 0$ (反偏) 时势垒升高, 多子过不去, 只有扩散到耗尽区边缘的少数载流子被结电场扫过去, 电流很小且几乎与电压无关。反偏时耗尽区加宽。若一侧重掺杂 (如 n^+p 结), 耗尽区几乎全在轻掺杂一侧, 宽度与 MOS 耗尽层的式 (6.9) 同形: $W = \sqrt{2\epsilon_s(V_{bi} - V)/(qN_A)}$ 。

7.2 器件结构与工作方式

在 MOS 电容的两侧各做一个重掺杂的 n^+ 区, 就得到 n 沟道 MOSFET (图 8)。两个 n^+ 区分别叫源 (S) 和漏 (D), 它们之间、栅下方的区域长 L 、宽 W , 叫沟道区; 衬底 (B, 也叫体) 通常与源相连。以源为电压参考点, 记 V_{GS} 、 V_{DS} 为栅、漏相对源的电压。

- $V_{GS} < V_T$: 源与漏之间是 p 型区, n^+ 源-p- n^+ 漏构成两个背靠背的 pn 结, 无论 V_{DS} 正负总有一个结反偏 (第 7.1 节), 几乎没有电流——开关关断。
- $V_{GS} > V_T$: 栅下表面反型为 n 型, 这层电子把源与漏连通——开关导通, 电流的大小由反型层电荷 (即栅压) 控制。

源、漏是电子的“仓库”: 重掺杂区中电子极多, 其准费米能级就是接在上面的金属电极的费米能级。取源端的电子准费米能级为能量零点, 则漏端 $E_{Fn}(\text{漏}) = E_{Fn}(\text{源}) - qV_{DS}$ 。沟道中记 $E_{Fn}(y) = -qV(y)$, $V(y)$ 从 $V(0) = 0$ 变到 $V(L) = V_{DS}$, 称为沟道电势 (电子准费米势)。

定义: MOSFET 的类型

- **n 沟道与 p 沟道**: p 型衬底、 n^+ 源漏、电子导电的叫 n 沟道 (nMOS); n 型衬底、 p^+ 源漏、空穴导电的叫 p 沟道 (pMOS), 其电压、电流符号全部相反。
- **增强型与耗尽型**: $V_{GS} = 0$ 时没有沟道、需要加栅压才导通的叫增强型 (n 沟道 $V_T > 0$); $V_{GS} = 0$ 时已有沟道 (例如在沟道区注入施主)、需加反向栅压才关断的叫耗尽型 (n 沟道 $V_T < 0$)。逻辑电路几乎只用增强型。

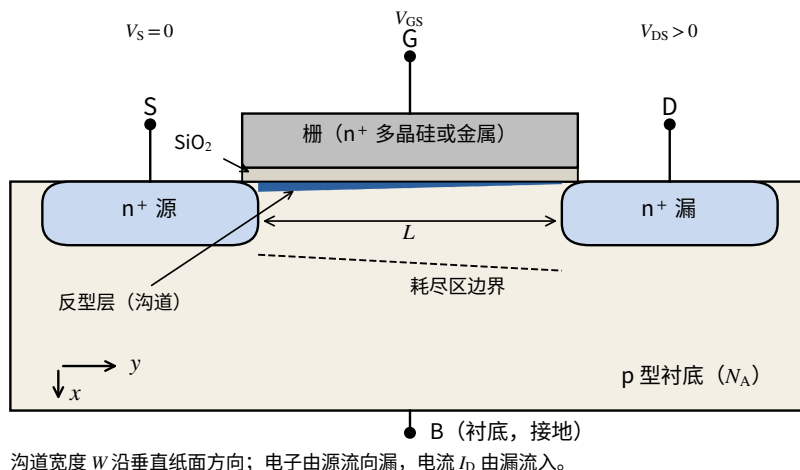


图 8 n 沟道增强型 MOSFET 的剖面 (示意, 未按比例)。坐标原点取在源端沟道起点的 Si/SiO₂ 界面上: x 垂直界面指向衬底, y 沿沟道从源指向漏。图中反型层向漏端变薄, 表示 V_{DS} 较大时漏端沟道电荷减少 (第 7.6 节); 漏端附近耗尽区加宽, 因为那里沟道电势更高。

7.3 沟道电流的一般表达式

推导: $I_D = -W\mu Q_n dV/dy$

第一步, 由第 5.4 节, 电子电流密度的 y 分量 $J_{n,y} = \mu n dE_{Fn}/dy$ 。沟道中 $E_{Fn} = -qV(y)$, 所以 $J_{n,y} = -q\mu n dV/dy$ 。

第二步, 对沟道横截面积分。设 μ 在反型层内为常数 (用有效迁移率 μ_{eff} 代表平均值), 并设电子准费米势在同一截面上不随深度 x 变化 (电子只沿 y 流动), 沿 $+y$ 方向的电流为

$$\begin{aligned} I_y &= W \int J_{n,y} dx = -q\mu W \left(\int n dx \right) \frac{dV}{dy} \\ &= \mu W Q_n \frac{dV}{dy}, \end{aligned}$$

其中 $Q_n = -q \int n dx < 0$ 是单位面积反型层电荷。

第三步, 电子从源流向漏, 电流从漏流入器件, 所以漏极电流 $I_D = -I_y$:

$$I_D = -W\mu Q_n(y) \frac{dV}{dy}. \quad (7.1)$$

第四步, 栅极不导电、沟道中没有产生与复合, 电流处处相同, I_D 与 y 无关。两边对 y 从 0 积分到 L , 并把 Q_n 看作 V 的函数 (这一步需要下一小节的渐变沟道近似), 换积分变量为 V :

$$I_D = \frac{W\mu}{L} \int_0^{V_{DS}} [-Q_n(V)] dV. \quad (7.2)$$

式 (7.2) 把问题化成了求 “沟道电势为 V 处的反型层电荷” $Q_n(V)$ 。它同时包含了漂移与扩散 (第 5.4 节), 用一维 MOS 的精确解计算 $Q_n(V)$ 再数值积分, 叫 *Pao-Sah* 模型 [29]; 本讲的数值图都以它为 “精确解”。

7.4 渐变沟道近似

前置知识：二维泊松方程

若电势依赖于 x 与 y ，在 xy 平面内取边长 dx 、 dy 的小矩形柱体（沿 z 方向长 W ），用高斯定理：穿出左右两面的电位移通量差为 $\varepsilon_s \frac{\partial \mathcal{E}_x}{\partial x} dx \cdot W dy$ ，上下两面为 $\varepsilon_s \frac{\partial \mathcal{E}_y}{\partial y} dy \cdot W dx$ ，二者之和等于柱内电荷 $\rho W dx dy$ 。用 $\mathcal{E}_x = -\partial\psi/\partial x$ 、 $\mathcal{E}_y = -\partial\psi/\partial y$ ：

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = -\frac{\rho}{\varepsilon_s}. \quad (7.3)$$

假设：渐变沟道近似（GCA）

- 沟道中横向（沿 y ）电场的变化远比纵向（沿 x ）缓慢： $|\partial^2 \psi / \partial y^2| \ll |\partial^2 \psi / \partial x^2|$ 。于是在每个 y 处，式 (7.3) 退化为第 6 节的一维泊松方程，局部的反型层电荷由局部的栅压与沟道电势决定。
- 适用条件：沟道长度远大于纵向的特征尺度， $L \gg t_{\text{ox}}, W_d$ ；并且只在离源、漏两端几个“自然长度” ℓ 以外成立（第 9.5 节）。夹断点附近不成立（第 7.6 节）。
- 衬底中的空穴没有电流，空穴准费米能级保持为衬底的 E_F ；电子准费米势在同一截面上为常数 $V(y)$ 。

局部的强反型条件 在沟道中 y 处，电子的准费米势为 V ，于是 $n = n_0 e^{(\psi-V)/U_T}$ （把式 (5.5) 中的 E_F 换成 E_{Fn} ）；空穴仍是 $p = p_0 e^{-\psi/U_T}$ 。表面电子浓度达到 N_A 需要 $\psi_s = 2\phi_F + V$ ：沟道电势每升高 1 V，能带就要多弯 1 V 才能反型。重复第 6.6 节的推导（栅相对衬底的电压是 V_{GS} ）：

$$\begin{aligned} Q_n(V) = & -C_{\text{ox}} [V_{GS} - V_{FB} - (2\phi_F + V) \\ & - \gamma \sqrt{2\phi_F + V}], \end{aligned} \quad (7.4)$$

$$\gamma \equiv \frac{\sqrt{2\varepsilon_s q N_A}}{C_{\text{ox}}}.$$

方括号中从栅压里依次扣除：平带电压、表面势 $2\phi_F + V$ 、支撑耗尽层电荷 $-\sqrt{2\varepsilon_s q N_A (2\phi_F + V)}$ 所需的氧化层电压；剩下的部分才用来感应反型电子。

7.5 线性区与饱和区：最简单的模型

下面的公式中 V_T 应理解为按某一区间外推得到的阈值 V_T^* （第 6.6 节检验框），否则会把电流高估 20%–40%；图 9 就是这样比较的。先忽略耗尽层电荷随 V 的变化（把式 (7.4) 中的 $\gamma \sqrt{2\phi_F + V}$ 换成 $\gamma \sqrt{2\phi_F}$ ），于是 $Q_n(V) = -C_{\text{ox}}(V_{GS} - V_T - V)$ ：沟道 y 处的“有效栅压”是栅与沟道之间的电压 $V_{GS} - V(y)$ 。代入式 (7.2) 积分：

$$\begin{aligned} I_D = \mu C_{\text{ox}} \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \\ (0 \leq V_{DS} \leq V_{GS} - V_T). \end{aligned} \quad (7.5)$$

- **线性区** ($V_{DS} \ll V_{GS} - V_T$)： $I_D \approx \mu C_{\text{ox}} (W/L) (V_{GS} - V_T) V_{DS}$ 。MOSFET 是一个阻值受栅压控制的电阻，电导等于反型层的薄层电导 $\mu|Q_n|$ 乘 W/L 。

- **饱和：**式 (7.5) 是 V_{DS} 的开口向下的抛物线，在

$$\begin{aligned} V_{DS,sat} &= V_{GS} - V_T, \\ I_{D,sat} &= \frac{\mu C_{ox} W}{2L} (V_{GS} - V_T)^2 \end{aligned} \quad (7.6)$$

处取极大。此时 $Q_n(L) = -C_{ox}(V_{GS} - V_T - V_{DS,sat}) = 0$ ：漏端的反型层消失，称为夹断。 $V_{DS} > V_{DS,sat}$ 时式 (7.5) 给出电流减小，这是错误的——它在漏端给出 $Q_n > 0$ ，已超出模型的适用范围。实际电流基本保持 $I_{D,sat}$ ，这就是 MOSFET 的饱和区， $I_{D,sat} \propto (V_{GS} - V_T)^2$ 叫平方律。

检验：量纲与极限

μC_{ox} 的单位是 $\text{m}^2/(\text{Vs}) \cdot \text{F}/\text{m}^2 = \text{A}/\text{V}^2$ ，乘 V^2 得 A。 $V_{DS} \rightarrow 0$ 时 $I_D \rightarrow 0$ ； $V_{GS} \rightarrow V_T$ 时 $I_{D,sat} \rightarrow 0$ 。标准器件取 $\mu_{\text{eff}} = 300 \text{ cm}^2/(\text{Vs})$ 、 $W/L = 10$ ： $\mu C_{ox} W/L = 1.04 \times 10^{-3} \text{ A}/\text{V}^2$ ， $V_{GS} - V_T = 1 \text{ V}$ 时 $I_{D,sat} = 0.52 \text{ mA}$ 。

7.6 夹断的物理图像：电流为何不再增加

推导：沟道中的电势与电荷分布

由式 (7.1)， $I_D dy = \mu W C_{ox} (V_{GS} - V_T - V) dV$ 。记 $V_{ov} \equiv V_{GS} - V_T$ （过驱动电压），从源 ($y = 0$, $V = 0$) 积分到 y ，再用式 (7.5) 消去 I_D ：

$$V_{ov} V - \frac{V^2}{2} = \frac{y}{L} \left(V_{ov} V_{DS} - \frac{V_{DS}^2}{2} \right).$$

在夹断 ($V_{DS} = V_{ov}$) 时解得 $V(y) = V_{ov} (1 - \sqrt{1 - y/L})$ ，于是

$$\begin{aligned} Q_n(y) &= -C_{ox} V_{ov} \sqrt{1 - \frac{y}{L}}, \\ \mathcal{E}_y &= -\frac{dV}{dy} = -\frac{V_{ov}}{2L \sqrt{1 - y/L}}. \end{aligned} \quad (7.7)$$

越靠近漏端，电子越少，横向电场越强，二者之积（正比于电流）保持不变；在 $y \rightarrow L$ 处 $Q_n \rightarrow 0$ 、 $|\mathcal{E}_y| \rightarrow \infty$ （图 9(b)）。“电子为零而电流不为零”的矛盾说明 GCA 在漏端失效：当 $|\mathcal{E}_y|$ 大到与纵向电场可比时， $\partial^2 \psi / \partial y^2$ 不能再忽略。实际情形是：在漏端附近一段长 ΔL 的“夹断区”里，电子离开界面、密度很低，但被很强的二维电场以很高的速度（不超过饱和速度，第 7.9 节）扫进漏区，电荷并不为零。

为什么电流饱和 V_{DS} 超过 $V_{DS,sat}$ 后，多出的电压 $V_{DS} - V_{DS,sat}$ 几乎全部降在很短的夹断区上；从源到夹断点的沟道仍然承受电压 $V_{DS,sat}$ ，长度仍约为 L ，因而送到夹断点的电流仍约为 $I_{D,sat}$ 。到达夹断点的电子会被强电场扫进漏区（就像反偏 pn 结收集扩散到结边缘的少数，第 7.1 节），再提高漏压只会让它们更快地落下去，不会让更多电子到来。这好比瀑布：落差再大，流量也由上游的河道决定。

7.7 体电荷修正与体效应

推导：体因子 m

保留式 (7.4) 中耗尽层电荷随 V 的变化。把 $\sqrt{2\phi_F + V}$ 在 $V = 0$ 附近展开到一阶（小参量 $V/2\phi_F$ ）：

$$\gamma\sqrt{2\phi_F + V} \approx \gamma\sqrt{2\phi_F} + \frac{\gamma}{2\sqrt{2\phi_F}}V.$$

代回式 (7.4)： $Q_n(V) \approx -C_{ox}(V_{GS} - V_T - mV)$ ，其中

$$m \equiv 1 + \frac{\gamma}{2\sqrt{2\phi_F}} = 1 + \frac{C_{dm}}{C_{ox}},$$

$$C_{dm} = \sqrt{\frac{\varepsilon_s q N_A}{2 \cdot 2\phi_F}} = \frac{\varepsilon_s}{W_{d,max}}. \quad (7.8)$$

（第二个等号： $\gamma/2\sqrt{2\phi_F} = \sqrt{2\varepsilon_s q N_A}/(2C_{ox}\sqrt{2\phi_F})$ ，与式 (6.10) 在 $\psi_s = 2\phi_F$ 处的值相同。）代入式 (7.2)：

$$I_D = \mu C_{ox} \frac{W}{L} \left[V_{ov} V_{DS} - \frac{m}{2} V_{DS}^2 \right],$$

$$V_{DS,sat} = \frac{V_{GS} - V_T}{m}, \quad (7.9)$$

$$I_{D,sat} = \frac{\mu C_{ox} W}{2mL} (V_{GS} - V_T)^2.$$

m 的物理意义：沟道电势升高 dV 时，耗尽层加深，多吸收了 $C_{dm}dV$ 的栅电荷，反型层电荷因此减少得比 $C_{ox}dV$ 更快。标准器件 $m = 1.29$ ，最简单的模型把饱和电流高估了 29%。不作线性化而直接积分式 (7.4) 也可以（结果含 $(2\phi_F + V_{DS})^{3/2}$ 项），对标准器件与式 (7.9) 相差不到 5%（scripts/fig_iv.py）。

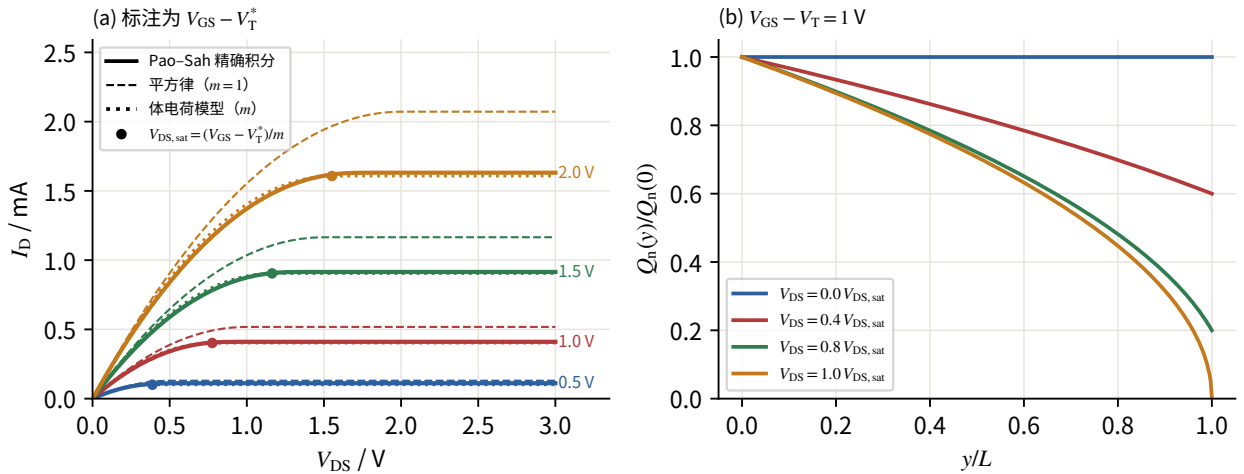


图 9 标准器件 ($W/L = 10$, $\mu_{eff} = 300 \text{ cm}^2/(\text{Vs})$) 的输出特性。(a) Pao-Sah 精确积分 (式 (7.2), Q_n 由一维 MOS 精确解给出) 与平方律 (7.5)、体电荷模型 (7.9) 的比较；为了只比较 m 的作用，后两者都用外推阈值 $V_T^* = 0.466 \text{ V}$ (图 6(c), $V_G = 1.2\text{--}2.0 \text{ V}$ 拟合)。 $V_{GS} - V_T^* \geq 1 \text{ V}$ 时体电荷模型与精确积分相差约 2%；近阈值时精确解更大，多出的是扩散电流。(b) 体电荷模型中沟道电荷沿沟道的分布： $V_{DS} \rightarrow V_{DS,sat}$ 时漏端夹断。

体效应 若衬底相对源加反偏 $V_{SB} > 0$ (例如多个晶体管串联、源极不接地时), 源端的沟道电势相对衬底已是 V_{SB} , 阈值条件变为 $\psi_s = 2\phi_F + V_{SB}$ 。栅相对衬底的电压为 $V_{GS} + V_{SB}$, 由式 (7.4) 令 $Q_n = 0$ 、 $V = V_{SB}$, 再减去 V_{SB} 换回相对源的栅压:

$$\begin{aligned} V_T(V_{SB}) &= V_{FB} + 2\phi_F \\ &\quad + \gamma\sqrt{2\phi_F + V_{SB}}, \\ V_T - V_{T0} &= \gamma(\sqrt{2\phi_F + V_{SB}} \\ &\quad - \sqrt{2\phi_F}). \end{aligned} \quad (7.10)$$

衬底像第二个栅 (“背栅”), 通过耗尽层电荷影响阈值。标准器件 $\gamma = 0.528 \text{ V}^{1/2}$, $V_{SB} = 1 \text{ V}$ 时阈值升高 0.23 V 。

7.8 跨导、输出电阻与沟道长度调制

定义：跨导与输出电阻

跨导 $g_m \equiv \partial I_D / \partial V_{GS}$ (V_{DS} 固定), 衡量栅压控制电流的能力; 输出电阻 $r_o \equiv (\partial I_D / \partial V_{DS})^{-1}$ (V_{GS} 固定)。用作放大器时, 小信号电压增益的上限 (本征增益) 为 $g_m r_o$ 。

由式 (7.9), 饱和区 $g_m = \mu C_{ox} W (V_{GS} - V_T) / (mL) = 2I_{D,sat} / (V_{GS} - V_T)$; 线性区 $g_m = \mu C_{ox} (W/L) V_{DS}$ 。若饱和区电流严格不随 V_{DS} 变化, $r_o = \infty$ 。实际上 V_{DS} 增大时夹断点向源移动, 有效沟道长度缩短为 $L - \Delta L$, 电流略有增加:

$$\begin{aligned} I_D &= I_{D,sat} \frac{L}{L - \Delta L} \\ &\approx I_{D,sat} \left(1 + \frac{\Delta L}{L} \right), \end{aligned} \quad (7.11)$$

这叫沟道长度调制。

定义：沟道长度调制系数与厄利电压

饱和区中把 I_D - V_{DS} 曲线近似为直线 $I_D \approx I_{D0}(1 + \lambda V_{DS})$, I_{D0} 是这条直线外推到 $V_{DS} = 0$ 的截距。于是

$$\lambda \equiv \frac{1}{I_{D0}} \frac{\partial I_D}{\partial V_{DS}}, \quad r_o = \frac{1}{\lambda I_{D0}},$$

$V_A \equiv 1/\lambda$ 叫厄利电压 (直线与 V_{DS} 轴交于 $-V_A$)。 $\lambda V_{DS} \ll 1$ 时 $I_{D0} \approx I_D \approx I_{D,sat}$, $\lambda \approx I_D^{-1} \partial I_D / \partial V_{DS}$ 。全讲 (含演示与实验题) 都用这一个定义。

假设：夹断区长度的粗略模型

把夹断区当作漏端 n^+p 结的一部分、一维的单边突变耗尽层 (第 7.1 节): 它承受电压 $V_{DS} - V_{DS,sat}$, 掺杂为 N_A , 由式 (6.9), $\Delta L \approx \sqrt{2\epsilon_s(V_{DS} - V_{DS,sat})/qN_A}$ 。这一模型忽略了栅对夹断区电场的影响与可动电子的电荷, 只给出量级 (通常高估 ΔL)。

由式 (7.11) 求导, $\lambda \approx \frac{1}{L} \frac{d\Delta L}{dV_{DS}} = \frac{\Delta L}{2L(V_{DS} - V_{DS,sat})}$ (注意不是割线 $\Delta L/[L(V_{DS} - V_{DS,sat})]$: $\Delta L \propto \sqrt{V_{DS} - V_{DS,sat}}$, 二者差 2 倍)。按此模型 λ 随工作点变化, 所以实际中 λ 是在给定工

作点附近由测量拟合的唯象参数。标准器件、 $L = 1\text{ }\mu\text{m}$ 、 $V_{\text{DS}} - V_{\text{DS,sat}} = 1\text{ V}$ 时 $\Delta L \approx 0.11\text{ }\mu\text{m}$, $\lambda \approx 0.06\text{ V}^{-1}$ 。由此

$$\begin{aligned} r_o &= \frac{1}{\lambda I_{\text{D0}}} \approx \frac{1}{\lambda I_{\text{D,sat}}}, \\ g_m r_o &\approx \frac{2}{\lambda(V_{\text{GS}} - V_{\text{T}})} \propto L. \end{aligned} \quad (7.12)$$

沟道越短, $\lambda \propto 1/L$ 越大, 本征增益越低——这是模拟电路设计中短沟道器件的一个代价 (例题 3)。

交互演示: 标准器件的输出特性族 $I_{\text{D}}-V_{\text{DS}}$ (阈值取外推值 $V_{\text{T}}^* = 0.466\text{ V}$)。可选长沟道电荷平方律或计入速度饱和的模型, 调节沟道长度 L 、体因子 m 与沟道长度调制系数 λ , 读出夹断电压与饱和电流; 下方小图给出所选工作点沿沟道的电荷分布 $Q_{\text{n}}(y)$, 可以看到漏端夹断。

网页版中此处为可调参数的交互动画 (演示编号 `mosfet-iv`)。

7.9 短沟道中的速度饱和

假设: 漂移速度模型

硅中电子的漂移速度在弱场下为 $\mu\mathcal{E}$, 在强场下趋于饱和值 $v_{\text{sat}} \approx 1 \times 10^7\text{ cm/s}$ [26]: 电子从电场获得的能量一旦接近光学声子的能量 (约 60 meV), 就会迅速发射光学声子而失去能量。(光学声子是晶格中相邻原子相对反向振动的振动模式的能量量子; 电子发射一个光学声子就把约 60 meV 交给晶格。)本讲用一个便于积分的插值式

$$\begin{aligned} v(\mathcal{E}) &= \frac{\mu\mathcal{E}}{1 + \mathcal{E}/\mathcal{E}_{\text{sat}}}, \\ \mathcal{E}_{\text{sat}} &\equiv \frac{v_{\text{sat}}}{\mu}. \end{aligned} \quad (7.13)$$

它在 $\mathcal{E} \ll \mathcal{E}_{\text{sat}}$ 与 $\mathcal{E} \gg \mathcal{E}_{\text{sat}}$ 两个极限正确, 中间区域只是近似 (实测曲线拐得更急)。反型层中取 $\mu_{\text{eff}} = 300\text{ cm}^2/(\text{V s})$, $\mathcal{E}_{\text{sat}} = 3.3 \times 10^4\text{ V/cm} = 3.3\text{ V}/\mu\text{m}$ 。

当 $L = 1\text{ }\mu\text{m}$ 、 $V_{\text{DS}} \sim 1\text{ V}$ 时, 沟道中的平均电场已达 $1\text{ V}/\mu\text{m}$, 与 $\mathcal{E}_{\text{sat}}$ 可比, 速度饱和不能忽略。

推导: 计入速度饱和的 $I-V$

第一步, 电流 = 宽度 $\times$ 面电荷 $\times$ 速度: $I_{\text{D}} = W(-Q_{\text{n}})v$, v 由式 (7.13) 给出, 其中 $\mathcal{E} = \text{d}V/\text{d}y$ (取大小)。

第二步, 去分母: $I_{\text{D}} \left(1 + \frac{1}{\mathcal{E}_{\text{sat}}} \frac{\text{d}V}{\text{d}y}\right) = W\mu(-Q_{\text{n}}) \frac{\text{d}V}{\text{d}y}$ 。

第三步, 两边对 y 从 0 积分到 L (I_{D} 为常数), 右边换元为对 V 积分, 用 $Q_{\text{n}} = -C_{\text{ox}}(V_{\text{ov}} - mV)$:

$$I_{\text{D}} \left(L + \frac{V_{\text{DS}}}{\mathcal{E}_{\text{sat}}}\right) = W\mu C_{\text{ox}} \left(V_{\text{ov}} V_{\text{DS}} - \frac{m}{2} V_{\text{DS}}^2\right),$$

即

$$I_D = \frac{\mu C_{ox} W}{L} \times (V_{ov} V_{DS} - m V_{DS}^2 / 2) \times \left(1 + \frac{V_{DS}}{\mathcal{E}_{sat} L}\right)^{-1}. \quad (7.14)$$

速度饱和的全部效果是一个分母：长沟道电流除以 $1 + V_{DS}/(\mathcal{E}_{sat} L)$ ，后者是“平均电场与临界场之比”。

第四步，饱和点。 I_D 对 V_{DS} 的极大值（附录 A.2）在

$$\begin{aligned} V_{DS,sat} &= \mathcal{E}_{sat} L (\sqrt{1+u} - 1), \\ u &\equiv \frac{2V_{ov}}{m\mathcal{E}_{sat} L}, \\ I_{D,sat} &= W C_{ox} v_{sat} \times (V_{ov} - m V_{DS,sat}). \end{aligned} \quad (7.15)$$

最后一式说明：在饱和点，漏端的电荷 $C_{ox}(V_{ov} - m V_{DS,sat})$ 恰以 v_{sat} 运动——饱和的物理意义从“漏端电荷为零”变成了“漏端速度达到饱和”。

检验：两个极限

- 长沟道 ($\mathcal{E}_{sat} L \gg V_{ov}$): 把根式展开到二阶, $V_{DS,sat} \rightarrow V_{ov}/m$, $I_{D,sat} \rightarrow \mu C_{ox} W V_{ov}^2 / (2mL)$, 回到式 (7.9)。
- 短沟道 ($\mathcal{E}_{sat} L \ll V_{ov}$): $V_{DS,sat} \approx \sqrt{2V_{ov}\mathcal{E}_{sat} L/m} \ll V_{ov}$,

$$I_{D,sat} \rightarrow W C_{ox} v_{sat} (V_{GS} - V_T), \quad (7.16)$$

电流与 L 无关、与过驱动电压成线性关系：全部反型电荷 $C_{ox} V_{ov}$ 都以饱和速度运动。跨导 $g_m \rightarrow W C_{ox} v_{sat}$ 也不再随栅压增加。

- 过渡长度：令 $L_c \equiv V_{ov}/(m\mathcal{E}_{sat})$ ，代入得 $I_{D,sat}/I_{D,sat}^{(\text{平方律})} = 2(2 - \sqrt{3}) = 0.536$ 。标准器件 $V_{ov} = 1\text{ V}$ 时 $L_c = 0.23\text{ }\mu\text{m}$ ；即使 $L = 1\text{ }\mu\text{m}$ ，饱和电流也已比平方律低 18%（图 10）。

注：更短的沟道：准弹道运输

今天的沟道长度只有 10–20 nm，比电子的平均自由程（几十纳米）还短，电子在沟道中几乎不散射（“弹道”运动），漂移速度模型本身失效。此时电流由源端势垒顶的电子注入速度决定， $I_D \approx W C_{ox} v_{inj} (V_{GS} - V_T)$ ， v_{inj} 与热运动速度同量级 [28]——形式与式 (7.16) 相同，但物理来源不同。

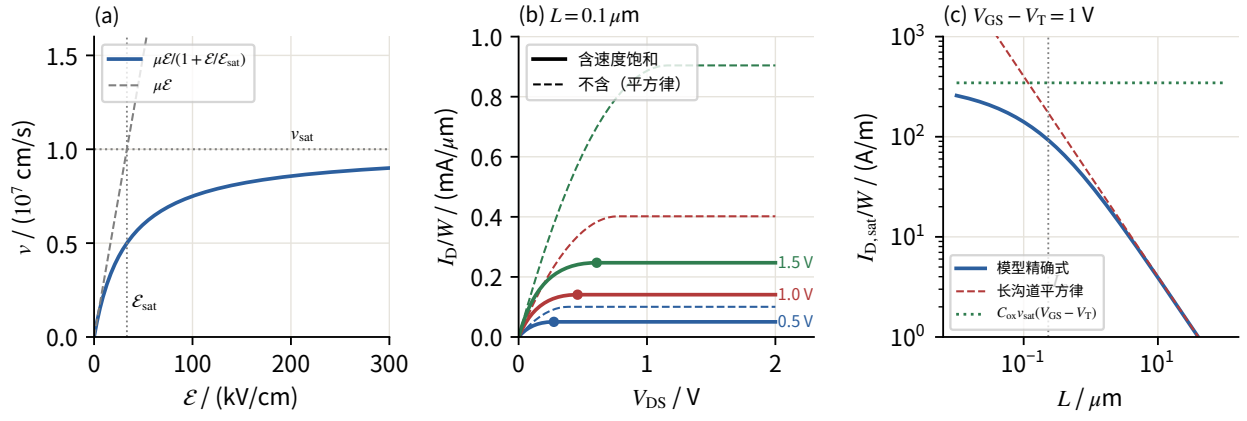


图 10 速度饱和（标准器件， $\mu_{\text{eff}} = 300 \text{ cm}^2/(\text{Vs})$ ， $v_{\text{sat}} = 1 \times 10^7 \text{ cm/s}$ ， $m = 1.29$ ）。(a) 漂移速度模型 (7.13)。 (b) $L = 0.1 \mu\text{m}$ 的输出特性：计入速度饱和（实线，式 (7.14)，圆点为 $V_{\text{DS,sat}}$ ）与平方律（虚线）；标注为 $V_{\text{GS}} - V_{\text{T}}$ ，实线的饱和电流近似等间距——线性律。 (c) $V_{\text{GS}} - V_{\text{T}} = 1 \text{ V}$ 时饱和电流随 L 的变化：式 (7.15) 从长沟道平方律过渡到短沟道极限 (7.16)，竖线为 L_{c} 。脚本 `scripts/fig_vsats.py` 用 `assert` 核对了两个极限与“漏端速度 = v_{sat} ”。

8 亚阈值区与玻尔兹曼极限

$V_{\text{GS}} < V_{\text{T}}$ 时晶体管并没有完全关断：表面处于弱反型，仍有少量电子，电流随栅压指数变化。这个“亚阈值电流”决定了晶体管关断时的漏电，是今天芯片功耗问题的核心。

8.1 弱反型时的沟道电荷

推导： $Q_{\text{n}} \propto e^{(\psi_{\text{s}} - V)/U_{\text{T}}}$

第一步，弱反型时 ($\phi_{\text{F}} < \psi_{\text{s}} < 2\phi_{\text{F}}$) 反型电子很少，栅压方程中可以忽略它们，表面势由耗尽区的电荷决定。这里用第一积分给出的式 (6.7)（比耗尽近似多一个 $-U_{\text{T}}$ ）： $V_{\text{GS}} = V_{\text{FB}} + \psi_{\text{s}} + \gamma\sqrt{\psi_{\text{s}} - U_{\text{T}}}$ （式 (6.3)）。 ψ_{s} 只由栅压决定，与沟道电势 V 无关，因而沿整个沟道相同。

第二步，表面附近 $\psi(x) \approx \psi_{\text{s}} - \mathcal{E}_{\text{s}}x$ ，电子浓度 $n = n_0 e^{(\psi - V)/U_{\text{T}}} = n_{\text{s}} e^{-\mathcal{E}_{\text{s}}x/U_{\text{T}}}$ ，其中 $n_{\text{s}} = n_0 e^{(\psi_{\text{s}} - V)/U_{\text{T}}}$ 是表面电子浓度。对 x 积分，反型层的有效厚度为 $U_{\text{T}}/\mathcal{E}_{\text{s}}$ ：

$$Q_{\text{n}} = -q \int_0^{\infty} n \, dx = -qn_{\text{s}} \frac{U_{\text{T}}}{\mathcal{E}_{\text{s}}}.$$

第三步，由式 (6.7)， $\mathcal{E}_{\text{s}} = |Q_{\text{d}}|/\epsilon_{\text{s}} = \sqrt{2qN_{\text{A}}(\psi_{\text{s}} - U_{\text{T}})/\epsilon_{\text{s}}}$ ；又 $n_0 = n_{\text{i}}^2/N_{\text{A}}$ 。整理后出现的组合正是对式 (6.7) 求导得到的耗尽层电容 $C_{\text{d}} \equiv -dQ_{\text{d}}/d\psi_{\text{s}}$ （它比式 (6.10) 多一个 $-U_{\text{T}}$ ，本节与第 8.3 节都用这一定义）：

$$\begin{aligned} Q_{\text{n}} &= -C_{\text{d}}(\psi_{\text{s}}) U_{\text{T}} \left(\frac{n_{\text{i}}}{N_{\text{A}}} \right)^2 \\ &\quad \times e^{(\psi_{\text{s}} - V)/U_{\text{T}}}, \\ C_{\text{d}} &= \sqrt{\frac{\epsilon_{\text{s}} q N_{\text{A}}}{2(\psi_{\text{s}} - U_{\text{T}})}}. \end{aligned} \tag{8.1}$$

第二步只保留了电势的一次项。附录 B.2 计入电势的二次项，结果是式 (8.1) 再乘以 $1 +$

$U_T/[2(\psi_s - U_T)]$ (式 (B.3)): 式 (8.1) 比精确值偏小 2%–3%, 对指数规律没有影响。这是本讲弱反型电荷的唯一写法; 下文的 C_d 都指式 (8.1) 中的定义。

8.2 亚阈值电流是扩散电流

把式 (8.1) 代入式 (7.2), ψ_s 与 V 无关, 积分只作用在 e^{-V/U_T} 上:

$$I_D = \mu \frac{W}{L} C_d U_T^2 \left(\frac{n_i}{N_A} \right)^2 \times e^{\psi_s/U_T} (1 - e^{-V_{DS}/U_T}). \quad (8.2)$$

- $V_{DS} \gtrsim 3U_T \approx 80 \text{ mV}$ 后电流与 V_{DS} 无关 (长沟道器件)。
- 这是扩散电流: ψ_s 沿沟道不变, 没有横向电场; 源端的面电子密度 $N(0) = |Q_n(0)|/q$, 漏端 $N(L) = N(0)e^{-V_{DS}/U_T}$, 二者之间线性分布 (扩散方程在电流不变时的解), 扩散电流 $I_D = qDW[N(0) - N(L)]/L$ 。用爱因斯坦关系 $D = \mu U_T$, 正好得到式 (8.2)。同一个公式 (7.2) 在强反型时给出漂移电流、在弱反型时给出扩散电流。

8.3 亚阈值摆幅与 60 mV/dec

定义: 亚阈值摆幅

亚阈值区中, 使漏极电流改变 10 倍所需的栅压变化: $S \equiv \left(\frac{d \log_{10} I_D}{dV_{GS}} \right)^{-1}$, 单位 mV/dec (毫伏每十倍)。S 越小, 晶体管从“关”到“开”越陡。

推导: $S = \ln 10 \cdot m U_T$

第一步, 式 (8.2) 中对栅压最敏感的是指数因子, $C_d \propto (\psi_s - U_T)^{-1/2}$ 的变化较慢, 先忽略: $d \ln I_D / d\psi_s \approx 1/U_T$ 。

第二步, 由电容分压式 (6.14) (弱反型时 $C_s \approx C_d$): $d\psi_s / dV_{GS} = 1/m$,

$$m = 1 + \frac{C_d + C_{it}}{C_{ox}}. \quad (8.3)$$

第三步, 链式法则: $d \ln I_D / dV_{GS} = 1/(m U_T)$, 换成以 10 为底的对数:

$$\begin{aligned} S &= \ln 10 \cdot U_T \left(1 + \frac{C_d + C_{it}}{C_{ox}} \right) \\ &\geq \ln 10 \cdot \frac{k_B T}{q} \\ &= 59.5 \text{ mV/dec (300 K)} \end{aligned} \quad (8.4)$$

检验: 与精确解比较

标准器件在 $V_{GS} = 0$ 处 $\psi_s = 0.583 \text{ V}$, $C_d/C_{ox} = 0.354$ (按式 (8.1) 的定义), 式 (8.4) 给出 $S = 80.6 \text{ mV/dec}$; Pao–Sah 精确积分给出 82.6 mV/dec (图 11(a))。2.5% 的差别来自第一步略去的 $C_d(\psi_s)$: 计入它, $d \ln I_D / d\psi_s = \frac{1}{U_T} - \frac{1}{2(\psi_s - U_T)}$, S 要除以 $1 - U_T/[2(\psi_s - U_T)] = 0.977$, 得 82.5 mV/dec , 与精确值相差 0.1%。注意 m 依赖于 ψ_s : 在 $\psi_s = 2\phi_F$ 处 $m = 1.29$ (式 (7.8)),

在亚阈值区中段 $m \approx 1.35$ 。

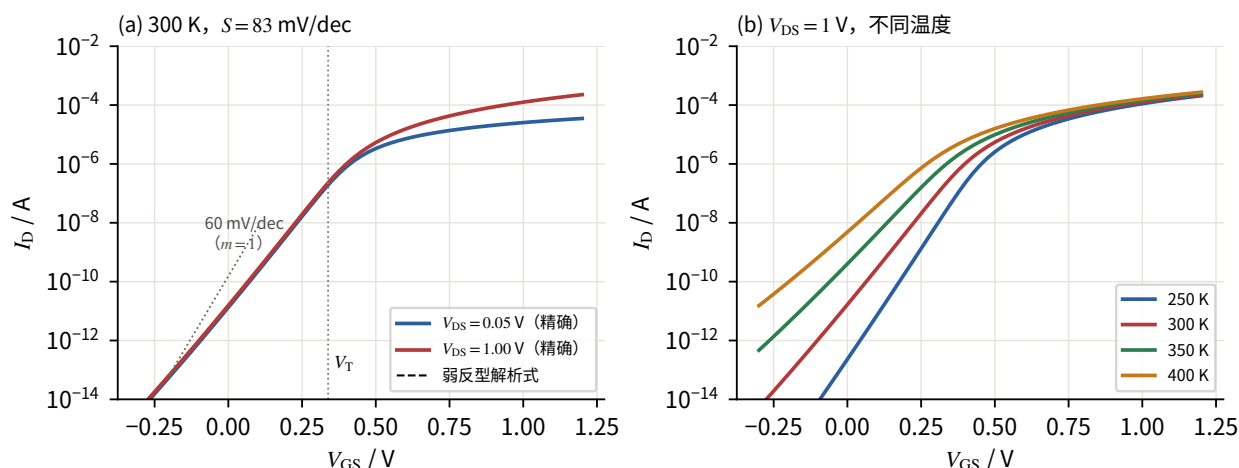


图 11 标准器件 ($W/L = 10$) 的转移特性, 对数坐标 (`scripts/fig_subth.py`)。 (a) 300 K: Pao-Sah 精确积分 (实线) 与弱反型解析式 (8.2) (虚线, 偏低 2%~4%; 乘以附录式 (B.3) 的修正因子后相差 < 1%); 两种 V_{DS} 的曲线在亚阈值区只差因子 $1 - e^{-V_{DS}/U_T}$ ($V_{DS} = 0.05$ V 时为 0.86), 对数坐标上几乎重合。点线为 $m = 1$ 时的理想斜率 60 mV/dec。 (b) 不同温度 ($V_{DS} = 1$ V, $n_i(T)$ 与 $E_g(T)$ 按文献公式 [26]): S 从 250 K 的 68 mV/dec 增大到 400 K 的 118 mV/dec; 阈值电压以约 -1.0 mV/K 降低 (n^+ 栅的功函数差也计入 $E_g(T)$, 方法同例题 2)。

玻尔兹曼极限的物理图像 从源到漏看, 沟道中的电子势能 (导带底) 在弱反型时比源区高出一个势垒。源区的电子按玻尔兹曼分布占据导带, 能量高于势垒顶的电子比例 $\propto e^{-E_b/k_B T}$, 只有它们能进入沟道——亚阈值电流本质上是越过势垒的热发射。栅压通过电容分压改变势垒高度, $dE_b = -q dV_{GS}/m$, 而 $m \geq 1$ (正电容串联分压, 分到的电压不会多于总电压)。因此电流随栅压的变化不可能快于 $e^{qV_{GS}/k_B T}$, 即每 $\ln 10 \cdot k_B T/q \approx 60$ mV 一个数量级。这个极限只来自“载流子按玻尔兹曼分布、靠热运动越过势垒”这一点, 与材料无关。降低温度可以减小 S (图 11(b)), 但芯片工作在室温以上。

交互演示: 转移特性 I_D - V_{GS} 的线性与对数坐标。 调节温度 T (阈值按约 -1.0 mV/K 随之变化, 与亚阈值特性图 (b) 一致)、耗尽层电容与界面陷阱电容之和与栅氧电容之比 $(C_d + C_{it})/C_{ox}$ 以及 300 K 时的阈值电压 (插值模型见附录 A), 读出亚阈值摆幅 S 、关断电流与开关电流比; 可与 60 mV/dec 的理想斜率比较。

网页版中此处为可调参数的交互动画 (演示编号 `mosfet-transfer`)。

8.4 关断电流: 电源电压与阈值电压的两难

把阈值处的电流记为 I_T , 亚阈值区中 $I_D \approx I_T 10^{(V_{GS}-V_T)/S}$, 于是 $V_{GS} = 0$ 时的关断电流

$$I_{off} \approx I_T 10^{-V_T/S}. \quad (8.5)$$

标准器件: $I_T = 2.4 \times 10^{-7}$ A, $V_T/S = 0.338/0.083 = 4.1$, $I_{off} \approx 2 \times 10^{-11}$ A (精确积分为 1.6×10^{-11} A)。

要点：为什么电源电压降不下去

- 开态电流 I_{on} 由过驱动电压 $V_{DD} - V_T$ 决定（平方律或线性律），要保持速度， $V_{DD} - V_T$ 不能太小；动态功耗 $\propto V_{DD}^2$ （第 9.2 节），要省电， V_{DD} 必须小。
- 两者同时满足就要降低 V_T ，而由式 (8.5)， V_T 每降低一个 S ($\geq 60 \text{ mV}$)，关断电流就增大 10 倍。一块有 10^9 个晶体管的芯片，每个关断时漏电 1 nA，静态电流就是 1 A。
- S 的下限 $\ln 10 \cdot k_B T / q$ 不随器件尺寸缩小，而且温度升高时 S 变大、 V_T 降低（例题 2），漏电更严重。这是第 9.4 节 Dennard 缩小失效的根源。

注：突破 60 mV/dec 的思路（定性）

- **隧穿场效应晶体管 (TFET)**：源区与沟道掺杂类型相反，载流子不是越过势垒，而是由栅压控制的带间隧穿从价带穿到导带。隧穿窗口截断了源区分布的高能“玻尔兹曼尾巴”，原理上 S 可以小于 60 mV/dec；难点是隧穿电流太小，开态电流不足 [30]。
- **负电容场效应晶体管**：在栅介质中加入一层铁电材料。铁电体在一定条件下表现为“负电容” $C_{fe} < 0$ ，与 C_{ox} 串联时可以使半导体表面分到的电压大于栅压的变化，相当于 $m < 1$ [31]。难点是稳定性与回滞。

两种思路分别改变了推导中的两个环节：前者改变“热发射”，后者改变“正电容分压”。

9 CMOS 与器件缩小

9.1 CMOS 反相器

把一只 nMOS 与一只 pMOS 串联在电源 V_{DD} 与地之间：nMOS 的源接地，pMOS 的源接 V_{DD} ，两管的栅连在一起作输入 V_{in} ，两管的漏连在一起作输出 V_{out} ，输出端接下一级的栅（负载电容 C_L ）。这就是 CMOS 反相器（互补 MOS）。

- $V_{in} = 0$ ：nMOS 的 $V_{GS} = 0 < V_{Tn}$ ，关断；pMOS 的 $V_{GS} = -V_{DD} < V_{Tp}$ ($V_{Tp} < 0$)，导通，输出被拉到 V_{DD} 。
- $V_{in} = V_{DD}$ ：nMOS 导通、pMOS 关断，输出被拉到 0。
- 两种稳态下总有一只管子关断，电源到地之间没有直流通路，静态功耗只来自关断管的漏电 $I_{off}V_{DD}$ 。这就是 Wanlass 与 Sah “纳瓦逻辑”的含义 [15]。

推导：翻转电压与噪声容限（平方律、 $\lambda = 0$ ）

记 $k_n = \mu_n C_{ox} W_n / L_n$ ， $k_p = \mu_p C_{ox} W_p / L_p$ ，取 $m = 1$ 。两管串联，电流相等： $I_{Dn}(V_{in}, V_{out}) = I_{Dp}(V_{DD} - V_{in}, V_{DD} - V_{out})$ （pMOS 用绝对值）。

第一步，翻转电压 V_M 定义为 $V_{out} = V_{in}$ 的点。此时两管都饱和 ($V_{DS} = V_M > V_M - V_{Tn}$)，令

饱和电流相等: $\frac{k_n}{2}(V_M - V_{Tn})^2 = \frac{k_p}{2}(V_{DD} - V_M - |V_{Tp}|)^2$, 取正根:

$$V_M = \frac{V_{Tn} + r(V_{DD} - |V_{Tp}|)}{1 + r}, \quad (9.1)$$

$$r = \sqrt{k_p/k_n}.$$

对称设计 ($r = 1$, $V_{Tn} = |V_{Tp}|$) 时 $V_M = V_{DD}/2$ 。因为空穴迁移率约为电子的 1/3, 对称设计要求 $W_p \approx 3W_n$ 。

第二步, 噪声容限。传输曲线斜率为 -1 的两点 $V_{IL} < V_M < V_{IH}$ 之间增益大于 1, 输入中的噪声会被放大。凡低于 V_{IL} 的输入都可靠地被当作 “0”, 所以低电平噪声容限 $NM_L = V_{IL} - V_{OL}$; 同理 $NM_H = V_{OH} - V_{IH}$ 。CMOS 输出满摆幅, $V_{OL} = 0$, $V_{OH} = V_{DD}$ 。

第三步, 求对称反相器的 V_{IL} (记 $V_T = V_{Tn} = |V_{Tp}|$, $k_n = k_p$)。在 V_{in} 略高于 V_T 处, nMOS 饱和、pMOS 处于线性区:

$$\frac{1}{2}(V_{in} - V_T)^2 = (V_{DD} - V_{in} - V_T)(V_{DD} - V_{out})$$

$$- \frac{1}{2}(V_{DD} - V_{out})^2.$$

两边对 V_{in} 求导 (V_{out} 是 V_{in} 的函数) 并令 $dV_{out}/dV_{in} = -1$: 左边为 $V_{in} - V_T$, 右边为 $-(V_{DD} - V_{out}) + (V_{DD} - V_{in} - V_T) - (V_{DD} - V_{out})$, 化简得 $V_{out} = V_{in} + V_{DD}/2$ 。代回电流相等的方程, 得到 V_{in} 的一次方程 (二次项相消), 解出 (scripts/verify_sympy.py 核对)

$$V_{IL} = \frac{3V_{DD} + 2V_T}{8},$$

$$V_{IH} = \frac{5V_{DD} - 2V_T}{8}, \quad (9.2)$$

$$NM_L = NM_H = \frac{3V_{DD} + 2V_T}{8}.$$

$V_{DD} = 1.2\text{ V}$ 、 $V_T = 0.35\text{ V}$ 时 $NM = 0.54\text{ V}$, 接近 $V_{DD}/2$: CMOS 的噪声容限很大。图 12(a) 用计入 m 、沟道长度调制与亚阈值电流的插值器件模型 (附录 A.3) 数值求解, 给出 $V_{IL} = 0.537\text{ V}$, 与式 (9.2) 的 0.538 V 几乎相同。转换区中两管同时导通, 有一个电流尖峰; 静态时电流只有漏电。

电源电压能降到多低 若 $V_{DD} < V_T$, 两管都工作在亚阈值区, 反相器仍然能工作。取 $I_D = I_0 e^{V_{GS}/mU_T} (1 - e^{-V_{DS}/U_T})$ (式 (8.2) 的形式), 对称器件在中点 $V_{in} = V_{out} = V_{DD}/2$ 处对电流相等的方程求导, 可得中点增益 (推导见 verify_sympy.py)

$$|A| = \frac{e^{V_{DD}/2U_T} - 1}{m}. \quad (9.3)$$

增益必须大于 1 才能恢复逻辑电平, 所以 $V_{DD} > 2U_T \ln(1 + m) \approx 43\text{ mV}$ ($m = 1.3$)。考虑到噪声容限等因素, Swanson 与 Meindl 1972 年估计实际可用的下限约为 $8k_B T/q \approx 0.2\text{ V}$ (据该文摘要) [32]; 这一下限如何随 m 变化, 可由式 (9.3) 自行估计 (例题 4)。图 12(b) 的数值解与式 (9.3) 在 $V_{DD} \leq 0.15\text{ V}$ 时相差不到 2%。

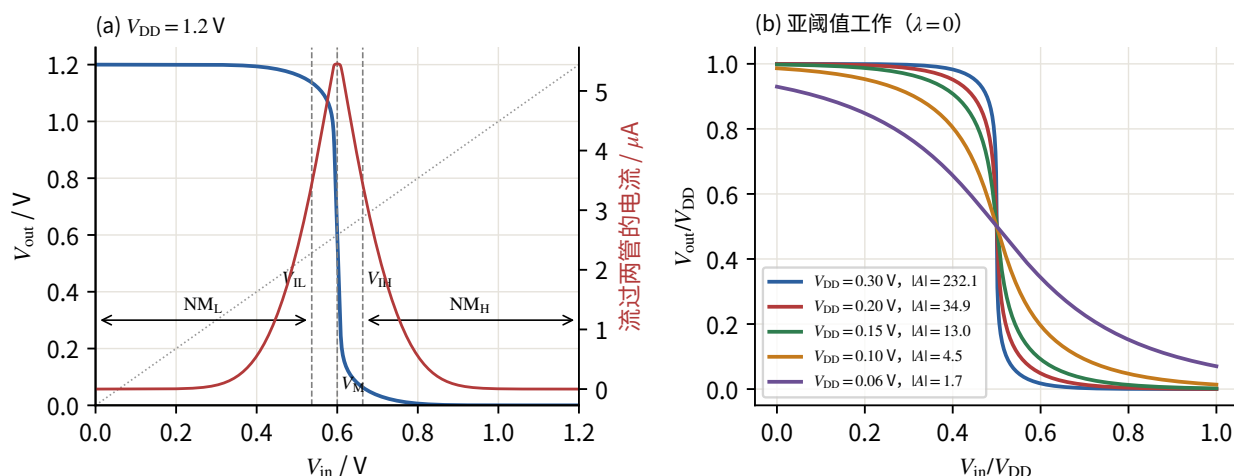


图 12 CMOS 反相器 ($k_n = k_p = 2 \times 10^{-4} \text{ A/V}^2$, $V_{Tn} = -V_{Tp} = 0.35 \text{ V}$, $m = 1.3$; 器件模型为把强反型体电荷平方律与弱反型指数律光滑连接的 EKV 型插值式, 见附录 A.3)。 (a) $V_{DD} = 1.2 \text{ V}$ 、 $\lambda = 0.2 \text{ V}^{-1}$ 的传输特性、噪声容限与流过两管的电流。 (b) 亚阈值工作 ($\lambda = 0$) 时归一化的传输特性: V_{DD} 降到几个 U_T 时增益趋于 1。

交互演示: CMOS 反相器的电压传输特性与功耗。 调节电源电压 V_{DD} 、两管阈值与强度比 k_p/k_n , 读出翻转电压 V_M 、噪声容限与中点增益; 给定负载电容、活动因子与时钟频率, 读出动态功耗 $\alpha C_L V_{DD}^2 f$ 与静态漏电功耗。

网页版中此处为可调参数的交互动画 (演示编号 mosfet-cmos)。

9.2 动态功耗与延迟

推导: $P = \alpha C_L V_{DD}^2 f$

第一步, 输出由 0 升到 V_{DD} (充电): 电源经 pMOS 向 C_L 输送电荷 $Q = C_L V_{DD}$, 电源做功 $\int V_{DD} i dt = V_{DD} Q = C_L V_{DD}^2$ 。电容最终储能 $\frac{1}{2} C_L V_{DD}^2$, 其余的 $\frac{1}{2} C_L V_{DD}^2$ 在 pMOS 中变成热——与 pMOS 的电阻大小、充电快慢都无关 (能量守恒)。

第二步, 输出由 V_{DD} 降到 0 (放电): 电源不做功, 电容储存的 $\frac{1}{2} C_L V_{DD}^2$ 在 nMOS 中变成热。

第三步, 一次完整的 “0→1→0” 消耗 $C_L V_{DD}^2$ 。若每个时钟周期中输出发生 0→1 跳变的概率为 α (活动因子), 时钟频率为 f , 平均功耗

$$P_{\text{dyn}} = \alpha C_L V_{DD}^2 f. \quad (9.4)$$

此外还有转换过程中两管同时导通的 “短路功耗” (通常占百分之十几以内) 与静态漏电功耗 $I_{\text{off}} V_{DD}$ 。

延迟 用开态电流 I_{on} ($V_{GS} = V_{DS} = V_{DD}$ 时的饱和电流) 把 C_L 放电到 $V_{DD}/2$, 所需时间

$$t_p \approx \frac{C_L V_{DD}/2}{I_{\text{on}}}. \quad (9.5)$$

这是一个量级估计: 放电过程中电流并非恒定 (后半段管子进入线性区), 系数约差 20%。 N 个 (奇数) 反相器首尾相连构成环形振荡器, 信号绕一圈要翻转两次才回到原状态, 振荡周期 $2Nt_p$, 频率 $f_{\text{osc}} = 1/(2Nt_p)$, 常用来测量工艺的门延迟。

数量级 取 $C_L = 1 \text{ fF}$ 、 $V_{DD} = 1 \text{ V}$ ，每次充放电耗能 1 fJ 。一块有 10^9 个门的芯片， $\alpha = 0.1$ 、 $f = 3 \text{ GHz}$ ，动态功耗为 $10^9 \times 0.1 \times 10^{-15} \times 1 \times 3 \times 10^9 = 300 \text{ W}$ ——这就是高性能处理器功耗在 100 W 量级、必须严格控制电压的原因。

9.3 Dennard 等比例缩小

推导：泊松方程的标度不变性

设把器件的所有尺寸缩小为 $1/\kappa$ ($x' = x/\kappa$)，所有电压缩小为 $1/\kappa$ ($\psi' = \psi/\kappa$)，掺杂浓度增大为 κ 倍 ($N' = \kappa N$)。在耗尽区中泊松方程为 $d^2\psi/dx^2 = qN_A/\epsilon_s$ 。设原器件的解为 $\psi(x)$ ，试探新器件的解 $\psi'(x') = \psi(\kappa x')/\kappa$ ：求两次导数， $d^2\psi'/dx'^2 = \kappa \psi''(\kappa x') = \kappa qN_A/\epsilon_s = qN'_A/\epsilon_s$ ，边界上的电压也恰为原来的 $1/\kappa$ ，所以 $\psi(\kappa x')/\kappa$ 就是新器件的解（反型层电子与可动载流子的分布另有 U_T 这一不缩小的尺度，见第 9.4 节）。于是所有电场 $\mathcal{E} \sim \psi/x$ 不变，所有耗尽层宽度按 $1/\kappa$ 缩小，器件的电场图样只是整体缩小 [20]。（式 (6.9) 直接验证： $W_d \propto \sqrt{\psi_s/N_A} \rightarrow \sqrt{(1/\kappa)/\kappa} = 1/\kappa$ 。）

由此逐项得到表 2：

表 2 Dennard 等比例缩小规则（缩小因子 $\kappa > 1$ ；“依据”一栏为本讲的推导出处）

物理量	缩放	依据
尺寸 L, W, t_{ox}, W_d ；电压 V	$1/\kappa$	设定 (W_d 由泊松方程不变性)
掺杂 N_A	κ	设定
电场 $\mathcal{E}$	1	泊松方程不变性
单位面积电容 C_{ox}	κ	ϵ_{ox}/t_{ox}
栅电容 $C_{ox}WL$ （负载电容）	$1/\kappa$	面积 $1/\kappa^2$ ，乘 C_{ox}
电流 I_D	$1/\kappa$	式 (7.6)： $C_{ox}(W/L)V^2$ ；式 (7.16) 亦然
延迟 $t_p \sim CV/I$	$1/\kappa$	式 (9.5)
单管功耗 VI	$1/\kappa^2$	
每次开关的能量 CV^2	$1/\kappa^3$	式 (9.4)
集成密度	κ^2	面积 $1/\kappa^2$
功率密度	1	$1/\kappa^2 \times \kappa^2$

结论：每一代器件更快 ($1/\kappa$)、更省（每次开关的能量 $1/\kappa^3$ ）、更密 (κ^2)，而芯片单位面积的发热不变。Dennard 等人用 $\kappa = 5$ 把一只沟道长 $5 \mu\text{m}$ 、栅氧 100 nm 的器件缩小为沟道长 $1 \mu\text{m}$ 、栅氧 20 nm 的器件 [20]。这条规则指导了此后约 30 年的集成电路发展。

9.4 缩小为何失效

表 2 假定所有电压都按 $1/\kappa$ 缩小，但器件中有几个电压是由材料和温度决定的，不随尺寸变化：

1. **亚阈值摆幅不缩小**。 $S \geq \ln 10 \cdot k_B T/q$ 只取决于温度。若 V_T 按 $1/\kappa$ 降低，由式 (8.5)，关断电流按 $10^{\Delta V_T/S}$ 指数增大。Dennard 等人在 1974 年的论文中已经指出亚阈值特性“不能按比例缩小” [20]。为了控制漏电， V_T 停在约 0.3 V ， V_{DD} 也停在约 1 V 。

2. **内建电势不缩小。** $2\phi_F$ 、源漏 pn 结的内建电势 $V_{bi} = U_T \ln(N_A N_D / n_i^2)$ (第 7.1 节) 都在 $E_g/q \approx 1\text{ V}$ 量级, 只随掺杂对数地变化, 耗尽层宽度因而缩小得比 $1/\kappa$ 慢, 需要更高的掺杂来补偿。
3. **栅氧化层隧穿。** 2002 年的 90 nm 工艺中栅氧已薄到约 1.2 nm (只有约 5 个原子层) [33]。电子直接隧穿氧化层的概率 $\propto e^{-2\kappa_t t_{ox}}$, $\kappa_t = \sqrt{2m_{ox}\Phi_B}/\hbar$ 。取 Si/SiO₂ 导带势垒 $\Phi_B \approx 3.1\text{ eV}$ [28]、氧化层中电子有效质量 $m_{ox} \approx 0.5m_0$, $\kappa_t \approx 6.4\text{ nm}^{-1}$: 氧化层每薄 0.2 nm, 隧穿电流约增大 $e^{2.6} \approx 13$ 倍。出路是改用介电常数高的材料 (铪基氧化物, 相对介电常数约 20): 在同样的 C_{ox} 下物理厚度可以大几倍。英特尔 2007 年的 45 nm 工艺首次在量产中采用高介电常数栅介质与金属栅, 等效氧化层厚度 1.0 nm [23]。
4. **结果。** 约 2004 年后电源电压基本不再降低, 由式 (9.4), 在功耗受限的条件下时钟频率无法继续提高, 单核性能的增长基本停止, 业界转向多核 [21]。尺寸仍在缩小 (摩尔定律延续), 但收益主要来自集成度。

9.5 短沟道效应、自然长度与多栅晶体管

沟道很短时, 源、漏的电场伸进沟道, 与栅“争夺”对沟道电势的控制: 阈值电压随 L 减小而降低, 随漏压增大而降低 (漏致势垒降低, DIBL), 亚阈值摆幅变差。描述这种效应的关键量是一个长度。

推导: 自然长度方程 (双栅薄体, 抛物线势近似)

考虑厚 t_{si} 的不掺杂硅薄层, 上下各有厚 t_{ox} 的氧化层与栅 (双栅), 栅电势记为 $V'_g (= V_{GS} - V_{FB})$ 。亚阈值区中可动电荷可以忽略, 硅中满足拉普拉斯方程 $\partial_x^2 \psi + \partial_y^2 \psi = 0$ (式 (7.3), $\rho = 0$)。

第一步 (近似): 在每个 y 处, 把 ψ 沿厚度方向 (x 从 $-t_{si}/2$ 到 $t_{si}/2$) 近似为抛物线 $\psi(x, y) = \psi_c(y) + a(y)x^2$ (对称, 所以没有一次项), ψ_c 为中心线电势。小参量是 t_{si}/L 。

第二步, 边界条件: 在 $x = t_{si}/2$ 处电位移法向分量连续。氧化层很薄, 其中电场近似均匀, 等于 $(V'_g - \psi_{surf})/t_{ox}$, $\psi_{surf} = \psi_c + at_{si}^2/4$ 为硅表面电势; 硅一侧的法向电场为 $-\partial_x \psi = -at_{si}$ 。连续条件为 $\varepsilon_s at_{si} = \varepsilon_{ox}(V'_g - \psi_{surf})/t_{ox}$ 。若 $\varepsilon_{ox} t_{si}/4 \ll \varepsilon_s t_{ox}$ (即 $t_{si} \ll 12t_{ox}$), 硅表面与中心的电势差可以忽略, $\psi_{surf} \approx \psi_c$ (附录 C 给出保留此项的结果与精确解), 于是 $a \approx \frac{\varepsilon_{ox}}{\varepsilon_s t_{si} t_{ox}} (V'_g - \psi_c)$ 。

第三步, 代入拉普拉斯方程在中心线 ($x = 0$) 处的形式 $2a + \psi_c'' = 0$:

$$\frac{d^2 \psi_c}{dy^2} = \frac{\psi_c - V'_g}{\ell^2}, \quad (9.6)$$

$$\ell = \sqrt{\frac{\varepsilon_s}{2\varepsilon_{ox}} t_{si} t_{ox}}.$$

ℓ 叫自然长度 (标度长度)。

式 (9.6) 与德拜屏蔽方程 (5.7) 形式相同: 栅“屏蔽”源、漏的影响, 偏离长沟道值 V'_g 的部分按 $e^{-y/\ell}$ 衰减。取源端 $\psi_c(0) = V_{bi}$ 、漏端 $\psi_c(L) = V_{bi} + V_{DS}$ (V_{bi} 为 n^+ 源漏与沟道之间的内建电

势，第 7.1 节；不掺杂硅体中约 0.5 V)，解为两个指数之和； $V_{DS} = 0$ 时

$$\begin{aligned}\frac{\psi_c(y) - V'_g}{V_{bi} - V'_g} &= \frac{\cosh[(y - L/2)/\ell]}{\cosh(L/2\ell)}, \\ \psi_{\min} - V'_g &= \frac{V_{bi} - V'_g}{\cosh(L/2\ell)} \\ &\approx 2(V_{bi} - V'_g)e^{-L/2\ell}.\end{aligned}\quad (9.7)$$

沟道中央电子势垒的顶部（ ψ 的极小值）比长沟道值低了 $\propto e^{-L/2\ell}$ 的一份；加上漏压，势垒进一步降低（DIBL）。栅对势垒顶的控制力 $d\psi_{\min}/dV'_g = 1 - \text{sech}(L/2\ell)$ ，由第 8.3 节的推理，亚阈值摆幅

$$S \approx \frac{60 \text{ mV/dec}}{1 - \text{sech}(L/2\ell)}.\quad (9.8)$$

检验：与二维数值解比较

scripts/scalelen.py 用有限差分直接解双栅结构的二维拉普拉斯方程（硅与氧化层分区取介电常数，栅与源漏取给定电势）。结果（图 13）：

- 中心线电势偏离长沟道值的部分确实按 $e^{-y/\ell}$ 衰减；拟合得到的衰减长度与附录 C 精确本征方程给出的 ℓ 相差 $< 1\%$ （ $t_{si} = 10 \text{ nm}$ 、 $t_{ox} = 1 \text{ nm}$ 时 $\ell = 4.92 \text{ nm}$ ；抛物线近似 (9.6) 给出 3.87 nm ，偏小约 20%）。
- 三种几何的 $\psi_{\min}$ 画成 L/ℓ 的函数落在同一条曲线上，比式 (9.7) 大一个几乎不变的因子 $c \approx 1.1\text{--}1.2$ （来自最低阶模式之外的投影系数；三种几何合起来拟合得 $c = 1.17$ ，演示 mosfet-scale 即用此值）；摆幅为 $60 \text{ mV/dec}/[1 - c \text{sech}(L/2\ell)]$ ， $L = 5\ell$ 时约 74 mV/dec ， $L = 3\ell$ 时约 119 mV/dec 。

结论：短沟道效应由 L/ℓ 这一个比值控制， $L \gtrsim 5\ell\text{--}10\ell$ 时可以忽略 [34]。

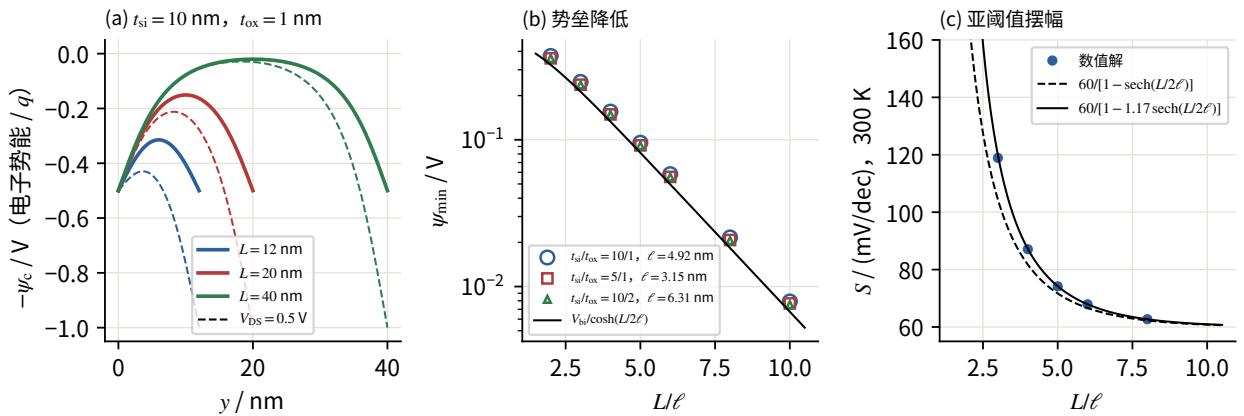


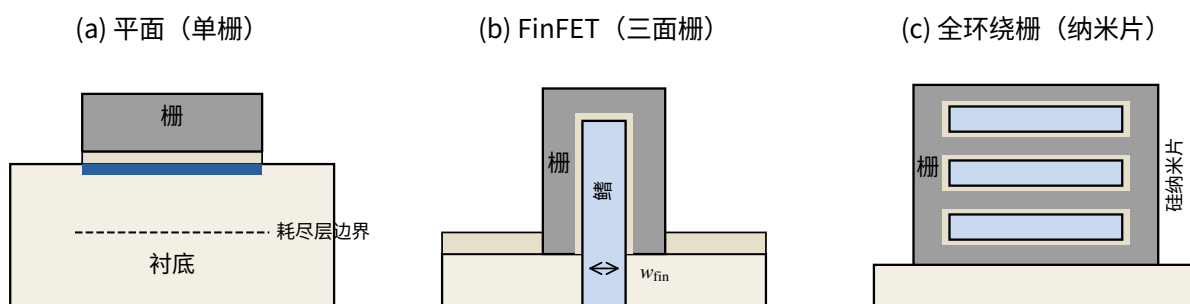
图 13 双栅全耗尽薄体 MOSFET 的短沟道效应（二维拉普拉斯方程数值解， $V'_g = 0$ ， $V_{bi} = 0.5 \text{ V}$ ）。(a) 沟道中心线上的电子势能 $-q\psi_c$ ： L 缩短时源漏之间的势垒降低，漏压使之进一步降低。(b) 势垒降低量随 L/ℓ 的变化：三种 t_{si}/t_{ox} 组合落在同一条线上，斜率与 $e^{-L/2\ell}$ 一致。(c) 亚阈值摆幅：数值解与 $60/[1 - c \text{sech}(L/2\ell)]$ 。

几何怎样决定 ℓ 式 (9.6) 的推导中，分母里的 2 来自上下两个栅各提供一份 ϵ_{ox}/t_{ox} 的“栅耦合”。推广到 N 面栅（截面近似为方形）， $\ell \approx \sqrt{\epsilon_s t_{si} t_{ox} / (N \epsilon_{ox})}$ ， N 叫“等效栅数”：单栅薄体 $N = 1$ ，

双栅 2，三面栅 (FinFET) 约 3，全环绕栅 4[34]。体硅平面器件没有薄体，耗尽层深度 W_d 起 t_{si} 的作用， $\ell \approx \sqrt{\varepsilon_s W_d t_{ox} / \varepsilon_{ox}}$ 。数值 ($\varepsilon_s / \varepsilon_{ox} = 3$ ，等效氧化层 1 nm)：

- 平面体硅， $W_d = 20 \text{ nm}$ ： $\ell \approx 7.7 \text{ nm}$ ，要求 $L \gtrsim 40 \text{ nm}$ ；而 W_d 要再减小就必须提高掺杂，迁移率与阈值涨落随之变差。
- FinFET，鳍宽 7 nm， $N = 3$ ： $\ell \approx 2.6 \text{ nm}$ ， $L \approx 15 \text{ nm}$ 也能保持良好栅控；全环绕栅 $N = 4$ 时 ℓ 再小约 13%，而且多片纳米片叠起来可以在同样的占地面积内提供更大的沟道宽度。

这就是从平面器件走向 FinFET (2011 年) 与全环绕栅 (2022 年) 的静电学原因 (图 14) [24, 25, 34]：栅从更多方向包住一块更薄的沟道， ℓ 变小，源漏电场被更有效地屏蔽。



剖面垂直于电流方向 (电流沿垂直纸面方向从源流向漏)；浅蓝为硅 ((a) 中深蓝为反型沟道)，米色为氧化层，灰色为栅。

图 14 三种晶体管结构垂直于电流方向的剖面 (示意，未按比例)。(a) 平面体硅：栅只在上方，沟道下方是很深的衬底。(b) FinFET：硅做成竖立的鳍，栅从两侧与顶部包住它。(c) 全环绕栅纳米片：多片很薄的硅片被栅完全包围。

交互演示：短沟道效应与自然长度。选择器件几何 (单栅、双栅、三面栅、全环绕栅) 并调节硅膜厚度、等效氧化层厚度与沟道长度，按自然长度模型计算中心线电势分布、势垒降低量、DIBL 与亚阈值摆幅。

网页版中此处为可调参数的交互动画 (演示编号 mosfet-scale)。

要点：CMOS 与缩小小结

- CMOS 反相器静态几乎不耗电；噪声容限约 $V_{DD}/2$ ；动态功耗 $\alpha C_L V_{DD}^2 f$ ，延迟 $\sim C_L V_{DD} / 2I_{on}$ 。
- Dennard 缩小来自泊松方程在 $(x, \psi, N) \rightarrow (x/\kappa, \psi/\kappa, \kappa N)$ 下不变；其失效来自不随尺寸缩小的 $k_B T/q$ (亚阈值摆幅)、 E_g/q (内建电势) 与隧穿。
- 短沟道效应由 L/ℓ 控制， $\ell \propto \sqrt{t_{si} t_{ox} / N}$ ；多栅结构减小 ℓ 。

10 决赛风格例题

例题中的数值均由 `scripts/example_numbers.py` 计算并用 `assert` 交叉核对。常数： $\varepsilon_0 = 8.854 \times 10^{-12} \text{ F/m}$ ， $\varepsilon_s = 11.7\varepsilon_0$ ， $\varepsilon_{ox} = 3.9\varepsilon_0$ ， $n_i = 1.0 \times 10^{10} \text{ cm}^{-3}$ (300 K)， $U_T = 25.85 \text{ mV}$

(300 K)。

例题 1：场效应为什么迟迟做不出来——表面态的屏蔽

1940 年代的场效应实验大致如下：在绝缘基片上蒸镀一层 n 型半导体薄膜，厚 $d_f = 1 \mu\text{m}$ ，施主浓度 $N_D = 10^{16} \text{ cm}^{-3}$ （全部电离，介电常数取 $11.7\epsilon_0$ ）；薄膜上方隔着厚 $t = 10 \mu\text{m}$ 、相对介电常数 6 的云母片放一块金属板，作为平行板电容器的另一极。在金属板上加 $V = 100 \text{ V}$ （相对薄膜为正），测量薄膜沿面方向电导的变化。（本题数值为示意，不对应某一次具体实验。）

1. 若薄膜表面没有表面态，感应电荷全部是可动电子，迁移率与原有电子相同，求感应电子的面密度与薄膜电导的相对变化 $\Delta G/G$ 。
2. 设表面有能量均匀分布、交换电子很快的表面态，密度为 D_s （单位面积、单位能量）。用小信号模型证明：可动电子只占感应电荷的比例 $f = C_{sc}/(C_{sc} + q^2 D_s)$ ，其中 C_{sc} 是半导体在平带附近的电容。求 C_{sc} ，并说明为什么感应电荷的总量仍由云母电容决定。
3. Bardeen 回忆，扣除薄膜迁移率低的因素后，余下的差距约为 10–20 倍。估计相应的 D_s ，并与硅表面原子的面密度（约 $6.8 \times 10^{14} \text{ cm}^{-2}$ ）比较。
4. 分别估计有、无表面态（取 $f = 1/20$ ）时薄膜表面势的变化。用“费米能级钉扎”解释结果。
5. 据 Shockley 回忆，Bardeen 的表面态模型引出了在液态空气温度下做场效应实验、以“冻结”表面态的建议，并由此得到了第一个正面结果 [5]。下面估计“冻结”的含义。设表面态能级在导带下 $\Delta E = 0.3 \text{ eV}$ ，向导带发射电子的时间常数 $\tau = 1/(\sigma_n v_{th} N_c e^{-\Delta E/k_B T})$ ，俘获截面 $\sigma_n = 10^{-15} \text{ cm}^2$ ，300 K 时 $v_{th} = 2.3 \times 10^7 \text{ cm/s}$ 、 $N_c = 3.2 \times 10^{19} \text{ cm}^{-3}$ （ $v_{th} \propto T^{1/2}$ ， $N_c \propto T^{3/2}$ ）。求 300 K 与 80 K（取 80 K 代表液态空气温度）时的 τ 。设实验在 80 K 下给栅加频率 100 Hz 的交变电压（本题的假设），说明为什么此时能“看见”场效应。
6. 现代热氧化 Si/SiO₂ 界面 $D_{it} \approx 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$ 。重算 (2) 中的 f ；并求它使 10 nm 栅氧 MOSFET 的亚阈值摆幅增加多少。若 $D_{it} = 10^{12} \text{ cm}^{-2} \text{ eV}^{-1}$ 又如何？

解答

(1) 云母电容 $C_g = 6\epsilon_0/t = 5.31 \times 10^{-6} \text{ F/m}^2$ 。外加电压几乎全部降在云母上（见 (2)），感应电荷 $\sigma = C_g V = 5.31 \times 10^{-4} \text{ C/m}^2$ ，面密度 $\sigma/q = 3.32 \times 10^{11} \text{ cm}^{-2}$ 。薄膜原有电子面密度 $N_D d_f = 10^{12} \text{ cm}^{-2}$ 。电导正比于可动电子总数， $\Delta G/G = 0.33$ ：效应很大，应当很容易测到。

(2) 小信号下，半导体的电荷随表面势变化 $\delta Q_{sc} = -C_{sc} \delta \psi_s$ （ ψ_s 升高，积累更多电子）；表面态的电荷 $\delta Q_{ss} = -q^2 D_s \delta \psi_s$ （式 (6.13)，能量宽 $q \delta \psi_s$ 的一段表面态被填满）。两者并联，由同一个 $\delta \psi_s$ 驱动，所以可动电荷所占比例

$$f = \frac{\delta Q_{sc}}{\delta Q_{sc} + \delta Q_{ss}} = \frac{C_{sc}}{C_{sc} + q^2 D_s}.$$

平带附近半导体的屏蔽长度是德拜长度，由第 6.8 节， $C_{sc} = \epsilon_s/L_D$ ， $L_D = \sqrt{\epsilon_s U_T/(q N_D)} = 40.9 \text{ nm}$ ， $C_{sc} = 2.53 \times 10^{-3} \text{ F/m}^2$ 。云母电容与 $(C_{sc} + C_{ss})$ 串联， $C_g/C_{sc} = 2 \times 10^{-3}$ ，所以外加电压几乎全部降在云母上，感应电荷总量 $\approx C_g V$ ，与表面态无关；表面态只改变这些电荷的去处。

(3) $f = 1/10$ 时 $q^2 D_s = 9C_{sc}$, $D_s = 1.4 \times 10^{13} \text{ cm}^{-2} \text{ eV}^{-1}$; $f = 1/20$ 时 $D_s = 3.0 \times 10^{13} \text{ cm}^{-2} \text{ eV}^{-1}$ 。这相当于每 1 eV 能量范围内有 2%~4% 的表面原子提供一个电子态。一个未经处理的表面每个原子都可能有悬挂键, 所以这个数值是完全合理的——表面态足以解释实验的失败。

(4) 无表面态时, 电荷全部进入积累层。积累层的精确第一积分 (式 (6.4) 中把空穴换成电子) 为 $|Q| = \sqrt{2\varepsilon_s k_B T N_D} \sqrt{e^{\psi_s/U_T} - \psi_s/U_T - 1}$, 令其等于 σ , 数值解得 $\psi_s = 0.094 \text{ V} \approx 3.6U_T$ (大信号, 线性近似不适用)。有表面态 ($f = 1/20$) 时, $C_{sc} + C_{ss} = 20C_{sc}$, $\Delta\psi_s \approx \sigma/(20C_{sc}) = 10.5 \text{ mV} = 0.41U_T$, 线性近似适用。表面态像一个巨大的电容挂在表面上, 外加电荷只能让表面处的费米能级相对能带移动约 10 meV——费米能级被“钉”在表面态能级附近, 能带几乎不变, 半导体中的载流子几乎不变。

(5) 300 K: $\tau = 1/[10^{-19} \text{ m}^2 \times 2.3 \times 10^5 \text{ m/s} \times 3.2 \times 10^{25} \text{ m}^{-3} \times e^{-0.3/0.02585}] = 1.5 \times 10^{-7} \text{ s}$ 。80 K: v_{th} 与 N_c 的前因子减小为 $(80/300)^2 = 0.071$ 倍, 指数因子变成 $e^{-0.3/0.00689}$, $\tau = 1.5 \times 10^8 \text{ s}$ (约 5 年)。室温下表面态在 $0.15 \mu\text{s}$ 内就跟上栅压变化, 任何频率的测量都看到被屏蔽的结果; 80 K 下表面态的占据在 100 Hz 的一个周期 (10 ms) 内完全“冻结”, 交变栅压感应的电荷只能由可动载流子承担 (交流的 $C_{ss} \rightarrow 0$, $f \rightarrow 1$), 交流电导调制就显现出来。(这里的推理只针对表面态; 低温下施主也会部分“冻析”, 实验材料需选浅能级杂质。)

(6) $D_{it} = 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$: $C_{it} = 1.6 \times 10^{-5} \text{ F/m}^2$, $f = 2.53/(2.53 + 0.016) = 0.994$ 。对 10 nm 栅氧, $C_{it}/C_{ox} = 0.0046$, 由式 (8.4), $\Delta S = \ln 10 \cdot U_T \cdot C_{it}/C_{ox} = 0.28 \text{ mV/dec}$, 可以忽略。 $D_{it} = 10^{12}$ 时 $C_{it}/C_{ox} = 0.46$, $\Delta S = 28 \text{ mV/dec}$ ——摆幅变差三分之一以上。热氧化把界面态降低了约三个数量级, 这是 MOSFET 能工作的前提。

评分要点 (共 30 分): (1) C_g 、感应电荷、 $\Delta G/G$ 4 分; (2) 并联电容模型与 f 4 分, $C_{sc} = \varepsilon_s/L_D$ 2 分, 说明总电荷由 C_g 决定 2 分; (3) D_s 及与原子密度的比较 4 分; (4) 无表面态用大信号精确式 3 分, 有表面态的 $\Delta\psi_s$ 与钉扎解释 3 分; (5) 两个温度的 τ 3 分, 频率论证 2 分; (6) f 与 ΔS 3 分。

例题 2: 阈值电压的设计、体效应与温度系数

设计一种 n 沟道 MOSFET: n^+ 多晶硅栅, 栅氧 $t_{ox} = 4 \text{ nm}$, 衬底均匀掺杂, 300 K 时要求 $V_T = 0.40 \text{ V}$ (按式 (6.11) 计算, 不计界面陷阱)。

1. 求所需的衬底掺杂 N_A , 以及 ϕ_F 、 γ 、 m 与最大耗尽层宽度。
2. 工艺在界面处留下 $5 \times 10^{10} \text{ q cm}^{-2}$ 的固定正电荷, 阈值变化多少?
3. 求 $V_{SB} = 0.5$ 、 1.0 V 时的阈值。把这只管子接成源极跟随器 (栅为输入, 源为输出并接恒流源, 衬底接地), 证明小信号增益 $A = dV_S/dV_G = [1 + \gamma/(2\sqrt{2\phi_F + V_{SB}})]^{-1}$, 并求 $V_{SB} = 0.8 \text{ V}$ 处的 A 。
4. 求阈值电压的温度系数 dV_T/dT 。设 $n_i \propto T^{3/2}e^{-E_g/2k_B T}$, 在 300 K 附近 $E_g(T) = E_{g0} - aT$, $a = 2.55 \times 10^{-4} \text{ eV/K}$, $E_g(300 \text{ K}) = 1.12 \text{ eV}$; n^+ 栅的 $\phi_{ms} = -(E_g/2q + \phi_F)$ 也随温度变化。
5. 芯片温度从 25°C 升到 85°C , 设阈值处的电流 I_T 不变, 亚阈值摆幅取 $\ln 10 \cdot mU_T$ (m 按 (1) 的值), 关断电流增大多少倍?

6. **阈值调整注入。**另一种设计：衬底均匀掺杂只取 $N'_A = 3 \times 10^{17} \text{ cm}^{-3}$ ，再在紧贴界面处注入一薄层受主（面密度 D_I ，深度远小于耗尽层宽度，可看作 $x = 0^+$ 处的面电荷 $-qD_I$ ）。(a) 用高斯定理与栅压方程证明：这层电荷只使阈值升高 qD_I/C_{ox} ，而不改变强反型时的耗尽层宽度，因而不改变 C_{dm} 与 m 。(b) 求使 $V_T = 0.40 \text{ V}$ 的 D_I 。(c) 比较两种设计的 m 、亚阈值摆幅与 $V_{\text{SB}} = 1 \text{ V}$ 时的体效应，说明工业上为什么这样做。(d) 实际注入层深度 $x_I = 10 \text{ nm}$ ，估计“面电荷”近似带来的阈值误差量级。

解答

(1) $C_{\text{ox}} = 8.63 \times 10^{-3} \text{ F/m}^2$ 。式 (6.11) 中 $V_{\text{FB}} = -(0.56 + \phi_F)$ ，于是 $V_T = \phi_F - 0.56 + \sqrt{4\varepsilon_s q N_A \phi_F}/C_{\text{ox}}$ ， $\phi_F = U_T \ln(N_A/n_i)$ 。这是 N_A 的超越方程，数值求解（二分法）得

$$N_A = 6.0 \times 10^{17} \text{ cm}^{-3}, \quad \phi_F = 0.463 \text{ V},$$

$$V_{\text{FB}} = -1.023 \text{ V}, \quad \frac{|Q_d|}{C_{\text{ox}}} = 0.497 \text{ V}.$$

$\gamma = \sqrt{2\varepsilon_s q N_A}/C_{\text{ox}} = 0.517 \text{ V}^{1/2}$ ， $m = 1 + \gamma/(2\sqrt{2\phi_F}) = 1.268$ ， $W_{\text{d,max}} = \sqrt{4\varepsilon_s \phi_F/qN_A} = 44.7 \text{ nm}$ 。检验： $-1.023 + 0.926 + 0.497 = 0.400$ 。

(2) 由式 (6.2)， $\Delta V_T = \Delta V_{\text{FB}} = -Q_f/C_{\text{ox}} = -1.602 \times 10^{-19} \text{ C} \times 5 \times 10^{14} \text{ m}^{-2}/8.63 \times 10^{-3} \text{ F/m}^2 = -9.3 \text{ mV}$ 。栅氧薄，固定电荷的影响就小。

(3) 由式 (7.10)： $V_T(0.5) = 0.520 \text{ V}$ ， $V_T(1.0) = 0.620 \text{ V}$ 。源极跟随器中电流由恒流源固定，饱和和电流公式中 $V_{\text{GS}} - V_T(V_{\text{SB}})$ 必须保持不变（衬底接地， $V_{\text{SB}} = V_S$ ）：

$$dV_G - dV_S = dV_T = \frac{dV_T}{dV_{\text{SB}}} dV_S,$$

$$\frac{dV_T}{dV_{\text{SB}}} = \frac{\gamma}{2\sqrt{2\phi_F + V_{\text{SB}}}}.$$

整理即得 $A = [1 + \gamma/(2\sqrt{2\phi_F + V_{\text{SB}}})]^{-1}$ （ $\lambda = 0$ 时的结果）。 $V_{\text{SB}} = 0.8 \text{ V}$ 时 $dV_T/dV_{\text{SB}} = 0.197$ ， $A = 0.836$ ：体效应使跟随器的增益明显小于 1。

(4) 第一步，求 $d\phi_F/dT$ 。 $\phi_F = \frac{k_B T}{q} [\ln N_A - \ln n_i]$ ， $\ln n_i = \text{常数} + \frac{3}{2} \ln T - \frac{E_{g0} - aT}{2k_B T}$ （ a 项与 T 无关，并入常数）。求导：

$$\frac{d\phi_F}{dT} = \frac{\phi_F}{T} - \frac{k_B T}{q} \left(\frac{3}{2T} + \frac{E_{g0}}{2k_B T^2} \right)$$

$$= \frac{1}{T} \left[\phi_F - \frac{3}{2} U_T - \frac{E_{g0}}{2q} \right].$$

$E_{g0} = 1.12 + 0.0765 = 1.196 \text{ eV}$ ，代入得 $d\phi_F/dT = (0.463 - 0.039 - 0.598)/300 = -0.58 \text{ mV/K}$ 。温度升高，本征浓度增大，费米能级向禁带中央靠拢。

第二步，对 $V_T = -\frac{E_g}{2q} - \phi_F + 2\phi_F + \gamma\sqrt{2\phi_F}$ 逐项求导， $dE_g/dT = -a$ ， $d(\gamma\sqrt{2\phi_F})/dT = \frac{\gamma}{\sqrt{2\phi_F}} \frac{d\phi_F}{dT} = 2(m-1) \frac{d\phi_F}{dT}$ ：

$$\frac{dV_T}{dT} = \frac{a}{2q} + (2m-1) \frac{d\phi_F}{dT}$$

$$= 0.127 - 1.536 \times 0.580 = -0.76 \text{ mV/K}.$$

（数值差分核对一致。）阈值随温度升高而降低，典型值为 -0.5 到 -2 mV/K 。

(5) 25 °C: $V_T = 0.401 \text{ V}$, $S = \ln 10 \times m \times 25.69 \text{ mV} = 75.0 \text{ mV/dec}$; 85 °C: $V_T = 0.401 - 60 \times 0.76 \text{ mV} = 0.354 \text{ V}$, $S = 90.1 \text{ mV/dec}$ 。由式 (8.5),

$$\begin{aligned} \frac{I_{\text{off}}(85^\circ\text{C})}{I_{\text{off}}(25^\circ\text{C})} &= 10^{0.401/0.0750 - 0.354/0.0901} \\ &= 10^{1.42} \approx 26. \end{aligned}$$

阈值降低与摆幅变大共同作用, 60 K 的温升使漏电增大约 26 倍; 实际中 I_T 还会随温度变化, 但不改变量级。漏电随温度指数增大、漏电又发热, 可能形成“热失控”, 这是芯片热设计必须考虑的。

(6)(a) 面电荷位于 $x = 0^+$, 在它与界面之间没有厚度, 所以它不产生电势差: 半导体中耗尽区 ($0 < x < W_d$) 的电势分布仍由体掺杂 N'_A 决定, $\psi_s = qN'_A W_d^2 / 2\epsilon_s$ 不变, 强反型时仍有 $W_{d,\max} = \sqrt{4\epsilon_s \phi_F / qN'_A}$, $C_{\text{dm}} = \epsilon_s / W_{d,\max}$ 与 m 都不变。但由高斯定理, 半导体中的总电荷多了 $-qD_I$, 栅上要多 $+qD_I$, 栅压方程 (6.3) 中多出 qD_I / C_{ox} 一项, 所以 $\Delta V_T = qD_I / C_{\text{ox}}$ (与固定正电荷的式 (6.2) 同理, 符号相反)。(b) 均匀掺杂 $3 \times 10^{17} \text{ cm}^{-3}$ 时 $\phi_F = 0.445 \text{ V}$, 式 (6.11) 给出 $V_T = 0.230 \text{ V}$ 。需补 0.170 V : $D_I = 0.170 \times C_{\text{ox}} / q = 9.2 \times 10^{11} \text{ cm}^{-2}$ 。(c) 注入设计: $\gamma = 0.366 \text{ V}^{1/2}$, $m = 1.194$, $W_{d,\max} = 62 \text{ nm}$; 均匀设计: $\gamma = 0.517$, $m = 1.268$ 。亚阈值摆幅 ($\ln 10 m U_T$) 71.1 对 75.5 mV/dec ; $V_{\text{SB}} = 1 \text{ V}$ 时阈值升高 0.158 对 0.220 V 。同样的阈值下, 注入设计的开关更陡、体效应更小——“阈值由界面附近的电荷定、静电学由体掺杂定”, 两者被分开调节。这正是 Dennard 1974 年论文采用离子注入的原因之一 [20]。(本题的注入层紧贴界面, 正在反型电子所在的一两纳米之内, 不能指望它减少杂质散射; 若把注入峰放在沟道下方, 做成表面低、深处高的“倒掺杂”, 还可减少沟道内的杂质散射, 但那已不是本题的表面薄层模型。)(d) 若注入电荷均匀分布在 $0 < x < x_I$ 内, 它自身就会在半导体里产生一段电势降: 可等效看作位于 $x_I/2$ 的面电荷, 它与界面之间多出 qD_I / ϵ_s 的电场, 贡献表面势 $\delta\psi \approx qD_I x_I / (2\epsilon_s) = 0.071 \text{ V}$ 。但这 0.071 V 并不是阈值误差。阈值判据固定为 $\psi_s = 2\phi_F$, 注入层分走了 $\delta\psi$, 体掺杂只需提供 $2\phi_F - \delta\psi$: 耗尽层略变窄, $|Q_d|$ 减小约 $C_{\text{dm}}\delta\psi$ (因 $d|Q_d|/d\psi_s = C_d$); 而注入电荷总量不变, 栅压方程中 qD_I / C_{ox} 一项不变。于是

$$\begin{aligned} \Delta V_T &\approx -\frac{C_{\text{dm}}}{C_{\text{ox}}} \delta\psi = -(m-1) \frac{qD_I x_I}{2\epsilon_s} \\ &= -0.194 \times 0.071 \text{ V} \approx -0.014 \text{ V}. \end{aligned}$$

用耗尽近似精确求解 (D_I 均匀放在 $0 < x < x_I$ 内, `example_numbers.py`) 得 $V_T = 0.386 \text{ V}$ ($x_I = 10 \text{ nm}$) 与 0.371 V (20 nm), 面电荷近似为 0.400 V 。所以面电荷近似对 10 nm 深的注入层误差约 0.01 V , 约为调整量 0.17 V 的 8%, 方向是实际阈值略偏低: 注入层内的电势差绝大部分被体耗尽电荷的减少抵消, 只剩 $C_{\text{dm}}/C_{\text{ox}}$ 那一份。设计时把剂量相应略加大即可。

评分要点 (共 38 分): (1) 列出超越方程并数值求解 4 分, γ 、 m 、 $W_{d,\max}$ 3 分; (2) 3 分; (3) 两个阈值 2 分, 跟随器增益推导 4 分, 数值 1 分; (4) $d\phi_F/dT$ 推导 4 分, dV_T/dT 的各项 (含 V_{FB} 的温度依赖) 4 分; (5) 两个温度下的 V_T 、 S 与倍数 5 分; (6) 面电荷不改变 W_d 的论证 3 分, ΔV_T 与 D_I 2 分, 两种设计的比较 2 分, 有限深度修正 1 分 (须指出 ψ_s 被注入层分走一部分 $\rightarrow W_d$ 减小 $\rightarrow \Delta V_T \approx -(m-1)qD_I x_I / 2\epsilon_s$, 只答 0.07 V 不得分)。

例题 3：短沟道器件的电流、增益与速度——哪种效应先出现

某工艺的 n 沟道 MOSFET: n^+ 多晶硅栅, $t_{\text{ox}} = 4 \text{ nm}$, 衬底均匀掺杂 $N_A = 10^{18} \text{ cm}^{-3}$, 按式 (6.11)、(7.8) 得 $V_T = 0.57 \text{ V}$ 、 $m = 1.34$ (自行核验); 反型层有效迁移率 $\mu_{\text{eff}} = 250 \text{ cm}^2/(\text{V s})$, $v_{\text{sat}} = 1 \times 10^7 \text{ cm/s}$, $W = 1 \mu\text{m}$, 工作在 $V_{\text{GS}} = V_{\text{DS}} = V_{\text{DD}} = 1.8 \text{ V}$ 。漂移速度用式 (7.13)。

1. 求 $\mathcal{E}_{\text{sat}}$ 。分别对 $L = 2 \mu\text{m}$ 与 $L = 0.18 \mu\text{m}$ 求 $V_{\text{DS,sat}}$ 与 $I_{\text{D,sat}}$, 与平方律 (7.9) 比较。速度饱和和使电流减半的沟道长度是多少?
2. 求 $L = 0.18 \mu\text{m}$ 时饱和区的跨导 g_m , 与长沟道公式 $\mu C_{\text{ox}} W V_{\text{ov}}/(mL)$ 及短沟道极限 $W C_{\text{ox}} v_{\text{sat}}$ 比较。
3. 源极接触与扩展区有串联电阻 $R_s = 200 \Omega$ (即 $R_s W = 200 \Omega \mu\text{m}$)。证明在一阶近似下 $I_D \approx I_{\text{D0}}/(1 + g_m R_s)$ 、外部跨导 $g_{m,\text{ext}} = g_m/(1 + g_m R_s)$, 并求 $L = 0.18 \mu\text{m}$ 时电流减小的比例。为什么说串联电阻对短沟道器件“更致命”?
4. 实测 $L = 2 \mu\text{m}$ 的器件 $\lambda = 0.02 \text{ V}^{-1}$ 。设 $\lambda \propto 1/L$ (第 7.8 节的估计), 按第 7.8 节的定义 $I_D = I_{\text{D0}}(1 + \lambda V_{\text{DS}})$ ($I_{\text{D0}} = I_{\text{D,sat}}(V_{\text{GS}})$) 自洽地求两种沟道长度在工作点处的本征增益 $g_m r_o$ 。与式 (7.12) 的“ $\propto L$ ”相比, 短沟道器件的增益降得更快还是更慢? 哪些因素起作用? 若沿用 $\lambda V_{\text{DS}} \ll 1$ 的近似, 结论会怎样?
5. 用平面体硅的自然长度 $\ell \approx \sqrt{\epsilon_s W_{\text{d,max}} t_{\text{ox}}/\epsilon_{\text{ox}}}$ 估计短沟道静电效应开始显著的沟道长度 (取 $L \approx 5\ell$)。把沟道从 $2 \mu\text{m}$ 缩到 $0.18 \mu\text{m}$ 的过程中, 速度饱和、沟道长度调制、短沟道静电效应依次在什么长度变得重要?
6. **本征速度**。栅压变化 dV_{GS} 时, 栅要向沟道提供电荷 dQ_{ch} ($Q_{\text{ch}} = W \int_0^L Q_n dy$ 为沟道总电荷), 定义栅-源电容 $C_{\text{gs}} = |dQ_{\text{ch}}/dV_{\text{GS}}|$; 晶体管能放大的最高频率 (电流增益降到 1 的频率) 约为 $f_T = g_m/(2\pi C_{\text{gs}})$ 。(a) 不计速度饱和, 利用式 (7.1) 把 dy 换成 dV , 证明饱和时 $Q_{\text{ch}} = -\frac{2}{3} W L C_{\text{ox}} V_{\text{ov}}$, 进而 $f_T = 3\mu V_{\text{ov}}/(4\pi m L^2)$, 并解释它与载流子渡越时间的关系。(b) 计入速度饱和, 写出 Q_{ch} 的积分表达式, 数值求两种长度的 $C_{\text{gs}}/(W L C_{\text{ox}})$ 与 f_T , 与 (a) 的公式及 $v_{\text{sat}}/(2\pi L)$ 比较。

解答

(1) $\mathcal{E}_{\text{sat}} = v_{\text{sat}}/\mu = 1 \times 10^5 \text{ m/s}/0.025 \text{ m}^2/(\text{V s}) = 4 \times 10^6 \text{ V/m} = 40 \text{ kV/cm}$; $C_{\text{ox}} = 8.63 \times 10^{-3} \text{ F/m}^2$; 核验 $\phi_F = 0.476 \text{ V}$, $\gamma = 0.667 \text{ V}^{1/2}$, $V_T = -1.036 + 0.952 + 0.651 = 0.567 \text{ V}$, $m = 1 + \gamma/(2\sqrt{2\phi_F}) = 1.342$; $V_{\text{ov}} = 1.23 \text{ V}$ 。由式 (7.15):

L	$\mathcal{E}_{\text{sat}} L$	$V_{\text{DS,sat}}$ (平方律)	$I_{\text{D,sat}}$ (平方律)	比值
$2 \mu\text{m}$	8.0 V	0.87 V (0.92 V)	0.055 mA (0.061 mA)	0.90
$0.18 \mu\text{m}$	0.72 V	0.64 V (0.92 V)	0.326 mA (0.679 mA)	0.48

两种长度都满足 $V_{\text{DS}} = 1.8 \text{ V} > V_{\text{DS,sat}}$, 处于饱和区。电流减半处由第 7.9 节, $L_c = V_{\text{ov}}/(m\mathcal{E}_{\text{sat}}) = 0.23 \mu\text{m}$ (此时比值为 0.536)。

(2) 对 $I_{\text{D,sat}}(V_{\text{ov}})$ 数值求导: $g_m = 0.405 \text{ mS}$ 。长沟道公式给出 1.10 mS , 短沟道极限为 $W C_{\text{ox}} v_{\text{sat}} = 0.863 \text{ mS}$: $0.18 \mu\text{m}$ 器件的跨导只有平方律的 37%, 也只有短沟道极限的一半

——它处在过渡区。

(3) 源极电阻上的压降使器件实际感受到的栅源电压为 $V_{GS} - I_D R_s$ 。把 $I_D(V_{GS} - I_D R_s)$ 展开到一阶: $I_D \approx I_{D0} - g_m I_D R_s$, 解得 $I_D \approx I_{D0}/(1 + g_m R_s)$; 对 V_{GS} 求导得 $g_{m,ext} = g_m/(1 + g_m R_s)$ (负反馈)。 $g_m R_s = 0.081$: 一阶式给出 0.302 mA, 与自洽数值解一致, 电流减小 7.4%, $g_{m,ext} = 0.375$ mS。缩小器件时 g_m 随 $1/L$ 增大 (长沟道) 或趋于 $WC_{ox}v_{sat}$ (短沟道), 而接触电阻 $R_s W$ 由接触与扩展区决定, 并不随 L 缩小; 在短沟道极限 $g_m R_s \rightarrow WC_{ox}v_{sat} R_s = 0.17$, 且 C_{ox} 每代增大, 这一比例越来越大。

(4) 由定义, $g_m = (1 + \lambda V_{DS})g_{m0}$, $g_{m0} \equiv dI_{D,sat}/dV_{GS}$; $r_o = 1/(\lambda I_{D0})$ 。于是

$$g_m r_o = \frac{1 + \lambda V_{DS}}{\lambda} \cdot \frac{g_{m0}}{I_{D0}}.$$

$L = 2 \mu\text{m}$: $\lambda V_{DS} = 0.036$, $g_{m0} = 0.085$ mS, $g_{m0}/I_{D0} = 1.54 \text{ V}^{-1}$, $r_o = 910 \text{ k}\Omega$, $g_m r_o = 80$ 。
 $L = 0.18 \mu\text{m}$: $\lambda = 0.02 \times 2/0.18 = 0.22 \text{ V}^{-1}$, $\lambda V_{DS} = 0.40$, $g_{m0} = 0.405$ mS (第 (2) 问), $g_{m0}/I_{D0} = 1.24 \text{ V}^{-1}$, $r_o = 13.8 \text{ k}\Omega$, $g_m r_o = 7.8$ 。比值 $80/7.8 = 10.2$, 略小于长度比 11.1: 增益大体仍按 $\propto L$ 下降, 甚至稍慢。两个因素方向相反: (i) 速度饱和使 g_{m0}/I_{D0} 从长沟道的 $2/V_{ov} = 1.62 \text{ V}^{-1}$ 向 $1/V_{ov}$ (式 (7.16) 中 $I \propto V_{ov}$) 下降, 这里降了 20%; (ii) $\lambda V_{DS} = 0.4$ 时 $(1 + \lambda V_{DS})$ 因子把 g_m 抬高 40% (长沟道只有 4%), 恰好把前者抵消有余。若沿用 $\lambda V_{DS} \ll 1$ 的近似 ($r_o \approx 1/(\lambda I_{D,sat})$ 、 $g_m \approx g_{m0}$), 得 77 与 5.6, 比值 14, 结论会变成“降得更快”—— $\lambda V_{DS} = 0.4$ 时这一近似已不可靠。结论对 CLM 的具体写法也不敏感: 用 `mosfet-iv` 演示的写法 $I_D = I_{D,sat}(1 + \lambda V_{DS})/(1 + \lambda V_{DS,sat})$ 数值求导, 得 79 与 7.3, 比值 10.8。真正的教训是: 短沟道器件的 λ 大到使 λV_{DS} 不再是小量, “增益 $\approx 2/(\lambda V_{ov})$ ” 这类小信号公式本身要重新审视。

(5) $W_{d,max} = \sqrt{4\epsilon_s \phi_F / q N_A} = 35 \text{ nm}$, $\ell = \sqrt{3 \times 35 \times 4} \text{ nm} = 20.5 \text{ nm}$, $5\ell \approx 0.10 \mu\text{m}$ 。依次: 速度饱和在 $2 \mu\text{m}$ 时已使电流降低 10%, 在 $L \approx 0.23 \mu\text{m}$ 时成为主导; 沟道长度调制按 $1/L$ 增长, 在 $0.18 \mu\text{m}$ 时 $\lambda V_{DS} \approx 0.4$, 已严重影响模拟增益; 短沟道静电效应要到 $L \lesssim 0.1 \mu\text{m}$ 才显著 (此时阈值随 L 、 V_{DS} 漂移)。所以对这一工艺, “先出现”的是速度饱和——这也是 1990 年代以来短沟道器件饱和电流近似 $\propto V_{ov}$ 的原因。

(6)(a) 由式 (7.1), $dy = W\mu(-Q_n)dV/I_D$, 所以

$$\begin{aligned} Q_{ch} &= W \int_0^L Q_n dy = -\frac{W^2 \mu}{I_D} \int_0^{V_{DS,sat}} Q_n^2 dV \\ &= -\frac{W^2 \mu C_{ox}^2}{I_D} \int_0^{V_{ov}/m} (V_{ov} - mV)^2 dV \\ &= -\frac{W^2 \mu C_{ox}^2 V_{ov}^3}{3mI_D}. \end{aligned}$$

代入 $I_D = \mu C_{ox} W V_{ov}^2 / (2mL)$, $Q_{ch} = -\frac{2}{3} W L C_{ox} V_{ov}$ (与 m 无关: 夹断后沟道电荷沿 y 呈 $\sqrt{1-y/L}$ 分布, 平均值为满值的 $2/3$)。于是 $C_{gs} = \frac{2}{3} W L C_{ox}$, $f_T = \frac{\mu C_{ox} W V_{ov} / (mL)}{2\pi \cdot \frac{2}{3} W L C_{ox}} = \frac{3\mu V_{ov}}{4\pi m L^2}$ 。 $L^2/(\mu V_{ov})$ 正是电子在平均电场 V_{ov}/L 下漂过沟道的渡越时间, 所以 $f_T \sim 1/(2\pi \times \text{渡越时间})$: 晶体管不可能比电子穿过沟道更快地响应。(b) 计入速度饱和, 由第 7.9 节第二步,

$$dy = [W\mu(-Q_n)/I_D - 1/\mathcal{E}_{\text{sat}}]dV,$$

$$Q_{\text{ch}} = -WC_{\text{ox}} \int_0^{V_{\text{DS,sat}}} (V_{\text{ov}} - mV) \times \left[\frac{W\mu C_{\text{ox}}(V_{\text{ov}} - mV)}{I_D} - \frac{1}{\mathcal{E}_{\text{sat}}} \right] dV,$$

对 V_{ov} 数值求导 (example_numbers.py): $L = 2\mu\text{m}$ 时 $C_{\text{gs}}/(WLC_{\text{ox}}) = 0.70$, $f_T = 1.1\text{ GHz}$ ((a) 式给 1.4 GHz); $L = 0.18\mu\text{m}$ 时 $C_{\text{gs}}/(WLC_{\text{ox}}) = 0.82$, $f_T = 50\text{ GHz}$, 而 (a) 式给 169 GHz , $v_{\text{sat}}/(2\pi L) = 88\text{ GHz}$ 。速度饱和时沟道电荷分布更均匀 (C_{gs} 向 WLC_{ox} 靠近), 渡越速度又被 v_{sat} 封顶, f_T 从 $\propto 1/L^2$ 变为接近 $\propto v_{\text{sat}}/L$ 。

评分要点 (共 45 分): (1) $\mathcal{E}_{\text{sat}}$ 、 V_T 、 m 的核验 3 分, 两种长度的 $V_{\text{DS,sat}}$ 、 $I_{\text{D,sat}}$ 及比较 6 分, L_c 2 分; (2) g_m 与两个极限 4 分; (3) 一阶推导 4 分, 数值 2 分, “更致命”的论证 3 分; (4) 自洽的两个本征增益 4 分, g_m/I_D 与 $(1 + \lambda V_{\text{DS}})$ 两个因素的分析及与近似结果的对比 3 分; (5) ℓ 的计算 2 分, 三种效应的排序 2 分; (6) 换元求 Q_{ch} 与 $2/3$ 因子 4 分, f_T 与渡越时间 2 分, 速度饱和情形的积分与数值 4 分。

例题 4: 降低电源电压——漏电、延迟与最小能量点

一条逻辑通路由 $N_d = 50$ 个相同的门串联, 时钟周期等于这条通路的总延迟 $T_{\text{clk}} = N_d t_p$ 。每个门的负载电容为 C , 活动因子 $\alpha = 0.05$, 延迟 $t_p = CV_{\text{DD}}/(2I_{\text{on}})$ (式 (9.5)); 关断的门在整个周期中漏电 I_{off} 。

1. 证明每个门每个周期的能量为 $E = \alpha CV_{\text{DD}}^2 \left[1 + \frac{N_d I_{\text{off}}}{2\alpha I_{\text{on}}} \right]$ 。
2. 若 $V_{\text{DD}} < V_T$ (亚阈值逻辑), I_{on} 与 I_{off} 都按 $I_T 10^{(V_{\text{GS}} - V_T)/S}$ 计算 (两者的 $1 - e^{-V_{\text{DS}}/U_T}$ 因子相同)。证明 E 与 V_T 无关, 并写出 $E(V_{\text{DD}})$ 。
3. 取 $S = 80\text{ mV/dec}$, 求使 E 取极小的 V_{DD} 及此时漏电能量与动态能量之比。注意: 该模型在 $V_{\text{DD}} \rightarrow 0$ 时给出 $E \rightarrow 0$, 为什么这不是真正的最优点? (提示: 第 9.1 节的增益条件。)
4. 设 $V_T = 0.35\text{ V}$, $I_T = 1\mu\text{A}$, $C = 1\text{ fF}$ 。求最小能量点的门延迟与时钟频率。与 $V_{\text{DD}} = 1.0\text{ V}$ ($I_{\text{on}} = 100\mu\text{A}$, 漏电可忽略) 相比, 能量与速度各变化多少倍? 用 101 级环形振荡器测这两种电压下的门延迟, 振荡频率各是多少?
5. 若有一种新器件的亚阈值摆幅为 30 mV/dec (假定电路在相应的更低电压下仍能工作), 最优电压与最小能量如何变化? 由此说明“突破 60 mV/dec ”对低功耗电路的意义。

解答

(1) 动态能量 αCV_{DD}^2 (式 (9.4) 除以 f)。漏电能量 $I_{\text{off}} V_{\text{DD}} T_{\text{clk}} = I_{\text{off}} V_{\text{DD}} \cdot N_d CV_{\text{DD}}/(2I_{\text{on}})$ 。相加并提出 αCV_{DD}^2 即得。漏电能量随 $I_{\text{off}}/I_{\text{on}}$ 增大, 而且通路越长 (N_d 大)、活动越少 (α 小), 漏电占比越大。

(2) $I_{\text{on}} = I_T 10^{(V_{\text{DD}} - V_T)/S}$, $I_{\text{off}} = I_T 10^{-V_T/S}$, 比值 $I_{\text{off}}/I_{\text{on}} = 10^{-V_{\text{DD}}/S}$, V_T 与 I_T 都消去:

$$E = \alpha C V_{\text{DD}}^2 \left[1 + B 10^{-V_{\text{DD}}/S} \right],$$

$$B = \frac{N_d}{2\alpha} = 500.$$

V_T 只影响速度, 不影响每个周期的能量。

(3) 令 $x = V_{\text{DD}} \ln 10 / S$, $E \propto x^2(1 + Be^{-x})$ 。求导: $dE/dx \propto x[2 - Be^{-x}(x - 2)]$, 极值条件 $Be^{-x}(x - 2) = 2$ 有两个根: $x = 2.03$ (极大) 与 $x^* = 7.16$ (极小)。于是 $V_{\text{opt}} = x^* S / \ln 10 = 0.249 \text{ V}$, 此时漏电能量 / 动态能量 $= Be^{-x^*} = 2/(x^* - 2) = 0.39$ 。模型中 $V_{\text{DD}} \rightarrow 0$ 时 $E \rightarrow 0$, 但电路早已失效: 由式 (9.3), V_{DD} 只有几个 U_T 时反相器的增益接近 1, 噪声容限消失, 逻辑电平不能恢复; 可靠工作要求中点增益足够大, 例如 $|A| \gtrsim 10$, 由式 (9.3), $V_{\text{DD}} \gtrsim 2U_T \ln(1 + 10m) \approx 0.14 \text{ V}$ ($m = S/(\ln 10 U_T) = 1.34$); Swanson 与 Meindl 估计实际可用的下限约为 $8k_B T/q = 0.21 \text{ V}$ (据该文摘要) [32]。 $V_{\text{opt}} = 0.25 \text{ V}$ 两个条件都满足 (此时 $|A| \approx 91$), 是可用范围内真正的最优点 (脚本在 $[8U_T, 0.6 \text{ V}]$ 上数值求极小核对)。这类“最小能量点”在实际亚阈值电路中确实存在 [35]。

(4) $I_{\text{on}} = 1 \mu\text{A} \times 10^{(0.249 - 0.35)/0.08} = 5.4 \times 10^{-8} \text{ A}$, $t_p = C V_{\text{opt}} / (2I_{\text{on}}) = 2.3 \text{ ns}$, 时钟频率 $1/(N_d t_p) = 8.7 \text{ MHz}$ 。 $V_{\text{DD}} = 1 \text{ V}$ 时 $t_p = 5 \text{ ps}$, 时钟 4 GHz , 每周期能量 $\approx \alpha C \times (1 \text{ V})^2$ 。所以最小能量点省能 $1/[0.249^2 \times 7.16/5.16] = 11.6$ 倍, 但慢了约 460 倍。环形振荡器 $f = 1/(2 \times 101 \times t_p)$: 最小能量点 2.2 MHz , 1 V 时 0.99 GHz 。亚阈值逻辑适合对速度要求不高、能量极其受限的场合 (传感器节点、植入器件)。

(5) x^* 只由 B 决定, 所以 $V_{\text{opt}} \propto S$, $E_{\text{min}} = \alpha C V_{\text{opt}}^2 x^* / (x^* - 2) \propto S^2$ 。 $S = 60 \text{ mV/dec}$ (理想 MOSFET): $V_{\text{opt}} = 0.187 \text{ V}$, E_{min} 降为 $S = 80$ 时的 0.56 倍。这一 V_{opt} 已低于 (3) 中引用的 $8k_B T/q$ 估计, 必须重新检查可靠性: $m = 1$ 时由式 (9.3), $|A| = e^{V_{\text{DD}}/2U_T} - 1 \approx 36$, 仍满足 $|A| \gtrsim 10$ (要求 $V_{\text{DD}} \gtrsim 2U_T \ln 11 = 0.12 \text{ V}$) ——可见 $8k_B T/q$ 不是普适下限, $|A| \gtrsim 10$ 所要求的电压随 m 变化。 $S = 30 \text{ mV/dec}$: $V_{\text{opt}} = 0.093 \text{ V}$, E_{min} 再降为 $1/4$; 这种器件不靠热发射工作, 式 (9.3) 与 $8k_B T/q$ 的估计都不再适用, 模型给出的只是“若电路在该电压下仍能工作”的理想极限。在这个模型里, 能量的下限完全由亚阈值摆幅决定: 对常规 MOSFET, 玻尔兹曼极限 $S \geq \ln 10 \cdot k_B T/q$ 直接给出了每次逻辑操作能量的下限 $\sim \alpha C (k_B T/q)^2 \times \text{常数}$ 。这就是隧穿晶体管、负电容晶体管等“陡峭摆幅”器件被寄予厚望的原因 (第 8.4 节注)。

评分要点 (共 30 分): (1) 两部分能量 4 分; (2) 比值与 V_T 无关 4 分; (3) 极值方程 4 分, 两根与极小 3 分, $V \rightarrow 0$ 的辨析 4 分; (4) 延迟、频率与两种电压的比较 6 分, 环振频率 2 分; (5) 标度 $V_{\text{opt}} \propto S$ 、 $E_{\text{min}} \propto S^2$ 及意义 3 分。

11 实验与数据处理：从电学测量提取 MOSFET 的参数

例题：实验题：一片硅片上的 MOS 电容与 MOSFET（本题使用模拟数据）

样品：同一硅片上的一只 MOS 电容（面积 $A = 100\text{ }\mu\text{m} \times 100\text{ }\mu\text{m}$ ）与两只 n 沟道 MOSFET（ $W = 50\text{ }\mu\text{m}$, $L_1 = 10\text{ }\mu\text{m}$, $L_2 = 20\text{ }\mu\text{m}$ ）， n^+ 多晶硅栅，p 型衬底均匀掺杂，衬底与源相连。全部测量在 300 K 下进行。常数： $\epsilon_s = 11.7\epsilon_0$, $\epsilon_{ox} = 3.9\epsilon_0$, $n_i = 1.0 \times 10^{10}\text{ cm}^{-3}$, $E_g = 1.12\text{ eV}$, $U_T = 25.85\text{ mV}$ 。

测量：

- A. MOS 电容的 1 MHz 电容： $V_G = -4\text{ V}$ （积累）时 $C_{acc} = 22.580\text{ pF}$ ； $V_G = 3\text{ V}$ （强反型）时 $C_{min} = 4.719\text{ pF}$ 。相对误差 0.2%。
- B. 两只 MOSFET 在 $V_{DS} = 50\text{ mV}$ 下的转移特性（表 3）。
- C. L_1 器件在 $V_{DS} = 50\text{ mV}$ 下的亚阈值特性（表 4）。
- D. L_1 器件在 $V_{GS} = 1.8、2.2\text{ V}$ 下的饱和区输出特性（表 5）。

电流测量的误差为 0.3%（相对）加 0.5 pA（绝对）。

说明：本题数据由 `scripts/exp_data.py` 生成：器件电流按渐变沟道近似下的 Pao-Sah 积分（式 (7.2)，沟道电荷用一维 MOS 精确解）计算，模型中包含能量均匀分布的界面陷阱、迁移率随栅压的退化 $\mu = \mu_0/[1 + \theta_0(V_{GS} - V_T)]$ 、源漏两端各 $R_{sd}/2$ 的串联电阻（自洽求解）；饱和区按经验式 $I_D = I_{D0}(1 + \lambda V_{DS})$ ；电容按第 6.8 节的高频模型；再按上述误差加入高斯随机误差。随机数种子事先固定为 20261027。

题量说明：(1)–(5) 为笔试主体（41 分，约 1.5 小时；作为 1 小时笔试时可只做 (1)–(4)）；(6)、(7) 为课后拓展，需要用到第 6.6 节检验框中的模型分析，不计入考试题量。

1. 由测量 A 求 C_{ox} 、 t_{ox} 、强反型时的耗尽层宽度 $W_{d,max}$ 与衬底掺杂 N_A （用 $W_{d,max} = \sqrt{4\epsilon_s\phi_F/qN_A}$ 迭代），并按式 (6.11)（不计氧化层电荷与界面陷阱）预测阈值电压。
2. 线性区模型取 $I_D = \beta V_{DS}(V_{GS} - V_T)/[1 + \theta_{eff}(V_{GS} - V_T)]$, $\beta = \mu_0 C_{ox} W/L$ 。(a) 说明 $V_{DS} = 50\text{ mV}$ 时这一模型略去了什么。(b) 证明“Y 函数” $Y \equiv I_D/\sqrt{g_m} = \sqrt{\beta V_{DS}}(V_{GS} - V_T)$ ，与 θ_{eff} 无关。(c) 用中心差分求 g_m ，对两只器件作 $Y-V_{GS}$ 直线拟合（ $V_{GS} \geq 1.0\text{ V}$ ），求 V_T 、 β 与 μ_0 。
3. 证明源漏串联电阻使 $\theta_{eff} = \theta_0 + \beta R_{sd}$ 。由两只器件的数据求 θ_0 与 R_{sd} 。
4. 由测量 C 求亚阈值摆幅 S 与 m 。利用式 (8.2) 与 (1)(2) 的结果，由 $V_{GS} = 0.10\text{ V}$ 处电流的大小自洽地估计此时的表面势 ψ_s ，进而求 C_d 与界面陷阱密度 D_{it} （计入第 8.3 节检验中 $C_d(\psi_s)$ 的修正）。
5. 由测量 D 求沟道长度调制系数 λ 、饱和区跨导、输出电阻与本征增益（取 $V_{GS} = 2.0\text{ V}$ 、 $V_{DS} = 3\text{ V}$ ）。用 (2)(3) 的参数预测饱和电流，检验是否自洽。
6. (课后拓展) 比较 (1) 与 (2) 的阈值电压，定量说明差别的来源。最大跨导外推法（在 g_m 最大处作切线外推到 $I_D = 0$ ）给出的阈值是多少？为什么不可靠？

7. (课后拓展) 讨论系统误差: (a) 积累区电容并不等于 C_{ox} ; (b) 强反型时表面势比 $2\phi_F$ 高 (第 6.6 节), 据此修正 N_A 与 D_{it} ; (c) 其他误差来源与改进方案。

表 3 测量 B: $V_{DS} = 50 \text{ mV}$ 时的漏极电流 (模拟数据)

V_{GS}/V	$I_D/\mu\text{A}$ ($L_1 = 10 \mu\text{m}$)	$I_D/\mu\text{A}$ ($L_2 = 20 \mu\text{m}$)
0.8	7.180	3.634
1.0	11.208	5.717
1.2	15.071	7.760
1.4	18.722	9.773
1.6	22.313	11.579
1.8	25.540	13.549
2.0	28.776	15.233
2.2	31.947	16.944
2.4	34.709	18.654
2.6	37.574	20.209
2.8	40.071	21.690
3.0	42.464	23.114

表 4 测量 C: L_1 器件在 $V_{DS} = 50 \text{ mV}$ 下的亚阈值电流 (模拟数据)

V_{GS}/V	I_D/A	V_{GS}/V	I_D/A
0.00	1.799×10^{-11}	0.25	1.192×10^{-8}
0.05	6.606×10^{-11}	0.30	4.300×10^{-8}
0.10	2.406×10^{-10}	0.35	1.431×10^{-7}
0.15	8.734×10^{-10}	0.40	3.960×10^{-7}
0.20	3.239×10^{-9}		

表 5 测量 D: L_1 器件的饱和区输出特性 (模拟数据)

V_{DS}/V	$I_D/\mu\text{A}$ ($V_{GS} = 1.8 \text{ V}$)	$I_D/\mu\text{A}$ ($V_{GS} = 2.2 \text{ V}$)
2.0	282.9	457.1
2.5	285.3	462.8
3.0	289.4	466.2
3.5	291.5	473.5
4.0	295.2	475.9
4.5	298.2	481.2
5.0	299.6	485.4

解答

数据处理由 `scripts/exp_fit.py` 完成；直线拟合用最小二乘法（附录 D），拟合参数的误差由残差估计每点误差后得到。

(1) $C_{\text{ox}} \approx C_{\text{acc}}/A = 2.258 \times 10^{-3} \text{ F/m}^2$, $t_{\text{ox}} = \varepsilon_{\text{ox}}/C_{\text{ox}} = 15.29 \text{ nm}$ 。由式 (6.15), $W_{\text{d,max}} = \varepsilon_{\text{s}}(A/C_{\text{min}} - 1/C_{\text{ox}}) = 173.6 \text{ nm}$ 。迭代：从 $N_{\text{A}} = 10^{17} \text{ cm}^{-3}$ 出发, $\phi_{\text{F}} = 0.417 \rightarrow 0.390 \rightarrow 0.388 \text{ V}$, $N_{\text{A}} = 3.57 \rightarrow 3.35 \rightarrow 3.33 \times 10^{16} \text{ cm}^{-3}$, 收敛到

$$N_{\text{A}} = (3.33 \pm 0.02) \times 10^{16} \text{ cm}^{-3},$$

$$\phi_{\text{F}} = 0.388 \text{ V}$$

(误差只计电容的 0.2%，用蒙特卡洛传递)。 $V_{\text{FB}} = -(0.56 + 0.388) = -0.948 \text{ V}$, $\gamma = 0.466 \text{ V}^{1/2}$, $V_{\text{T}} = -0.948 + 0.777 + 0.466\sqrt{0.777} = 0.239 \text{ V}$ 。

(2)(a) 线性区的完整式是式 (7.9): $I_{\text{D}} = \beta[(V_{\text{GS}} - V_{\text{T}}) - mV_{\text{DS}}/2]V_{\text{DS}}$ 。题给模型略去了 $mV_{\text{DS}}/2 \approx 0.03 \text{ V}$ ，相当于把阈值移高 0.03 V； V_{DS} 越小越好，但电流太小又会增大相对误差，50 mV 是常用的折中。

(b) 记 $V_{\text{ov}} = V_{\text{GS}} - V_{\text{T}}$, $g_{\text{m}} = \frac{dI_{\text{D}}}{dV_{\text{GS}}} = \frac{\beta V_{\text{DS}}}{(1 + \theta_{\text{eff}} V_{\text{ov}})^2}$, 所以 $\sqrt{g_{\text{m}}} = \sqrt{\beta V_{\text{DS}}}/(1 + \theta_{\text{eff}} V_{\text{ov}})$, $I_{\text{D}}/\sqrt{g_{\text{m}}} = \sqrt{\beta V_{\text{DS}}} V_{\text{ov}}$: 分母中的 $1 + \theta_{\text{eff}} V_{\text{ov}}$ 恰好消去 [36]。

(c) g_{m} 用相邻两点的中心差分 (得到 $V_{\text{GS}} = 1.0\text{--}2.8 \text{ V}$ 共 10 个点), 拟合 $Y\text{--}V_{\text{GS}}$ 直线 (图 15(a)):

器件	V_{T}/V	$\beta/(10^{-4} \text{ A/V}^2)$	$\mu_0 = \beta L/(WC_{\text{ox}})$
$L_1 = 10 \mu\text{m}$	0.500 ± 0.017	4.80 ± 0.10	$425 \text{ cm}^2/(\text{V s})$
$L_2 = 20 \mu\text{m}$	0.499 ± 0.015	2.402 ± 0.046	$426 \text{ cm}^2/(\text{V s})$

两只器件的 V_{T} 一致, $\beta_1/\beta_2 = 2.000$ 与 L_2/L_1 相符, 说明有效沟道长度可取版图长度, Y 函数确实消除了串联电阻的影响 (两只器件的 R_{sd} 相同而 β 不同)。 μ_0 约为体硅电子迁移率的 1/3, 是反型层迁移率的典型值。

(3) 源端电阻 $R_{\text{sd}}/2$ 使本征栅压变为 $V_{\text{GS}} - I_{\text{D}}R_{\text{sd}}/2$, 线性区中 $I_{\text{D}}R_{\text{sd}}/2 \sim 4 \text{ mV} \ll V_{\text{ov}}$, 可以忽略；两端电阻使本征漏源电压变为 $V_{\text{DS}} - I_{\text{D}}R_{\text{sd}}$, 这一项不能忽略 (与 V_{DS} 本身可比)。于是 $I_{\text{D}}(1 + \theta_0 V_{\text{ov}}) = \beta V_{\text{ov}}(V_{\text{DS}} - I_{\text{D}}R_{\text{sd}})$, 移项得 $I_{\text{D}}[1 + (\theta_0 + \beta R_{\text{sd}})V_{\text{ov}}] = \beta V_{\text{ov}}V_{\text{DS}}$, 即 $\theta_{\text{eff}} = \theta_0 + \beta R_{\text{sd}}$ 。由 (2) 的模型, $\beta V_{\text{DS}}V_{\text{ov}}/I_{\text{D}} - 1 = \theta_{\text{eff}}V_{\text{ov}}$, 对 $V_{\text{GS}} \geq 1.2 \text{ V}$ 的点作过原点附近的直线 (图 15(b)): $\theta_{\text{eff},1} = (0.162 \pm 0.002) \text{ V}^{-1}$, $\theta_{\text{eff},2} = (0.117 \pm 0.002) \text{ V}^{-1}$ 。两个方程两个未知数:

$$R_{\text{sd}} = \frac{\theta_{\text{eff},1} - \theta_{\text{eff},2}}{\beta_1 - \beta_2} = (187 \pm 12) \Omega,$$

$$\theta_0 = \theta_{\text{eff},1} - \beta_1 R_{\text{sd}} = (0.072 \pm 0.005) \text{ V}^{-1}.$$

这里两只不同长度的器件提供了两个独立的观测量 $\theta_{\text{eff},1}$ 、 $\theta_{\text{eff},2}$, 才能分开两个未知数；只用一只器件无法区分迁移率退化与串联电阻。

(4) 对 $V_{\text{GS}} = 0\text{--}0.25 \text{ V}$ 的 6 个点拟合 $\log_{10} I_{\text{D}}\text{--}V_{\text{GS}}$ 直线 (更高的点已接近阈值, 曲线弯曲), 斜率的倒数 $S = (88.7 \pm 0.1) \text{ mV/dec}$, $m = S/(\ln 10 \cdot U_{\text{T}}) = 1.490$ 。

估计 ψ_{s} : 拟合直线在 $V_{\text{GS}} = 0.10 \text{ V}$ 处 $I_{\text{D}} = 2.40 \times 10^{-10} \text{ A}$ 。由式 (8.2) (ψ_{s} 换成 $\psi_{\text{s}} - U_{\text{T}}$ 计算 C_{d}), 以 $\mu_0 = 425 \text{ cm}^2/(\text{V s})$ 、 $N_{\text{A}} = 3.33 \times 10^{16} \text{ cm}^{-3}$ 代入, ψ_{s} 只出现在 $e^{\psi_{\text{s}}/U_{\text{T}}}$ 与 C_{d} 中, 迭代两三次即收敛: $\psi_{\text{s}} = 0.626 \text{ V}$, 介于 ϕ_{F} 与 $2\phi_{\text{F}}$ 之间 (弱反型), $C_{\text{d}}/C_{\text{ox}} = 0.301$ 。修正因

子 $1/[1 - U_T/2(\psi_s - U_T)] = 1.022$, 所以

$$\frac{C_{it}}{C_{ox}} = \frac{m}{1.022} - 1 - \frac{C_d}{C_{ox}} = 0.156,$$

$$D_{it} = \frac{0.156 C_{ox}}{q} = 2.2 \times 10^{11} \text{ cm}^{-2} \text{ eV}^{-1}.$$

若不作修正则得 2.7×10^{11} : 在这种精度上, $C_d(\psi_s)$ 前因子的效应不能忽略。 S 的统计误差只贡献 $\pm 0.03 \times 10^{11}$, 结果的不确定度主要来自 N_A (见 (7))。

(5) 对两组数据分别拟合 $I_D = I_{D0} + (dI_D/dV_{DS})V_{DS}$, $\lambda = (dI_D/dV_{DS})/I_{D0}$: $V_{GS} = 1.8 \text{ V}$ 时 $I_{D0} = 271.3 \mu\text{A}$, $\lambda = (0.0215 \pm 0.0010) \text{ V}^{-1}$; 2.2 V 时 $I_{D0} = 438.9 \mu\text{A}$, $\lambda = (0.0214 \pm 0.0009) \text{ V}^{-1}$; 两者一致, “厄利电压” $1/\lambda \approx 47 \text{ V}$ 。 $V_{DS} = 3 \text{ V}$ 处 $g_m \approx (I_{2.2} - I_{1.8})/0.4 \text{ V} = 0.446 \text{ mS}$, $V_{GS} = 2.0 \text{ V}$ 处 $I_D \approx 378 \mu\text{A}$, $r_o = (\lambda I_{D0})^{-1} = 131 \text{ k}\Omega$, 本征增益 $g_m r_o \approx 58$ 。

自洽性: 用 (2)(3) 的 β_1 、 $\theta_{\text{eff},1}$ 、 V_T 与 (1) 的 $m = 1.264$, 按体电荷平方律 $I_{D,\text{sat}} = \beta V_{\text{ov}}^2/[2m(1 + \theta_{\text{eff}} V_{\text{ov}})]$ 预测 $V_{GS} = 1.8$ 、 2.2 V 的饱和电流为 265、431 μA , 与 I_{D0} 相差约 2%。(饱和区中 θ_{eff} 的串联电阻部分并不严格适用, 2% 的偏差在预期之内。)

(6) Y 函数阈值 0.500 V 比 (1) 的预测 0.239 V 高 0.26 V。逐项分析:

- 用 (7) 修正后的 N_A 、 t_{ox} , 式 (6.11) 给出 0.277 V (+0.04 V)。
- 界面陷阱: 若陷阱的电中性点在禁带中央 ($\psi_s = \phi_F$), 到阈值 ($\psi_s = 2\phi_F$) 时陷阱多带了 $qD_{it}\phi_F$ 的负电荷, 阈值升高 $qD_{it}\phi_F/C_{ox} \approx 0.06 \text{ V}$, 得 0.337 V。
- 强反型表面势的偏移: 记 $k = 1 + (C_d + C_{it})/C_{ox} \approx 1.42$, 强反型时 $Q_n/C_{ox} \approx -[V_{GS} - V_T - k\Delta(V_{GS})]$, $\Delta = 2U_T \ln(C_{ox}V_{\text{ov}}/\sqrt{2\varepsilon_s k_B T N_A})$ (第 6.6 节)。在拟合区间中点 $V_{\text{ov}} \approx 1.5 \text{ V}$ 处作切线外推到零, 由于 $d\Delta/dV_{GS} = 2U_T/V_{\text{ov}}$, 截距移动 $k(\Delta - 2U_T) = 1.42 \times (0.151 - 0.052) \approx 0.14 \text{ V}$, 得 0.48 V。
- 线性区的 $mV_{DS}/2 \approx 0.03 \text{ V}$, 得 0.51 V, 与 Y 函数的 0.500 V 相差 0.01 V。

最大跨导外推法: g_m 在 $V_{GS} = 1.0 \text{ V}$ 处最大 (迁移率退化与串联电阻使 g_m 很早就开始下降), 切线外推给出 L_1 : 0.432 V, L_2 : 0.446 V——两只本应相同的器件差了 14 mV, 而且都比 Y 函数低, 因为 θ_{eff} 使 $I_D(V_{GS})$ 向下弯, 切线的截距左移, 左移量又与 θ_{eff} (含 βR_{sd} , 与 L 有关) 有关。Y 函数在一阶上不受这两者影响。

(7) (a) 积累层也有有限的电容: $C_{s,\text{acc}} \approx |Q_s|/2U_T$ (式 (6.6)), $|Q_s| \approx C_{\text{acc}}|V_G - V_{\text{FB}}| = 6.9 \times 10^{-3} \text{ C/m}^2$, $C_{s,\text{acc}} \approx 0.13 \text{ F/m}^2$ 。由 $1/C_{ox} = 1/C_{\text{acc}} - 1/C_{s,\text{acc}}$, $C_{ox} = 2.297 \times 10^{-3} \text{ F/m}^2$, $t_{\text{ox}} = 15.03 \text{ nm}$ (修正 1.7%)。

(b) 强反型 ($V_G = 3 \text{ V}$) 时 $\psi_s = 2\phi_F + \Delta$, $\Delta = 2U_T \ln(C_{ox}(3 - V_T)/\sqrt{2\varepsilon_s k_B T N_A}) \approx 0.18 \text{ V}$; 高频小信号对应的耗尽层宽度由式 (6.7) 为 $W_d = \sqrt{2\varepsilon_s(\psi_s - U_T)/qN_A}$ 。用修正后的 C_{ox} 重新迭代 (Δ 也随 N_A 更新), 得 $N_A = 3.99 \times 10^{16} \text{ cm}^{-3}$, 比 (1) 大 20%——远大于 0.6% 的统计误差: 这一测量的误差完全由模型决定。用修正后的 N_A 、 C_{ox} 重做 (4): $\psi_s = 0.633 \text{ V}$, $C_d/C_{ox} = 0.321$, $D_{it} = 2.0 \times 10^{11} \text{ cm}^{-2} \text{ eV}^{-1}$ 。(生成数据所用的“真值”为 $t_{\text{ox}} = 15.00 \text{ nm}$, $N_A = 4.0 \times 10^{16} \text{ cm}^{-3}$, $D_{it} = 2.0 \times 10^{11}$, $\mu_0 = 450 \text{ cm}^2/(\text{Vs})$, $\theta_0 = 0.080 \text{ V}^{-1}$, $R_{\text{sd}} = 200 \Omega$, $\lambda = 0.020 \text{ V}^{-1}$: 修正后的 t_{ox} 、 N_A 、 D_{it} 与真值相差 2% 以内; R_{sd} 、 θ_0 、 λ 分别偏离 1.1、1.7、1.5 个标准差。)

(c) 其他系统误差与改进：

- μ_0 偏低约 5%：强反型时 $-dQ_n/dV_{GS}$ 不是 C_{ox} 而约是 $0.97C_{ox}$ (反型层电容, 第 6.6 节), Y 函数的斜率把它算进了 β 。改进：用“分离 $C-V$ ”法 (源漏短接, 测栅到沟道的电容并积分) 直接测 $Q_n(V_{GS})$, 再由线性区电导求迁移率。
- 有效沟道长度、宽度与版图尺寸的差别：用多种 L (不止两种) 作 $1/\beta-L$ 图可以同时求出 ΔL 与 R_{sd} 。
- 温度： $S \propto T$, V_T 有 $\sim -1\text{mV/K}$ 的温度系数 (例题 2), 测量中器件自热、环境温度变化都会带来误差；在两个温度下测 S , 可以检验 m 是否与温度无关, 从而检验 D_{it} 的模型。
- D_{it} 的结果依赖于“能量均匀分布”与“中性点在禁带中央”两个假设；更直接的方法是比较高、低频 $C-V$ 曲线 (界面陷阱只在低频下响应)。

评分要点 (共 50 分)：(1) t_{ox} 、 $W_{d,max}$ 4 分, N_A 的迭代 3 分, V_T 预测 2 分；(2) 略去项 2 分, Y 函数的证明 3 分, 拟合与 μ_0 5 分；(3) θ_{eff} 公式的推导 3 分, 两器件联立 4 分；(4) S 、 m 3 分, ψ_s 的自洽估计 3 分, D_{it} 3 分；(5) λ 、 r_o 、增益 4 分, 自洽检验 2 分；(6) 阈值差别的四项 4 分, 最大跨导法的问题 2 分；(7) 两项修正 2 分, 其余讨论 1 分。

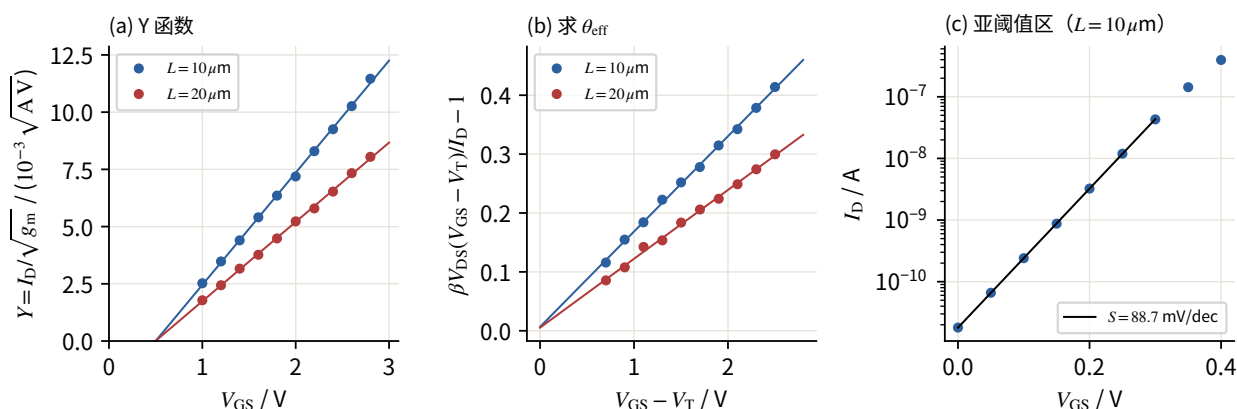


图 15 实验题的数据处理 (scripts/exp_fit.py)。(a) Y 函数与直线拟合, 两条直线交横轴于同一点 ($V_T \approx 0.50\text{ V}$), 斜率之比为 $\sqrt{2}$; (b) $\beta V_{DS}V_{ov}/I_D - 1$ 与 V_{ov} 的直线, 斜率即 θ_{eff} , 短沟道器件斜率更大 (串联电阻的贡献 βR_{sd} 更大); (c) 亚阈值区与 $S = 88.7\text{ mV/dec}$ 的拟合直线。

12 小结：推导链

1. **载流子统计**: 玻尔兹曼因子 + 泡利原理 $\Rightarrow$ 费米-狄拉克分布 (4.1); 平衡时费米能级处处相等。态密度 (4.3) $\Rightarrow n = N_c e^{-(E_c - E_F)/k_B T}$ (式 (4.4)) $\Rightarrow np = n_i^2$ (式 (4.6)) $\Rightarrow n = n_i e^{(E_F - E_i)/k_B T}$ (式 (4.8)) $\Rightarrow$ 费米势 $\phi_F = U_T \ln(N_A/n_i)$ (式 (4.10))。
2. **输运与静电**: 漂移 + 扩散 (5.1), 平衡零电流 $\Rightarrow$ 爱因斯坦关系 (5.2) $\Rightarrow J_n = \mu_n n dE_{Fn}/dx$; 能带图即电势图 (5.4); 高斯定理 $\Rightarrow$ 泊松方程 (5.6) $\Rightarrow$ 德拜屏蔽 (5.7)。

3. **MOS 电容**: 功函数差与固定电荷 $\Rightarrow V_{FB}$ (式 (6.2)); 高斯定理 $\Rightarrow$ 栅压方程 (6.3); 泊松方程第一积分 $\Rightarrow Q_s(\psi_s)$ (式 (6.5)) $\Rightarrow$ 耗尽近似 (6.9)、强反型判据 $\psi_s = 2\phi_F \Rightarrow V_T$ (式 (6.11)) 与 $Q_n = -C_{ox}(V_G - V_T)$ (式 (6.12)); 阈值上偏 0.1 V 量级, $V_G - V_T \lesssim 0.5$ V 时 $|Q_n|$ 被高估 30% 以上; 界面陷阱 $\Rightarrow$ 电容分压 (6.14); 串联电容 (6.15) $\Rightarrow C-V$ 曲线。
4. **MOSFET**: 准费米能级 $\Rightarrow I_D = -W\mu Q_n dV/dy$ (式 (7.1)) $\Rightarrow$ (渐变沟道近似) $I_D = (W\mu/L) \int (-Q_n) dV$ (式 (7.2)) $\Rightarrow$ 平方律 (7.5)–(7.6) 与夹断 $\Rightarrow$ 体因子 m 与体效应 (7.9)、(7.10) $\Rightarrow$ 沟道长度调制 (7.11); 速度饱和 (7.13) $\Rightarrow I_{D,sat} \rightarrow WC_{ox}v_{sat}(V_{GS} - V_T)$ (式 (7.16))。
5. **亚阈值**: 弱反型电荷 $\propto e^{(\psi_s - V)/U_T}$ (式 (8.1)) $\Rightarrow$ 扩散电流 (8.2) $\Rightarrow S = \ln 10 \cdot U_T(1 + (C_d + C_{it})/C_{ox}) \geq 60 \text{ mV/dec}$ (式 (8.4)) $\Rightarrow I_{off} = I_T 10^{-V_T/S}$ (式 (8.5))。
6. **CMOS 与缩小**: 反相器 V_M 、噪声容限 (9.1)、(9.2), 亚阈值增益 (9.3); 能量守恒 $\Rightarrow P = \alpha C_L V_{DD}^2 f$ (式 (9.4)); 泊松方程的标度不变性 $\Rightarrow$ Dennard 规则 (表 2); S 与内建电势不缩小、隧穿 $\Rightarrow$ 失效; 二维拉普拉斯方程 $\Rightarrow$ 自然长度 (9.6) $\Rightarrow$ 短沟道效应 $\propto e^{-L/2\ell}$ (式 (9.7)、(9.8)) $\Rightarrow$ 多栅结构。

要点：一句话

栅压通过电容分压弯曲半导体表面的能带；能带每多弯 U_T ，表面电子就多 e 倍。弯够 $2\phi_F$ 之后表面势被钉住，栅压全部用来积累反型电子，电流随栅压平方（或线性）增长；弯不够时，电子要靠热运动越过势垒，电流随栅压指数增长，但不可能快于每 $\ln 10 \cdot k_B T/q \approx 60 \text{ mV}$ 十倍——这一条玻尔兹曼极限，最终决定了芯片的电源电压与功耗。

参考文献

- [1] Julius Edgar Lilienfeld. Method and apparatus for controlling electric currents. U.S. Patent 1,745,175 (filed Oct. 8, 1926; Canadian priority Oct. 22, 1925; granted Jan. 28, 1930), 1930. URL <https://patents.google.com/patent/US1745175A/en>.
- [2] Computer History Museum. 1926: Field effect semiconductor device concepts patented. The Silicon Engine (online exhibit), accessed Oct. 2026, . URL <https://www.computerhistory.org/siliconengine/field-effect-semiconductor-device-concepts-patented/>.
- [3] Oskar Heil. Improvements in or relating to electrical amplifiers and other control arrangements and devices. British Patent 439,457 (convention date in Germany Mar. 2, 1934; application date in UK Mar. 4, 1935, No. 6815 (per the printed specification; CHM gives Mar. 5); complete specification accepted Dec. 6, 1935), 1935. URL <https://worldwide.espacenet.com/patent/search?q=pn%3DGB439457A>.
- [4] Computer History Museum. 1947: Invention of the point-contact transistor. The Silicon Engine (online exhibit), accessed Oct. 2026, . URL <https://www.computerhistory.org/siliconengine/invention-of-the-point-contact-transistor/>.

- [5] William Shockley. Transistor technology evokes new physics. Nobel Lecture, December 11, 1956, 1956. URL <https://www.nobelprize.org/uploads/2018/06/shockley-lecture.pdf>.
- [6] John Bardeen. Semiconductor research leading to the point contact transistor. Nobel Lecture, December 11, 1956, 1956. URL <https://www.nobelprize.org/uploads/2018/06/bardeen-lecture.pdf>.
- [7] J. Bardeen. Surface states and rectification at a metal semi-conductor contact. *Physical Review*, 71:717–727, 1947. doi: 10.1103/PhysRev.71.717.
- [8] J. Bardeen and W. H. Brattain. The transistor, a semi-conductor triode. *Physical Review*, 74:230–231, 1948. doi: 10.1103/PhysRev.74.230.
- [9] W. Shockley and G. L. Pearson. Modulation of conductance of thin films of semi-conductors by surface charges. *Physical Review*, 74:232–233, 1948. doi: 10.1103/PhysRev.74.232.
- [10] W. Shockley. A unipolar “field-effect” transistor. *Proceedings of the IRE*, 40(11):1365–1376, 1952. doi: 10.1109/JRPROC.1952.273964.
- [11] M. M. Atalla, E. Tannenbaum, and E. J. Scheibner. Stabilization of silicon surfaces by thermally grown oxides. *Bell System Technical Journal*, 38(3):749–783, 1959. doi: 10.1002/j.1538-7305.1959.tb03907.x.
- [12] Computer History Museum. 1960: Metal oxide semiconductor (MOS) transistor demonstrated. The Silicon Engine (online exhibit), accessed Oct. 2026, . URL <https://www.computerhistory.org/siliconengine/metal-oxide-semiconductor-mos-transistor-demonstrated/>.
- [13] Dawon Kahng. A historical perspective on the development of MOS transistors and related devices. *IEEE Transactions on Electron Devices*, 23(7):655–657, 1976. doi: 10.1109/T-ED.1976.18468.
- [14] Dawon Kahng. Electric field controlled semiconductor device. U.S. Patent 3,102,230 (filed May 31, 1960; granted Aug. 27, 1963), 1963. URL <https://patents.google.com/patent/US3102230A/en>.
- [15] F. M. Wanlass and C. T. Sah. Nanowatt logic using field-effect metal-oxide semiconductor triodes. In *1963 IEEE International Solid-State Circuits Conference, Digest of Technical Papers*, pages 32–33, 1963. doi: 10.1109/ISSCC.1963.1157450.
- [16] Computer History Museum. 1963: Complementary MOS circuit configuration is invented. The Silicon Engine (online exhibit), accessed Oct. 2026, . URL <https://www.computerhistory.org/siliconengine/complementary-mos-circuit-configuration-is-invented/>.
- [17] Gordon E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8):114–117, 1965. doi: 10.1109/JPROC.1998.658762. Reprinted in *Proc. IEEE* 86, 82–85 (1998).

- [18] Gordon E. Moore. Progress in digital integrated electronics. In *International Electron Devices Meeting, Technical Digest*, pages 11–13, 1975.
- [19] Computer History Museum. 1965: “Moore’s Law” predicts the future of integrated circuits. The Silicon Engine (online exhibit), accessed Oct. 2026, . URL <https://www.computerhistory.org/siliconengine/moores-law-predicts-the-future-of-integrated-circuits/>.
- [20] Robert H. Dennard, Fritz H. Gaensslen, Hwa-Nien Yu, V. Leo Rideout, Ernest Bassous, and Andre R. LeBlanc. Design of ion-implanted MOSFET’s with very small physical dimensions. *IEEE Journal of Solid-State Circuits*, 9(5):256–268, 1974. doi: 10.1109/JSSC.1974.1050511.
- [21] National Research Council. *The Future of Computing Performance: Game Over or Next Level?* The National Academies Press, Washington, DC, 2011. doi: 10.17226/12980.
- [22] Mark Bohr. A 30 year retrospective on Dennard’s MOSFET scaling paper. *IEEE Solid-State Circuits Society Newsletter*, 12(1):11–13, 2007. doi: 10.1109/N-SSC.2007.4785534.
- [23] K. Mistry et al. A 45nm logic technology with high-k+metal gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging. In *IEEE International Electron Devices Meeting, Technical Digest*, pages 247–250, 2007. doi: 10.1109/IEDM.2007.4418914.
- [24] Intel Corporation. Intel reinvents transistors using new 3-D structure. Press release (May 4), 2011. URL <https://www.intc.com/news-events/press-releases/detail/655/intel-reinvents-transistors-using-new-3-d-structure>.
- [25] Samsung Electronics. Samsung begins chip production using 3nm process technology with GAA architecture. Press release (June 30), 2022. URL <https://news.samsungsemiconductor.com/global/samsung-begins-chip-production-using-3nm-process-technology-with-gaa-architecture/>.
- [26] Ioffe Institute. NSM archive: Physical properties of semiconductors — silicon (Si). Basic parameters, band structure and electrical properties at 300 K; accessed Oct. 2026. URL <https://www.ioffe.ru/SVA/NSM/Semicond/Si/>.
- [27] S. M. Sze and Kwok K. Ng. *Physics of Semiconductor Devices*. Wiley, Hoboken, NJ, 3rd edition, 2007. ISBN 978-0-471-14323-9.
- [28] Yuan Taur and Tak H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, Cambridge, 2nd edition, 2009. ISBN 978-0-521-83294-1.
- [29] H. C. Pao and C. T. Sah. Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors. *Solid-State Electronics*, 9(10):927–937, 1966. doi: 10.1016/0038-1101(66)90068-2.
- [30] Adrian M. Ionescu and Heike Riel. Tunnel field-effect transistors as energy-efficient electronic switches. *Nature*, 479:329–337, 2011. doi: 10.1038/nature10679.

- [31] Sayeef Salahuddin and Supriyo Datta. Use of negative capacitance to provide voltage amplification for low power nanoscale devices. *Nano Letters*, 8(2):405–410, 2008. doi: 10.1021/nl071804g.
- [32] Richard M. Swanson and James D. Meindl. Ion-implanted complementary MOS transistors in low-voltage circuits. *IEEE Journal of Solid-State Circuits*, 7(2):146–153, 1972. doi: 10.1109/JSSC.1972.1050260.
- [33] S. Thompson et al. A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low k ILD, and 1 μm^2 SRAM cell. In *IEEE International Electron Devices Meeting, Digest*, pages 61–64, 2002. doi: 10.1109/IEDM.2002.1175779.
- [34] J.-P. Colinge. The new generation of SOI MOSFETs. *Romanian Journal of Information Science and Technology*, 11(1):3–15, 2008. URL <http://www.romjist.ro/content/pdf/01-colinge.pdf>.
- [35] Benton H. Calhoun, Alice Wang, and Anantha Chandrakasan. Modeling and sizing for minimum energy operation in subthreshold circuits. *IEEE Journal of Solid-State Circuits*, 40(9):1778–1786, 2005. doi: 10.1109/JSSC.2005.852162.
- [36] G. Ghibaudo. New method for the extraction of MOSFET parameters. *Electronics Letters*, 24(9):543–545, 1988. doi: 10.1049/el:19880369.
- [37] D. J. Frank, Y. Taur, and H.-S. P. Wong. Generalized scale length for two-dimensional effects in MOSFETs. *IEEE Electron Device Letters*, 19(10):385–387, 1998. doi: 10.1109/55.720194.

A 计算细节与插值模型

A.1 态密度积分与有效态密度

需要的积分是 $I_{1/2} \equiv \int_0^\infty x^{1/2} e^{-x} dx$ 。令 $x = t^2$, $dx = 2t dt$:

$$I_{1/2} = 2 \int_0^\infty t^2 e^{-t^2} dt.$$

分部积分 ($u = t$, $dv = te^{-t^2} dt$, $v = -\frac{1}{2}e^{-t^2}$): $2 \int_0^\infty t^2 e^{-t^2} dt = 2 \left[-\frac{t}{2} e^{-t^2} \right]_0^\infty + \int_0^\infty e^{-t^2} dt = \frac{\sqrt{\pi}}{2}$, 最后一步用了高斯积分 $\int_0^\infty e^{-t^2} dt = \sqrt{\pi}/2$ (把它的平方写成平面上的二重积分, 改用极坐标即得)。代入第 4.3 节:

$$\begin{aligned} n &= \frac{1}{2\pi^2} \left(\frac{2m^* k_B T}{\hbar^2} \right)^{3/2} \frac{\sqrt{\pi}}{2} e^{-(E_c - E_F)/k_B T} \\ &= 2 \left(\frac{m^* k_B T}{2\pi \hbar^2} \right)^{3/2} e^{-(E_c - E_F)/k_B T}, \end{aligned}$$

其中用了 $\frac{1}{2\pi^2} \cdot 2^{3/2} \cdot \frac{\sqrt{\pi}}{2} = 2 \cdot (2\pi)^{-3/2}$ 。再用 $\hbar = h/2\pi$ 即得式 (4.4)。同样的积分给出导带电子的平均动能 $\langle E - E_c \rangle = k_B T I_{3/2}/I_{1/2} = \frac{3}{2} k_B T$ ($I_{3/2} = \frac{3}{2} I_{1/2}$, 分部积分), 与经典理想气体相同 (图 3(b))。

数值: $m^* = m_0$ 、 $T = 300\text{ K}$ 时 $2(2\pi m_0 k_B T / h^2)^{3/2} = 2.51 \times 10^{25} \text{ m}^{-3} = 2.51 \times 10^{19} \text{ cm}^{-3}$;
 $N_c \propto T^{3/2}$ 。

A.2 速度饱和模型的饱和点

由第 7.9 节第三步, $I_D(L + V_{DS}/\mathcal{E}_{\text{sat}}) = W\mu C_{\text{ox}}g(V_{DS})$, $g(V) = V_{\text{ov}}V - mV^2/2$ 。两边对 V_{DS} 求导:

$$\frac{dI_D}{dV_{DS}} \left(L + \frac{V_{DS}}{\mathcal{E}_{\text{sat}}} \right) + \frac{I_D}{\mathcal{E}_{\text{sat}}} = W\mu C_{\text{ox}}(V_{\text{ov}} - mV_{DS}).$$

在极大值处 $dI_D/dV_{DS} = 0$, 所以 $I_{D,\text{sat}} = W\mu\mathcal{E}_{\text{sat}}C_{\text{ox}}(V_{\text{ov}} - mV_{DS,\text{sat}}) = WC_{\text{ox}}v_{\text{sat}}(V_{\text{ov}} - mV_{DS,\text{sat}})$, 这就是式 (7.15) 的最后一式。把它代回原式:

$$\begin{aligned} \mathcal{E}_{\text{sat}}(V_{\text{ov}} - mV) \left(L + \frac{V}{\mathcal{E}_{\text{sat}}} \right) &= V_{\text{ov}}V - \frac{m}{2}V^2 \\ \implies \frac{m}{2}V^2 + m\mathcal{E}_{\text{sat}}LV - V_{\text{ov}}\mathcal{E}_{\text{sat}}L &= 0, \end{aligned}$$

取正根即得式 (7.15) 的第一式 ($u = 2V_{\text{ov}}/(m\mathcal{E}_{\text{sat}}L)$)。物理解释: 漏端局部有 $I_D = W|Q_n(L)|v(\mathcal{E}_L)$, $|Q_n(L)| = C_{\text{ox}}(V_{\text{ov}} - mV_{DS})$; 与上面的 $I_{D,\text{sat}}$ 比较得 $v(\mathcal{E}_L) = v_{\text{sat}}$, 在式 (7.13) 中这要求 $\mathcal{E}_L \rightarrow \infty$ ——与长沟道模型漏端电场发散 (式 (7.7)) 同理, 模型在饱和点的漏端都失效, 但给出的电流是可靠的。两个极限的展开: 令 $u = 2V_{\text{ov}}/(m\mathcal{E}_{\text{sat}}L)$, $\sqrt{1+u} - 1 = u/2 - u^2/8 + \dots$ 。 $u \ll 1$ 时 $V_{DS,\text{sat}} \approx V_{\text{ov}}/m - V_{\text{ov}}^2/(2m^2\mathcal{E}_{\text{sat}}L)$, $I_{D,\text{sat}} \approx \frac{\mu C_{\text{ox}}W}{2mL}V_{\text{ov}}^2 \left(1 - \frac{V_{\text{ov}}}{m\mathcal{E}_{\text{sat}}L} \right)$; $u \gg 1$ 时得式 (7.16)。

A.3 CMOS 计算与演示所用的插值器件模型

图 12、第 9.1 节的数值与 mosfet-transfer、mosfet-cmos 两个演示, 需要一个从亚阈值区连续过渡到强反型区的器件模型。本讲用 EKV 型插值式 (`scripts/cmoslib.py`):

$$\begin{aligned} I_D &= 2mkU_T^2 \left[F\left(\frac{V_P}{U_T}\right) - F\left(\frac{V_P - V_{DS}}{U_T}\right) \right] (1 + \lambda V_{DS}), \\ F(z) &= \ln^2(1 + e^{z/2}), \quad V_P = \frac{V_{GS} - V_T}{m}, \end{aligned}$$

其中 $k = \mu C_{\text{ox}}W/L$, 源极接地 ($V_S = 0$)。两个极限:

- **强反型** ($z \gg 1$): $\ln(1 + e^{z/2}) \approx z/2$, $F \approx z^2/4$ 。若 $V_P > V_{DS}$ (线性区), $I_D \approx \frac{mk}{2}[V_P^2 - (V_P - V_{DS})^2] = k[(V_{GS} - V_T)V_{DS} - \frac{m}{2}V_{DS}^2]$, 即体电荷模型 (7.9); 若 $V_{DS} > V_P$, 第二项趋于 0, $I_D \approx k(V_{GS} - V_T)^2/(2m)$, 即饱和电流, 饱和电压 $V_P = (V_{GS} - V_T)/m$ 。
- **弱反型** ($z \ll -1$): $\ln(1 + e^{z/2}) \approx e^{z/2}$, $F \approx e^z$, $I_D \approx 2mkU_T^2 e^{(V_{GS}-V_T)/mU_T} (1 - e^{-V_{DS}/U_T})$, 与式 (8.2) 同形, 摆幅恰为 $\ln 10 \cdot mU_T$ 。

中间区域只是光滑插值, 并非由第一性原理推出; 本讲只把它用于反相器的半定量计算, 器件物理的结论都以 Pao-Sah 精确积分为准。因子 $(1 + \lambda V_{DS})$ 使饱和区满足第 7.8 节的定义 $\lambda = I_{D0}^{-1} \partial I_D / \partial V_{DS}$ 。

B 泊松方程第一积分的推广与弱反型电荷

B.1 沟道电势不为零时的第一积分

MOSFET 沟道中电子的准费米势为 V (相对衬底), 电子浓度 $n = n_0 e^{(\psi-V)/U_T}$, 空穴仍为 $p = p_0 e^{-\psi/U_T}$ 。为使深处体内 ($\psi = 0$) 电中性, 把电荷密度写成

$$\rho = q \left[p_0 \left(e^{-\psi/U_T} - 1 \right) - n_0 e^{-V/U_T} \left(e^{\psi/U_T} - 1 \right) \right]$$

(体内电子浓度 $n_0 e^{-V/U_T}$ 比 p_0 小十几个数量级, 这样写只差可以忽略的量)。重复第 6.4 节的步骤:

$$\begin{aligned} \left(\frac{d\psi}{dx} \right)^2 &= \frac{2qp_0 U_T}{\epsilon_s} F_V^2(\psi), \\ F_V^2 &= \left(e^{-\psi/U_T} + \frac{\psi}{U_T} - 1 \right) \\ &\quad + \frac{n_0}{p_0} e^{-V/U_T} \left(e^{\psi/U_T} - \frac{\psi}{U_T} - 1 \right). \end{aligned} \quad (\text{B.1})$$

电子项现在是 $e^{(\psi-2\phi_F-V)/U_T}$: 强反型需要 $\psi_s \approx 2\phi_F + V$, 这就是式 (7.4) 的依据。反型层电荷 $Q_n = -q \int (n - n_{\text{bulk}}) dx$ 用变量代换 $dx = d\psi / |d\psi/dx|$ 化为对 ψ 的积分:

$$Q_n = -qn_0 e^{-V/U_T} \int_0^{\psi_s} \frac{e^{\psi/U_T} - 1}{|d\psi/dx|} d\psi, \quad (\text{B.2})$$

$|d\psi/dx|$ 由式 (B.1) 给出。本讲的“精确解” (图 6、9、11 与实验题的数据) 都用式 (B.1)、(B.2) 与栅压方程 (6.3) 数值计算 (`scripts/mos1d.py`), 并用有限差分直接解泊松方程核对 (表面电场相差 $< 4 \times 10^{-4}$)。

B.2 弱反型电荷的更精确形式

弱反型时, 电子只分布在表面附近很薄的一层内, 可以在这一层内把电势展开到二阶。由泊松方程, 这一层中电荷几乎只有电离受主, $\psi'' = qN_A/\epsilon_s$, 所以

$$\begin{aligned} \psi(x) &\approx \psi_s - \mathcal{E}_s x + \frac{qN_A}{2\epsilon_s} x^2, \\ \mathcal{E}_s^2 &= \frac{2qN_A(\psi_s - U_T)}{\epsilon_s}, \end{aligned}$$

后一式来自第一积分 (B.1) 略去电子项后的耗尽区形式 (比耗尽近似多出 $-U_T$)。电子浓度 $n = n_s \exp \left[-\frac{\mathcal{E}_s x}{U_T} + \frac{qN_A x^2}{2\epsilon_s U_T} \right]$, 第二项相对第一项是小量 (在 $x \sim U_T/\mathcal{E}_s$ 处, 比值为 $U_T/[4(\psi_s - U_T)]$)。展开 $e^\epsilon \approx 1 + \epsilon$ 后逐项积分 ($\int_0^\infty x^2 e^{-x/a} dx = 2a^3$, $a = U_T/\mathcal{E}_s$):

$$\begin{aligned} \int_0^\infty n dx &\approx n_s \frac{U_T}{\mathcal{E}_s} \left[1 + \frac{qN_A U_T}{\epsilon_s \mathcal{E}_s^2} \right] \\ &= n_s \frac{U_T}{\mathcal{E}_s} \left[1 + \frac{U_T}{2(\psi_s - U_T)} \right]. \end{aligned}$$

整理得

$$Q_n \approx -C_d U_T \left(\frac{n_i}{N_A} \right)^2 e^{(\psi_s - V)/U_T} \times \left[1 + \frac{U_T}{2(\psi_s - U_T)} \right], \quad (B.3)$$

$$C_d = \sqrt{\frac{\varepsilon_s q N_A}{2(\psi_s - U_T)}}.$$

其中 C_d 即正文式 (8.1) 的定义，所以式 (B.3) 就是正文式乘以方括号。与式 (B.2) 的数值积分比较（标准器件， $\psi_s = 0.45\text{--}0.75\text{ V}$ ）：式 (B.3) 的误差 $< 0.4\%$ ；正文式 (8.1)（去掉方括号）偏小 $1.9\%\text{--}3.4\%$ ；若再把 C_d 中的 $\psi_s - U_T$ 换成 ψ_s （耗尽近似），偏小 $3.7\%\text{--}6.5\%$ 。注意：直接把第一积分 (6.5) 按电子项展开，得到的“ Q_s （有电子） $-Q_s$ （无电子）”恰是去掉方括号的形式——它不等于电子电荷，因为电子的出现还改变了同一表面势下耗尽层的电荷。

C 自然长度：改进的抛物线近似与精确本征方程

保留表面与中心的电势差 第 9.5 节第二步的连续条件 $\varepsilon_s a t_{\text{si}} = \varepsilon_{\text{ox}}(V'_g - \psi_c - a t_{\text{si}}^2/4)/t_{\text{ox}}$ 可以精确解出 a ： $a = \varepsilon_{\text{ox}}(V'_g - \psi_c)/[t_{\text{si}}(\varepsilon_s t_{\text{ox}} + \varepsilon_{\text{ox}} t_{\text{si}}/4)]$ 。代入 $\psi'_c = -2a$ ：

$$\ell^2 = \frac{\varepsilon_s t_{\text{si}} t_{\text{ox}}}{2\varepsilon_{\text{ox}}} + \frac{t_{\text{si}}^2}{8}. \quad (C.1)$$

分离变量的精确解 拉普拉斯方程是线性的，偏离长沟道值的部分 $\phi \equiv \psi - V'_g$ 满足 $\nabla^2 \phi = 0$ ，在栅上（ $x = \pm d$ ， $d = t_{\text{si}}/2 + t_{\text{ox}}$ ） $\phi = 0$ 。试探 $\phi = X(x)e^{-y/\ell}$ ： $X'' + X/\ell^2 = 0$ 。对称的最低阶解：硅中 $X = \cos(x/\ell)$ （ $|x| < t_{\text{si}}/2$ ），氧化层中 $X = B \sin[(d - |x|)/\ell]$ （在栅上为零）。在界面 $x = t_{\text{si}}/2$ 处 ϕ 连续、 $\varepsilon \partial \phi / \partial x$ 连续：

$$\cos \frac{t_{\text{si}}}{2\ell} = B \sin \frac{t_{\text{ox}}}{\ell}, \quad \varepsilon_s \sin \frac{t_{\text{si}}}{2\ell} = \varepsilon_{\text{ox}} B \cos \frac{t_{\text{ox}}}{\ell}.$$

两式相除消去 B ：

$$\varepsilon_s \tan \frac{t_{\text{si}}}{2\ell} \tan \frac{t_{\text{ox}}}{\ell} = \varepsilon_{\text{ox}} \quad (C.2)$$

（取 $t_{\text{si}}/2\ell < \pi/2$ 的根）[37]。两个正切都取小角近似 $\tan z \approx z$ ，即回到式 (9.6)。比较（ $t_{\text{si}}/t_{\text{ox}}$ ，单位 nm）：

$t_{\text{si}}/t_{\text{ox}}$	式 (9.6)	式 (C.1)	式 (C.2)	二维数值解
10/1	3.87	5.24	4.92	4.93
5/1	2.74	3.26	3.15	3.17
10/2	5.48	6.52	6.31	6.34

（二维数值解的衰减长度由 `scripts/scalelen.py` 拟合， $L = 80\text{ nm}$ ，拟合区间 $8 < y < 25\text{ nm}$ 。）简单抛物线近似偏小约 20%，改进的近似偏大约 5%，精确本征方程与数值解相差 $< 1\%$ 。各种近似的指数规律 $e^{-L/2\ell}$ 相同，只是 ℓ 的数值不同；文献中常用的“等效栅数”公式（第 9.5 节）属于简单抛物线近似一类，用于比较不同几何的趋势。

D 最小二乘直线拟合

对数据 (x_i, y_i) ($i = 1, \dots, N$) 拟合 $y = a + bx$, 使 $\chi^2 = \sum_i w_i (y_i - a - bx_i)^2$ 最小 ($w_i = 1/\sigma_i^2$; 各点误差未知时先取 $w_i = 1$)。令 $\partial\chi^2/\partial a = \partial\chi^2/\partial b = 0$, 记 $\Sigma_0 = \sum w_i$, $\Sigma_x = \sum w_i x_i$, $\Sigma_y = \sum w_i y_i$, $\Sigma_{xx} = \sum w_i x_i^2$, $\Sigma_{xy} = \sum w_i x_i y_i$, $D_\Sigma = \Sigma_0 \Sigma_{xx} - \Sigma_x^2$:

$$b = \frac{\Sigma_0 \Sigma_{xy} - \Sigma_x \Sigma_y}{D_\Sigma}, \quad a = \frac{\Sigma_{xx} \Sigma_y - \Sigma_x \Sigma_{xy}}{D_\Sigma}.$$

a 、 b 是 y_i 的线性组合, 按误差传递得 $\sigma_a^2 = \Sigma_{xx}/D_\Sigma$, $\sigma_b^2 = \Sigma_0/D_\Sigma$, 协方差 $\text{cov}(a, b) = -\Sigma_x/D_\Sigma$ 。若各点误差未知, 用残差估计 $\sigma^2 = \sum (y_i - a - bx_i)^2/(N-2)$, 再令 $w_i = 1/\sigma^2$ 。

实验题中阈值电压是直线与横轴的交点 $x_0 = -a/b$ 。 a 与 b 相关, 误差传递必须计入协方差:

$$\sigma_{x_0}^2 = \left(\frac{\sigma_a}{b}\right)^2 + \left(\frac{a\sigma_b}{b^2}\right)^2 - \frac{2a}{b^3} \text{cov}(a, b).$$

数据点集中在 x_0 一侧时 a 、 b 强相关, 忽略协方差会明显高估 σ_{x_0} 。 $\lambda = b/a$ 的误差同理。