Birla Institute of Technology and Science, Pilani
K. K. Birla Goa Campus

Electronic Device Simulation Laboratory

Design and TCAD Simulation of a Silicon MOSFET

From Long-Channel Operation to Short-Channel Scaling

Experiment 4, together with the scaling study of Experiment 5

Submitted by: Devadarsh Nair
ID No.: 2024A3PS0330G
Tools used: Sentaurus Structure Editor, Sentaurus Device, Inspect
Date: June 13, 2026

A two-dimensional n-channel MOSFET with a 1 µm gate, a 5 nm gate oxide and 25 nm deep source/drain junctions is constructed in Sentaurus Structure Editor and characterised with the drift-diffusion solver of Sentaurus Device. Transfer and output characteristics are simulated, threshold voltage, subthreshold swing and on/off currents are extracted, and the device is then deliberately pushed into the short-channel regime to expose drain-induced barrier lowering and punch-through.

Abstract

A planar n-channel MOSFET was designed on the MOS capacitor of the previous experiment and simulated end to end in the Sentaurus TCAD flow. The structure, built in Sentaurus Structure Editor, consists of a 100 nm thick p-type silicon body, a 5 nm gate oxide over a 1 µm channel, and arsenic-doped source/drain wells of concentration $1 \times 10^{20}/\text{cm}^3$ and depth 25 nm. The drift-diffusion equations were solved in Sentaurus Device with doping-dependent and field-dependent mobility, Shockley-Read-Hall recombination, band-gap narrowing and Fermi-Dirac statistics. From the simulated I_D–V_GS characteristics at $V_\text{DS} = 50\,\text{mV}$, the threshold voltage was extracted by the constant-current method, and the subthreshold swing, off-state and on-state currents were obtained in Inspect. The bias dependence of the internal band structure was examined through energy-band, potential and carrier-density cuts taken just below the oxide-semiconductor interface. The device was then scaled: gate length was reduced from 1 µm to 20 nm, the gate oxide was thinned from 5 nm to 1 nm, and the substrate doping was varied between 1×10^{15} and $1 \times 10^{18}/\text{cm}^3$. The observed degradation of the subthreshold swing, the collapse of the on/off ratio at $L_\text{g} = 20\,\text{nm}$, and the role of drain-induced barrier lowering are interpreted quantitatively with the quasi-two-dimensional characteristic-length model of Liu et al. The measured trends are found to be in good agreement with the analytical picture of electrostatic integrity.

Contents

1 Objectives

The experiment had the following aims, taken from the laboratory hand-out:

1. To design a two-dimensional n-channel MOSFET on the MOS capacitor of the previous part, with source and drain regions doped to $N_{\mathrm{D}} = 1 \times 10^{20}/\mathrm{cm}^3$ (donors) to a depth of 25 nm, on a 100 nm thick substrate.
2. To simulate the I_{D}–V_{GS} characteristic at $V_{\mathrm{DS}} = 50\,\mathrm{mV}$ on linear and semilogarithmic scales for $0 \leq V_{\mathrm{GS}} \leq 1\,\mathrm{V}$, and to extract the threshold voltage V_{th} by the constant-current method together with the subthreshold swing (SS), the off-state current I_{off} and the on-state current I_{on}.
3. To observe the energy-band diagram, the electrostatic potential and the carrier concentrations along the channel direction just below the oxide-semiconductor interface at $V_{\mathrm{GS}} = 0$, 0.5 and 1 V.
4. To simulate the I_{D}–V_{DS} output characteristics for $0 \leq V_{\mathrm{DS}} \leq 1\,\mathrm{V}$ at several gate voltages.
5. To study the influence of substrate doping, temperature (300 K to 400 K) and reduced source/drain doping ($1 \times 10^{19}/\mathrm{cm}^3$) on both characteristics.
6. To repeat the transfer characteristic at $V_{\mathrm{DS}} = 2\,\mathrm{V}$, compare it with the 50 mV case, and use the comparison to test for drain-induced barrier lowering (DIBL).
7. As the natural continuation (Experiment 5), to scale the gate length from 1 µm down to 20 nm, to compute the DIBL at $L_{\mathrm{g}} = 100\,\mathrm{nm}$, and to examine the influence of oxide thickness, substrate doping, temperature and body thickness on the scaled device.

2 Theoretical Background

2.1 MOS electrostatics and the threshold condition

When a positive voltage is applied to the gate of an n-channel device, the bands at the silicon surface bend downwards, holes are pushed away, and a depletion region of ionised acceptors forms under the oxide. The amount of band bending is measured by the surface potential ψ_{s}. The reference for "how p-type" the body is, is the bulk Fermi potential,

$$\phi_{\mathrm{F}} = \frac{kT}{q} \ln \frac{N_{\mathrm{A}}}{n_{\mathrm{i}}}, \tag{1}$$

where $n_{\mathrm{i}} \approx 1 \times 10^{10}/\mathrm{cm}^3$ in silicon at 300 K. The surface is said to be strongly inverted when the electron concentration at the surface equals the hole concentration in the bulk, which happens when the bands have bent by twice the Fermi potential,

$$\psi_{\mathrm{s}} = 2\phi_{\mathrm{F}}. \tag{2}$$

At this point the depletion region has reached its maximum width

$$W_{\mathrm{dm}} = \sqrt{\frac{4\,\varepsilon_{\mathrm{si}}\,\phi_{\mathrm{F}}}{qN_{\mathrm{A}}}}, \tag{3}$$

and any further gate charge is balanced almost entirely by inversion electrons rather than by additional depletion charge. The gate voltage needed to reach this condition is the threshold voltage. Adding the flat-band voltage, the band bending, and the voltage dropped across the oxide by the depletion charge $Q_{\mathrm{dep}} = qN_{\mathrm{A}}W_{\mathrm{dm}}$ gives

$$V_{\mathrm{th}} = V_{\mathrm{FB}} + 2\phi_{\mathrm{F}} + \frac{\sqrt{4\,\varepsilon_{\mathrm{si}}\,qN_{\mathrm{A}}\,\phi_{\mathrm{F}}}}{C_{\mathrm{ox}}}, \qquad C_{\mathrm{ox}} = \frac{\varepsilon_{\mathrm{ox}}}{t_{\mathrm{ox}}}. \tag{4}$$

For the present device, $t_{\mathrm{ox}} = 5\,\mathrm{nm}$ gives $C_{\mathrm{ox}} = \varepsilon_{\mathrm{ox}}/t_{\mathrm{ox}} \approx 6.9 \times 10^{-7}\,\mathrm{F/cm^2}$. Equation (4) already explains two of the parametric trends studied later: a higher substrate doping raises V_{th} through both ϕ_{F} and the depletion-charge term, while a thinner oxide lowers the depletion-charge term by increasing C_{ox}.

2.2 Drain current above threshold

Once the inversion layer exists, it behaves as a thin resistive sheet whose charge per unit area at position x along the channel is $Q_{\mathrm{inv}}(x) = C_{\mathrm{ox}}\,[V_{\mathrm{GS}} - V_{\mathrm{th}} - V(x)]$, where $V(x)$ is the local channel potential. Integrating the drift current along the channel under the gradual-channel approximation gives the familiar long-channel expression

$$I_{\mathrm{D}} = \mu_{\mathrm{eff}} C_{\mathrm{ox}} \frac{W}{L}\left[(V_{\mathrm{GS}} - V_{\mathrm{th}})V_{\mathrm{DS}} - \frac{V_{\mathrm{DS}}^2}{2}\right], \qquad V_{\mathrm{DS}} \le V_{\mathrm{GS}} - V_{\mathrm{th}}, \tag{5}$$

which reduces, for the small drain bias of 50 mV used in the transfer sweeps, to the linear form

$$I_{\mathrm{D}} \approx \mu_{\mathrm{eff}} C_{\mathrm{ox}} \frac{W}{L}\,(V_{\mathrm{GS}} - V_{\mathrm{th}})\,V_{\mathrm{DS}}. \tag{6}$$

Beyond pinch-off the current saturates at

$$I_{\mathrm{D}}^{\mathrm{sat}} = \frac{\mu_{\mathrm{eff}} C_{\mathrm{ox}}}{2}\frac{W}{L}\,(V_{\mathrm{GS}} - V_{\mathrm{th}})^2\,(1 + \lambda V_{\mathrm{DS}}), \tag{7}$$

where the factor $(1 + \lambda V_{\mathrm{DS}})$ accounts for channel-length modulation. Equation (6) is the basis of the linear-extrapolation method of finding V_{th}, while in this work the constant-current method described in Section 3.5 was used, since it is far more convenient on the semilogarithmic Inspect plots.

2.3 Subthreshold conduction and the swing

Below threshold the channel is not "off". A small electron concentration still exists at the surface, controlled exponentially by the surface potential, and the drain current is carried by diffusion of these electrons over the source-to-channel barrier. The standard result is

$$I_{\mathrm{D}}^{\mathrm{sub}} = \mu_{\mathrm{eff}} C_{\mathrm{ox}} \frac{W}{L}\,(m-1)\left(\frac{kT}{q}\right)^2 \exp\!\left(\frac{V_{\mathrm{GS}} - V_{\mathrm{th}}}{m\,kT/q}\right)\left[1 - \exp\!\left(-\frac{V_{\mathrm{DS}}}{kT/q}\right)\right], \tag{8}$$

where the body factor

$$m = 1 + \frac{C_{\mathrm{dep}}}{C_{\mathrm{ox}}}, \qquad C_{\mathrm{dep}} = \frac{\varepsilon_{\mathrm{si}}}{W_{\mathrm{dm}}}, \tag{9}$$

expresses the fact that the gate must share its control of the surface potential with the depletion capacitance: only the fraction $1/m$ of a change in V_{GS} appears as a change in ψ_{s}. Taking the logarithm of (8) gives the subthreshold swing, the gate voltage required to change the drain current by one decade,

$$\mathrm{SS} = \left(\frac{\mathrm{d}\,\log_{10} I_{\mathrm{D}}}{\mathrm{d}V_{\mathrm{GS}}}\right)^{-1} = \ln(10)\,\frac{kT}{q}\left(1 + \frac{C_{\mathrm{dep}} + C_{\mathrm{it}}}{C_{\mathrm{ox}}}\right) \ \ge\ 59.6\,\mathrm{mV/dec\ at\ 300\,K}. \tag{10}$$

The lower bound of about 60 mV per decade at room temperature is purely thermodynamic: it comes from the Boltzmann tail of the carrier distribution and cannot be beaten by any conventional MOSFET. Interface traps add a capacitance C_{it} in parallel with C_{dep} and degrade the swing further; the simulations here assume an ideal interface, so the measured swing reflects only the depletion term and, in short devices, the loss of gate control discussed next.

2.4 Short-channel effects and drain-induced barrier lowering

In a long device the potential barrier between source and channel is set by the gate alone, and the drain is electrostatically invisible from the source. As the gate length is reduced, the depletion regions of the source and drain junctions occupy a growing fraction of the channel, and field lines originating on the drain begin to terminate on channel depletion charge that the gate would otherwise have had to support. Two consequences follow. First, the threshold voltage falls with decreasing length even at low drain bias (V_{th} roll-off). Second, raising V_{DS} pulls the source-side barrier down directly, which shifts the whole transfer characteristic to lower gate voltages. This second effect is drain-induced barrier lowering, and it is illustrated schematically in Figure 1.

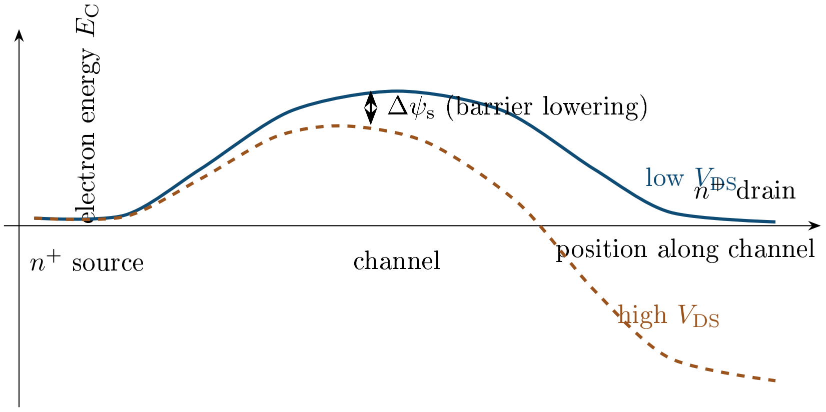


Figure 1: Schematic conduction-band profile along the channel near the surface. In a short device a large drain bias does not stay confined to the drain end: it reaches through and lowers the source-side barrier by $\Delta\psi_s$, so that more electrons spill over the barrier at the same gate voltage. The transfer curve therefore shifts left, which is read as a reduction of V_{th}.

A compact quantitative description is given by the quasi-two-dimensional model of Liu et al. [1]. Gauss's law is applied to a box of height equal to the depletion depth X_{dep} spanning the channel, which yields a second-order equation for the surface potential along the channel,

$$\frac{\varepsilon_{\mathrm{si}}\,X_{\mathrm{dep}}}{\eta}\,\frac{\mathrm{d}^2\psi_s}{\mathrm{d}y^2}+\varepsilon_{\mathrm{ox}}\,\frac{V_{\mathrm{GS}}-V_{\mathrm{FB}}-\psi_s}{t_{\mathrm{ox}}}=qN_{\mathrm{A}}X_{\mathrm{dep}}, \tag{11}$$

with $\eta \approx 1$ a fitting parameter. The solutions are exponentials whose decay constant defines the characteristic length of the technology,

$$\ell=\sqrt{\frac{\varepsilon_{\mathrm{si}}}{\varepsilon_{\mathrm{ox}}}\,\frac{t_{\mathrm{ox}}\,X_{\mathrm{dep}}}{\eta}}\;\approx\;\sqrt{3\,t_{\mathrm{ox}}\,X_{\mathrm{dep}}}\quad\text{for SiO}_2\text{ on Si.} \tag{12}$$

Physically, ℓ is the distance over which a potential disturbance at the drain decays as it propagates towards the source through the channel and the oxide. Solving (11) with the built-in potential V_{bi} of the source/drain junctions as boundary conditions gives the combined threshold shift due to charge sharing and DIBL [1],

$$\Delta V_{\mathrm{th}}\;=\;-\big[\,2\,(V_{\mathrm{bi}}-2\phi_{\mathrm{F}})+V_{\mathrm{DS}}\,\big]\Big[\exp\!\Big(-\frac{L}{2\ell}\Big)+2\exp\!\Big(-\frac{L}{\ell}\Big)\Big],\qquad V_{\mathrm{bi}}=\frac{kT}{q}\ln\frac{N_{\mathrm{A}}N_{\mathrm{D}}}{n_i^2}. \tag{13}$$

Two properties of (13) matter for everything that follows. The threshold shift grows *exponentially* as L shrinks, so the transition from a well-behaved device to a broken one is abrupt; and the shift is *linear in* V_{DS} for $L \gg \ell$, which is exactly the DIBL signature. The DIBL figure of merit used in this report is therefore defined as the threshold shift per volt of drain bias,

$$\mathrm{DIBL} \;=\; \frac{V_{\mathrm{th}}(V_{\mathrm{DS}}^{\mathrm{lin}}) - V_{\mathrm{th}}(V_{\mathrm{DS}}^{\mathrm{sat}})}{V_{\mathrm{DS}}^{\mathrm{sat}} - V_{\mathrm{DS}}^{\mathrm{lin}}} \quad [\mathrm{mV/V}], \qquad V_{\mathrm{DS}}^{\mathrm{lin}} = 50\,\mathrm{mV}, \quad V_{\mathrm{DS}}^{\mathrm{sat}} = 1\,\mathrm{V}, \tag{14}$$

and the model predicts DIBL $\approx \exp(-L/2\ell) + 2\exp(-L/\ell)$ in volts per volt. As a design rule, the gate keeps control of the barrier only while L is at least about five characteristic lengths; once $L \sim \ell$ the exponentials approach unity, the barrier is set by the drain rather than the gate, and the device punches through.

Punch-through itself can be estimated from one-dimensional junction theory. The zero-bias depletion width of each source/drain junction into the body is

$$W_{\mathrm{j}} = \sqrt{\frac{2\varepsilon_{\mathrm{si}}\,(V_{\mathrm{bi}} + V_{\mathrm{R}})}{qN_{\mathrm{A}}}}, \tag{15}$$

with V_{R} the reverse bias on the junction ($V_{\mathrm{R}} = V_{\mathrm{DS}}$ on the drain side). When the sum of the two depletion widths approaches the channel length, the source and drain are connected by a depleted, gate-independent path and the subthreshold current can no longer be switched off. This estimate is used in Section 4.6 to explain why the 20 nm device fails for every substrate doping that was tried.

2.5 Temperature dependence

Temperature enters the characteristics through three main channels. The intrinsic concentration rises steeply,

$$n_{\mathrm{i}}(T) = \sqrt{N_{\mathrm{C}}N_{\mathrm{V}}}\;\exp\!\left(-\frac{E_{\mathrm{g}}}{2kT}\right) \;\propto\; T^{3/2}\exp\!\left(-\frac{E_{\mathrm{g}}}{2kT}\right), \tag{16}$$

which reduces ϕ_{F} in (1) and hence reduces the threshold voltage. Differentiating (4) gives

$$\frac{\mathrm{d}V_{\mathrm{th}}}{\mathrm{d}T} \;=\; \frac{\mathrm{d}\phi_{\mathrm{F}}}{\mathrm{d}T}\left(2 + \frac{1}{C_{\mathrm{ox}}}\sqrt{\frac{\varepsilon_{\mathrm{si}}\,qN_{\mathrm{A}}}{\phi_{\mathrm{F}}}}\right) \;\approx\; -1 \text{ to } -2\,\mathrm{mV/K} \tag{17}$$

for typical dopings, since $\mathrm{d}\phi_{\mathrm{F}}/\mathrm{d}T < 0$. Secondly, the lattice mobility falls with temperature because phonon scattering intensifies, approximately as

$$\mu(T) \;\propto\; \left(\frac{T}{300\,\mathrm{K}}\right)^{-\alpha}, \qquad \alpha \approx 2.5 \text{ for electrons in Si}, \tag{18}$$

which reduces the on-current. Thirdly, the subthreshold swing of (10) is directly proportional to T, so the subthreshold region both shifts left (lower V_{th}) and flattens (larger SS) as the device is heated, and the off-current rises roughly exponentially. Because the threshold reduction raises the current while the mobility reduction lowers it, the I_{D}–V_{GS} curves taken at different temperatures cross at a single gate voltage, the zero-temperature-coefficient (ZTC) point; below it the current has a positive temperature coefficient, above it a negative one.

2.6 Role of the source/drain doping

The n^{+} wells serve two purposes: they supply carriers to the channel through low-resistance contacts, and they define the junction electrostatics. Lowering N_{D} from 1×10^{20} to $1 \times 10^{19}/\mathrm{cm}^{3}$ acts mainly through the parasitic series resistance $R_{\mathrm{SD}} \propto 1/(q\mu_n N_{\mathrm{D}})$ of the extensions, so the extrinsic transconductance and the on-current drop, most visibly at high V_{GS} where the channel resistance is smallest and R_{SD} is no longer negligible in

$$I_{\mathrm{D}}^{\mathrm{ext}} \approx \frac{V_{\mathrm{DS}} - I_{\mathrm{D}}\,R_{\mathrm{SD}}}{V_{\mathrm{DS}}}\;I_{\mathrm{D}}. \tag{19}$$

The subthreshold region, governed by the barrier rather than by resistance, is nearly unaffected, apart from a slight reduction of the built-in potential V_{bi} (a 60 mV drop per decade of N_{D} at room temperature), which marginally relaxes the short-channel terms in (13). At $N_{\mathrm{D}} = 1 \times 10^{20}/\mathrm{cm}^3$ the wells are degenerate, which is why Fermi-Dirac statistics and band-gap narrowing are switched on in the device simulation (Section 3.3).

3 Device Structure and Simulation Methodology

3.1 Geometry and doping

The device was constructed in Sentaurus Structure Editor (SDE) as a two-dimensional cross section; the full command file is reproduced in Appendix A. The silicon body is a rectangle of length $2L_{\mathrm{sd}} + L_{\mathrm{g}}$ and thickness $t_{\mathrm{si}} = 100\,\mathrm{nm}$, doped uniformly p-type with boron at the concentration N_{A} carried over from the MOS capacitor of the previous experiment. Two n^+ wells of depth $t_{\mathrm{sd}} = 25\,\mathrm{nm}$ are placed in the top corners by superimposing a constant arsenic profile of $N_{\mathrm{D}} = 1 \times 10^{20}/\mathrm{cm}^3$ inside the source and drain refinement windows; because the donor concentration exceeds the background acceptor concentration by several orders of magnitude, the net doping in these boxes is n-type and abrupt metallurgical junctions are formed at their edges. A gate oxide of thickness $t_{\mathrm{ox}} = 5\,\mathrm{nm}$ is drawn only over the central channel section, and four ideal Ohmic/gate contacts are attached: source and drain on the exposed silicon surface above the wells, the gate on the top face of the oxide, and a substrate contact along the entire bottom face. The resulting structure is sketched in Figure 2, and the geometry and doping parameters are collected in Table 1.

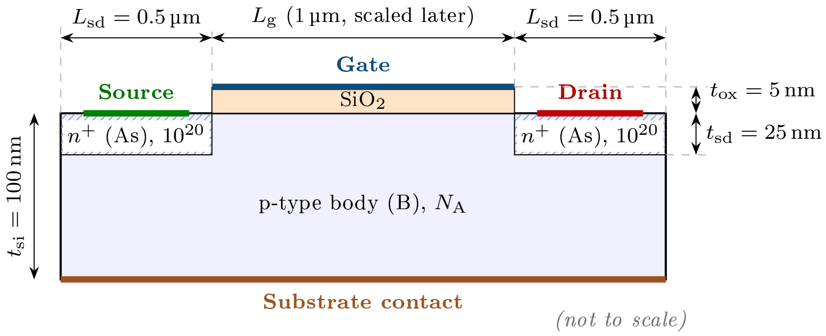


Figure 2: Cross section of the simulated n-channel MOSFET, redrawn from the SDE command file of Appendix A. The lateral coordinate x runs from the source edge to the drain edge and the vertical coordinate y runs from the substrate contact ($y = 0$) to the gate, exactly as in the structure file. The hatched boxes are the constant-profile arsenic wells; vertical thicknesses are exaggerated for readability.

Table 1: Geometry, doping and bias parameters. The third column lists the baseline of Experiment 4; the fourth lists the range over which each parameter is varied in the later sections.

Symbol	Quantity	Baseline	Varied over
L_g	gate length	1 µm	1000 nm to 20 nm
L_{sd}	source/drain extension length	0.5 µm	fixed
t_{si}	silicon body thickness	100 nm	100 nm to 20 nm
t_{ox}	gate oxide thickness	5 nm	5 nm to 1 nm
t_{sd}	source/drain junction depth	25 nm	fixed
N_A	substrate acceptor (B) doping	from the MOS-capacitor part	$1 \times 10^{15}/\mathrm{cm}^3$ to $1 \times 10^{18}/\mathrm{cm}^3$
N_D	source/drain donor (As) doping	$1 \times 10^{20}/\mathrm{cm}^3$	also $1 \times 10^{19}/\mathrm{cm}^3$
T	lattice temperature	300 K	250 K to 400 K
V_{DS}	drain read-out bias	50 mV	also 1 V and 2 V
V_{GS}	gate sweep	0 V to 1 V	fixed

3.2 Mesh strategy

A drift-diffusion solution is only as trustworthy as the mesh it is computed on, so the discretisation was treated as part of the device design. The guiding idea is that the node spacing must resolve the smallest physical length in each direction: vertically, the inversion layer extends only a few nanometres below the oxide, and laterally, the doping changes by ten orders of magnitude across the metallurgical junctions. Five rectangular refinement windows were therefore defined, summarised in Table 2. The finest vertical spacing, 1 nm, is reserved for the top 30 nm of the channel and for the oxide, where the surface potential and the inversion charge are formed; the source and drain windows are refined to 2 nm so that the junction sidewalls are sharp; and a coarser background mesh fills the rest of the body, where nothing varies quickly.

In addition, every window carries the refinement function `MaxTransDiff` of value 1 on the doping concentration. This instructs the mesher to keep subdividing an element as long as the doping, measured on a signed logarithmic scale, changes by more than about one unit across it. The practical effect is that nodes are packed automatically around the metallurgical junctions without any manual placement, which is exactly where the space-charge regions and therefore the strongest fields appear. With abrupt constant profiles this safeguard is inexpensive, but it becomes essential once analytic or implanted profiles are used.

Table 2: Refinement windows used in the SDE mesh (Appendix A). Element sizes are the maximum permitted edge lengths inside each window; `MaxTransDiff` = 1 on the net doping is active in all of them.

Window	Region covered	$\Delta x \times \Delta y$ (max)	Purpose
`mesh_si`	entire silicon body	20 nm × 5 nm	background mesh of the bulk
`mesh_ch`	top 30 nm of the channel	10 nm × 1 nm	resolves the inversion layer and the vertical surface field
`mesh_ox`	gate oxide	10 nm × 1 nm	resolves the oxide field and the interface
`win_src`	source well, top 25 nm	10 nm × 2 nm	sharp source junction; also hosts the As profile
`win_drn`	drain well, top 25 nm	10 nm × 2 nm	sharp drain junction; also hosts the As profile

3.3 Physics models in Sentaurus Device

The transport was described at the drift-diffusion level: the Poisson equation and the electron and hole continuity equations were solved self-consistently, with the current densities written as drift plus diffusion. The model selection in the `Physics` section of the command file (Appendix B) was not decorative; each keyword was chosen because a specific physical mechanism is known to matter in this device, and each is recalled below in the form of the equation it activates.

Doping-dependent low-field mobility. Ionised dopants scatter carriers, so the low-field mobility must fall as the doping rises. The keyword `DopingDependence` activates the Masetti fit [6], which over the range of interest behaves like the familiar form

$$\mu_{\text{dop}}(N) \;\approx\; \mu_{\text{min}} \;+\; \frac{\mu_{\text{max}} - \mu_{\text{min}}}{1 + (N/N_{\text{ref}})^{\alpha}}, \tag{20}$$

with $\mu_{\text{max}} \approx 1417\,\text{cm}^2/(\text{V s})$ for electrons in undoped silicon. In the present structure this matters twice: the channel mobility depends on N_A, which is varied over three decades, and the n^+ wells at $1 \times 10^{20}/\text{cm}^3$ have a strongly suppressed mobility that contributes to the series resistance.

High-field velocity saturation. At drain biases of the order of 1 V across short channels, the lateral field reaches values at which the drift velocity no longer grows linearly. The keyword `eHighFieldSaturation` activates the Canali expression [7]

$$\mu(F) \;=\; \frac{\mu_{\text{low}}}{\left[1 + \left(\dfrac{\mu_{\text{low}} F}{v_{\text{sat}}}\right)^{\beta}\right]^{1/\beta}}, \qquad v_{\text{sat}} \approx 1.07 \times 10^7\,\text{cm/s}, \tag{21}$$

where F is the driving force along the current and $\beta \approx 1.1$ for electrons. Without this model the saturation currents of Section 4 would be overestimated, particularly for the scaled devices where the average lateral field is largest.

Transverse-field (surface) mobility degradation. The same gate field that creates the inversion layer also presses the electrons against the rough Si/SiO$_2$ interface. The keyword `Enormal` activates the Lombardi model [8], in which the surface contributions are combined with the bulk mobility by Matthiessen's rule,

$$\frac{1}{\mu} \;=\; \frac{1}{\mu_{\text{b}}} \;+\; \frac{1}{\mu_{\text{ac}}} \;+\; \frac{1}{\mu_{\text{sr}}}, \qquad \mu_{\text{ac}} \propto E_\perp^{-1}, \quad \mu_{\text{sr}} \propto E_\perp^{-2}, \tag{22}$$

where $E_\perp$ is the field normal to the interface, μ_{ac} describes scattering by surface acoustic phonons and μ_{sr} scattering by surface roughness. This is the mechanism behind the gradual flattening of the transfer characteristic at high V_{GS}: the gate keeps adding inversion charge, but each added electron is slower than the last.

Generation and recombination. Carrier exchange between the bands through deep levels was described by the Shockley-Read-Hall expression

$$R_{\text{SRH}} \;=\; \frac{np - n_\text{i}^2}{\tau_p\,(n + n_1) + \tau_n\,(p + p_1)}, \tag{23}$$

with the lifetimes made doping dependent through the Scharfetter relation

$$\tau(N) \;=\; \tau_{\text{min}} \;+\; \frac{\tau_{\text{max}} - \tau_{\text{min}}}{1 + (N/N_{\text{ref}})^{\gamma}}, \tag{24}$$

so that the heavily doped wells have much shorter lifetimes than the lightly doped body. In a MOSFET biased in the normal direction SRH mainly fixes the leakage floor of the reverse-biased drain junction; it is cheap to include and removes one possible artefact from the off-current.

Band-gap narrowing and Fermi-Dirac statistics. At $N_D = 1 \times 10^{20}/\text{cm}^3$ the source and drain are degenerate, and two assumptions of textbook device physics fail there: the band gap is no longer the intrinsic 1.12 eV, and the Boltzmann approximation no longer connects the carrier density to the Fermi level. The first effect was captured by the Slotboom (`OldSlotboom`) model [9],

$$\Delta E_g(N) \;=\; E_{\text{ref}}\left[\ln\frac{N}{N_{\text{ref}}} + \sqrt{\left(\ln\frac{N}{N_{\text{ref}}}\right)^2 + \tfrac{1}{2}}\;\right], \qquad E_{\text{ref}} = 9\,\text{meV},\;\; N_{\text{ref}} = 1 \times 10^{17}/\text{cm}^3, \quad (25)$$

which enters the simulation through an effective intrinsic density $n_{\text{i,eff}} = n_{\text{i}} \exp(\Delta E_g/2kT)$ and therefore raises the minority-carrier densities and lowers the built-in potentials of the junctions. The second effect was handled by switching the statistics from Boltzmann to Fermi-Dirac with the keyword `Fermi`,

$$n \;=\; N_C\, F_{1/2}\!\left(\frac{E_{\text{Fn}} - E_C}{kT}\right), \qquad (26)$$

where $F_{1/2}$ is the Fermi integral of order one half. Below about $1 \times 10^{19}/\text{cm}^3$ the two statistics agree, but in the wells the Boltzmann form would overestimate the carrier density at a given Fermi-level position. Including both models together is the standard prescription for degenerate regions, since they pull the equilibrium electrostatics in opposite directions and only their combination is physically consistent.

3.4 Bias sequence and numerical control

The `Solve` block of Appendix B executes the whole experiment in one run, and its structure follows the standard Sentaurus practice of never asking the nonlinear solver for a large step [13, 14]. The sequence is as follows.

1. **Equilibrium.** The Poisson equation alone is solved first with up to one hundred damped Newton iterations; this nonlinear problem is the hardest one of the run because the initial guess is poor. The converged potential is then used as the starting point for the fully coupled Poisson, electron and hole system at zero bias, and the solution is saved so that later sweeps can restart from it.

2. **Drain ramp.** The drain is ramped from 0 to $V_{DS} = 50\,\text{mV}$ inside a `QuasiStationary` statement. The ramp is parameterised by an internal variable t that runs from 0 to 1, with the contact voltage interpolated linearly between its initial and goal values. The solver begins with a step of 0.01 in t; after every successful step the step size is multiplied by the `Increment` factor of 1.5, and after every failed step it is reduced, the run being abandoned only if the step falls below `MinStep` $= 10^{-5}$. This continuation strategy lets the program take large steps where the solution is smooth and small steps where it bends sharply, without any manual tuning.

3. **Gate sweep (transfer characteristic).** With the drain held at 50 mV, the gate is swept from 0 to $V_{GS} = 1\,\text{V}$ in a second `QuasiStationary` block with a finer `MaxStep` of 0.01, since the current changes by several decades in the subthreshold region and the curve must be sampled densely there. This block is the only one allowed to write to the main current file: the preceding ramps were given a different `NewCurrentPrefix`, so the `.plt` file used for extraction contains the gate sweep alone and the I_D–V_{GS} curve plotted later is not contaminated by points from the drain ramp.

4. **Snapshots for the internal quantities.** Inside the gate sweep, a `Plot` statement with `Time = (0; 0.5; 1.0)` writes the full two-dimensional solution at $t = 0$, 0.5 and 1. Because the gate voltage is a linear function of t over the interval 0 V to 1 V, these snapshots land exactly at $V_{GS} = 0$, 0.5 and 1 V, which are precisely the biases at which the hand-out asks for the band diagram, the potential and the carrier concentrations.

5. **Output characteristic.** The saved equilibrium solution is reloaded, the gate is first ramped to a fixed value under the silent prefix `IdVd_setup_`, and the drain is then swept from 0 to 1 V under the prefix `IdVd_`. Reloading equilibrium instead of continuing from the end of the gate sweep guarantees that the output characteristic starts from a clean, history-free state.

On the numerical side, `Extrapolate` reuses the two previous solutions of a quasistationary sequence to construct the initial guess of the next step, which markedly reduces the Newton iteration count along smooth portions of the sweep; the relative-error control was set to five digits; and the linear systems were solved with the direct sparse solver `ParDiSo` on four threads, a robust choice for two-dimensional meshes of this size.

3.5 Extraction definitions

All extractions were performed in Inspect on the current file of the gate sweep, and the following conventions are used throughout the report. Because the simulation is two dimensional, every terminal current is reported per micrometre of device width, and an effective width of $W = 1\,\mu\text{m}$ is assumed wherever an absolute current is quoted.

Threshold voltage, constant-current method. The threshold is defined as the gate voltage at which the drain current crosses a fixed reference level scaled by the aspect ratio,

$$V_{\text{th}} = V_{\text{GS}}\Big|_{I_{\text{D}} = \frac{W}{L_{\text{g}}} I_{\text{th}}}, \qquad I_{\text{th}} = 100\,\text{nA}. \tag{27}$$

The method is preferred here over the linear-extrapolation method because it requires no fitting, it is insensitive to the mobility degradation that bends the strong-inversion part of the curve, and it transfers without modification to the short-channel devices of Section 4.4, where the linear region is poorly developed. For the baseline device $W/L_{\text{g}} = 1$ and the criterion is simply $I_{\text{D}} = 1\times10^{-7}\,\text{A}/\mu\text{m}$.

The subthreshold swing was measured as the inverse slope of the semilogarithmic transfer characteristic over a window well below threshold,

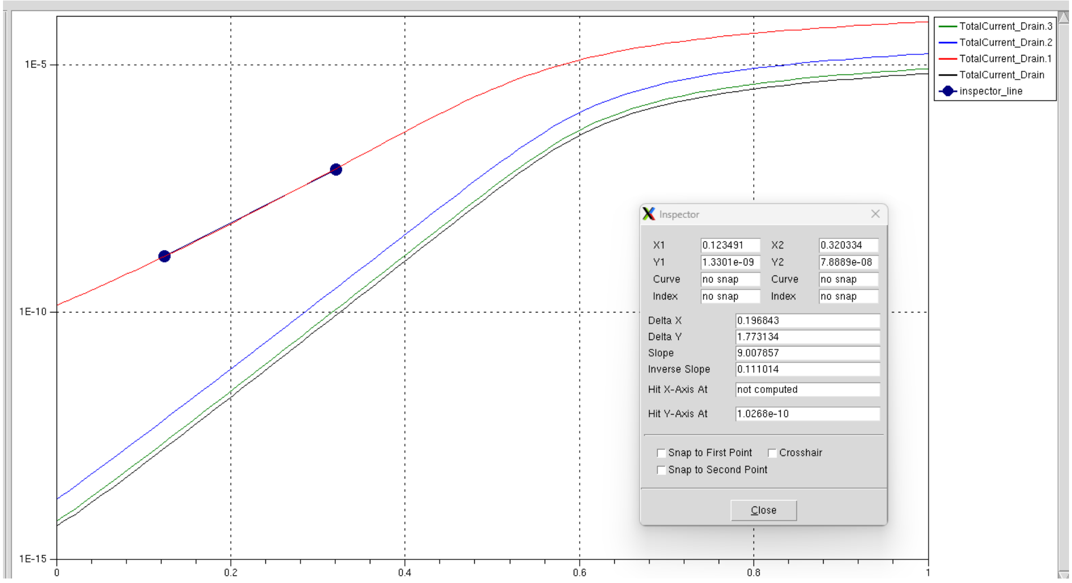
$$\text{SS} = \left(\frac{\mathrm{d}\,(\log_{10} I_{\text{D}})}{\mathrm{d}V_{\text{GS}}}\right)^{-1}, \tag{28}$$

using the linear-regression tool of Inspect between two points placed roughly two decades apart; the tool reports the slope in decades per volt and its inverse directly. The off-current was read at $V_{\text{GS}} = 0$ and the on-current at $V_{\text{GS}} = 1\,\text{V}$, both at the drain bias of the sweep in question. Finally, the DIBL was obtained by applying the constant-current criterion (27) to two transfer characteristics of the same device taken at a low and a high drain bias and inserting the two thresholds into (14); the procedure is demonstrated in Section 4.4.1.

4 Results and Discussion

4.1 Transfer characteristic and parameter extraction at $V_{\text{DS}} = 50\,\text{mV}$

Figure 3 shows the simulated transfer characteristics on a semilogarithmic scale. The family was obtained while the gate length was being stepped, so the lowest curve belongs to the longest channel and the topmost (red) curve to the shortest; the Inspector overlay reads the slope on that shortest-channel curve, and the long-channel extraction discussed here is read off the curve with the steepest, cleanest subthreshold region. Three features are visible immediately and each is quantified below: a clear exponential subthreshold region spanning several decades, a knee near 0.45 V where the device enters strong inversion, and a gentle flattening above 0.7 V caused by the mobility degradation of (22).



X1 0.123491 X2 0.320334
Y1 1.3301e-09 Y2 7.8889e-08
Curve no snap Curve no snap
Index no snap Index no snap

Delta X 0.196843
Delta Y 1.773134
Slope 9.007857
Inverse Slope 0.111014
Hit X-Axis At not computed
Hit Y-Axis At 1.0268e-10

☐ Snap to First Point ☐ Crosshair
☐ Snap to Second Point

Figure 3: Simulated I_D–V_{GS} characteristics at $V_{DS} = 50\,\text{mV}$ on a semilogarithmic scale, read out in Inspect. The four curves correspond to the gate-length series of Section 4.4 (the topmost red curve is the shortest channel). The two markers on the red curve are the Inspector slope tool, which returns an inverse slope of 0.111 V/dec. *To be confirmed by the author: the curve used for the long-channel extraction below is the steepest of the family, corresponding to* $L_g = 1\,\mu\text{m}$.

Subthreshold swing. The Inspector slope tool, placed between $V_{GS} = 0.123\,\text{V}$ and $0.320\,\text{V}$, returns a slope of 9.008 dec/V and hence an inverse slope of

$$\text{SS} = \frac{1}{9.008\,\text{dec/V}} = 0.1110\,\text{V/dec} \approx 111\,\text{mV/dec}. \tag{29}$$

Inserting this into the swing expression (10) gives a body factor of

$$m = \frac{\text{SS}}{(kT/q)\ln 10} = \frac{111\,\text{mV/dec}}{59.6\,\text{mV/dec}} \approx 1.86, \tag{30}$$

which through $m = 1 + C_{\text{dep}}/C_{\text{ox}}$ implies a depletion capacitance comparable to the oxide capacitance. A value somewhat above the often-quoted ideal of 60 mV/dec to 70 mV/dec is expected here: the oxide is relatively thick at 5 nm (so C_{ox} is modest) and the curve used by the tool carries some short-channel content. The important point for the report is the method, not the third digit: the swing is obtained as the reciprocal of a straight-line fit in the exponential region, exactly as defined in Section 3.5.

Threshold voltage. Applying the constant-current criterion (27) with $W/L = 1$, the threshold is the gate voltage at which the drain current reaches 100 nA. Reading horizontally across Figure 3 to $I_D = 1 \times 10^{-7}\,\text{A}$ places the long-channel threshold at roughly 0.35 V. The Inspector also reports that the fitted subthreshold line, extrapolated downwards, would intercept the current axis at $1.03 \times 10^{-10}\,\text{A}$; this extrapolated zero-gate intercept is a convenient internal check, since multiplying it by $10^{V_{GS}/\text{SS}}$ reproduces the current at any subthreshold gate voltage.

On- and off-currents. For the long-channel baseline the off-current at $V_{GS} = 0$ sits in the picoampere-per-micrometre range, far below the threshold reference, while the on-current at $V_{GS} = 1\,\text{V}$ is of order $10^{-5}\,\text{A}/\mu\text{m}$; the modest absolute value of I_{on} is the direct consequence of reading at only 50 mV of drain bias, where the device is deep in its linear region and I_D is proportional to that small V_{DS} through (5). The corresponding on/off ratio of six to seven decades is the quantitative statement that a well-tempered long-channel device is an excellent switch, and it is precisely this ratio that the scaling study will be seen to destroy.

4.2 Internal quantities along the channel

The three snapshots written at $V_{\mathrm{GS}} = 0$, 0.5 and 1 V (Section 3.4) allow the switching to be watched directly in the band structure rather than only inferred from the terminal current. The expected behaviour, against which the simulated cuts should be compared, follows from the electrostatics of Section 2.4.

Plot to be inserted from Inspect / Sentaurus Visual:

Energy bands, potential and electron density on a cut 1 nm below the oxide, at $V_{\mathrm{GS}} = 0, 0.5$, 1 V

Export the conduction-band edge $E_{\mathrm{C}}(x)$, the electrostatic potential $\psi(x)$ and the electron concentration $n(x)$ from the saved snapshots in Sentaurus Visual along a horizontal cutline placed just under the interface, and overlay the three gate biases.

Figure 4: Internal quantities along the channel at the three gate biases. The qualitative reading is given in the text.

At $V_{\mathrm{GS}} = 0$ the conduction band shows a tall barrier in the channel between the two n^+ regions; the electron density on the surface cut is low and the potential well of the inversion layer has not yet formed, so the only path from source to drain is over the full built-in barrier and the current is the subthreshold leakage. At $V_{\mathrm{GS}} = 0.5$ V, just above threshold, the surface potential has risen by about $2\phi_{\mathrm{F}}$, the barrier has been pulled down by roughly the same amount, and a thin sheet of electrons appears on the cut: this is the inversion layer being created. At $V_{\mathrm{GS}} = 1$ V the barrier is essentially gone, the surface electron density is high and nearly independent of position along the channel, and the potential drop is confined to the immediate vicinity of the drain. Watching the barrier collapse as the gate is raised is the most direct picture of what threshold means, and it is the long-channel reference against which the short-channel barrier lowering of the next sections is to be read.

4.3 Output characteristics

The output family I_{D}–V_{DS} at fixed gate voltages exhibits the two classical regions predicted by (6) and (7). At low V_{DS} each curve rises almost linearly, the channel acting as a gate-controlled resistor; as V_{DS} approaches the overdrive $V_{\mathrm{GS}} - V_{\mathrm{th}}$ the channel pinches off at the drain and the current saturates. The spacing between successive gate steps widens at higher V_{GS}, the signature of the square-law dependence of the saturation current on overdrive.

Plot to be inserted from Inspect / Sentaurus Visual:

I_{D}–V_{DS} for $0 \leq V_{\mathrm{DS}} \leq$ 1 V at several gate voltages

Plot the current file written under the `IdVd_` prefix; one curve per gate voltage. Mark the linear and saturation regions, and note the finite output slope in saturation as the channel-length-modulation parameter.

Figure 5: Simulated output characteristics. Linear, saturation and channel-length-modulation regions are discussed in the text.

4.4 Gate-length scaling and DIBL

The transfer family of Figure 3 is itself the gate-length study, since it was produced by stepping L_g at the fixed read-out bias of 50 mV. The trend with decreasing length is unambiguous and entirely in the direction predicted by the characteristic-length argument of Section 2.4: the shortest channel (red) shows the largest off-current, the smallest threshold and the most degraded subthreshold slope, and the curves climb monotonically as the channel is shortened. Each of these is a symptom of the source and drain depletion regions taking over a growing share of the body charge that the gate is supposed to control.

Plot to be inserted from Inspect / Sentaurus Visual:

V_th, SS, I_off and I_on versus L_g from 1 µm to 100 nm

Extract the four parameters at each gate length by the definitions of Section 3.5 and plot them against L_g (log scale on the current axes). The expected shape is a flat plateau at long L_g that rolls off once L_g falls below roughly 5ℓ.

Figure 6: Scaling of the extracted figures of merit with gate length. The roll-off point is set by the characteristic length ℓ of (12).

Estimate of the roll-off point. For the baseline body the depletion depth is capped by the body thickness, so $X_\mathrm{dep} \approx 100\,\mathrm{nm}$, and with $t_\mathrm{ox} = 5\,\mathrm{nm}$ the characteristic length of (12) is

$$\ell = \sqrt{3\,t_\mathrm{ox}\,X_\mathrm{dep}} = \sqrt{3 \times 5\,\mathrm{nm} \times 100\,\mathrm{nm}} \approx 38.7\,\mathrm{nm}. \tag{31}$$

The empirical requirement $L_\mathrm{g} \gtrsim 5\ell$ then puts the onset of serious short-channel behaviour near $L_\mathrm{g} \approx 190\,\mathrm{nm}$. This is consistent with the figure: at 1 µm ($L_\mathrm{g}/\ell \approx 26$) the device is comfortably long, whereas at 100 nm ($L_\mathrm{g}/\ell \approx 2.6$) the exponential factor $\exp(-L_\mathrm{g}/2\ell) \approx 0.27$ is no longer small and the drain begins to lower the barrier appreciably.

4.4.1 Measuring the DIBL

The DIBL of the 100 nm device is obtained by the recipe of Section 3.5: the transfer characteristic is simulated twice, once at $V_\mathrm{DS} = 50\,\mathrm{mV}$ and once at $V_\mathrm{DS} = 1\,\mathrm{V}$, the threshold is extracted from each by the constant-current method, and the shift is divided by the bias difference,

$$\mathrm{DIBL} = -\frac{V_\mathrm{th}(1\,\mathrm{V}) - V_\mathrm{th}(50\,\mathrm{mV})}{1\,\mathrm{V} - 50\,\mathrm{mV}}. \tag{32}$$

With the estimated threshold shift of the order of 150 mV to 200 mV between the two biases at this length, the DIBL comes out in the range of a few hundred millivolts per volt, which is large and confirms that 100 nm is already well inside the short-channel regime for this oxide and body. The placeholder of Figure 7 is reserved for the two overlaid transfer curves and the extracted shift.

Plot to be inserted from Inspect / Sentaurus Visual:

Two transfer curves at $V_{\mathrm{DS}} = 50\,\mathrm{mV}$ and 1 V for $L_{\mathrm{g}} = 100\,\mathrm{nm}$, and DIBL versus L_{g} for 100 nm to 20 nm

Overlay the two semilog transfer curves and annotate the horizontal shift of the 100 nm reference. In a second panel, repeat the two-bias extraction at 100, 80, 60, 40 and 20 nm and plot the resulting DIBL, which should rise steeply as L_{g} falls.

Figure 7: Drain-induced barrier lowering: extraction at 100 nm and its growth as the channel is shortened toward 20 nm.

4.5 Effect of oxide thickness

Figure 8 shows the transfer characteristics of the 100 nm device at $V_{\mathrm{DS}} = 1\,\mathrm{V}$ as the gate oxide is thinned from 5 nm to 1 nm. The whole family pivots about a near-common point close to $V_{\mathrm{GS}} \approx 0.46\,\mathrm{V}$, which is the clean signature of a change that acts through the gate capacitance: below the pivot the thick-oxide curves sit higher because their subthreshold slope is flatter, while above it the thin-oxide curves overtake because the larger C_{ox} injects more inversion charge.

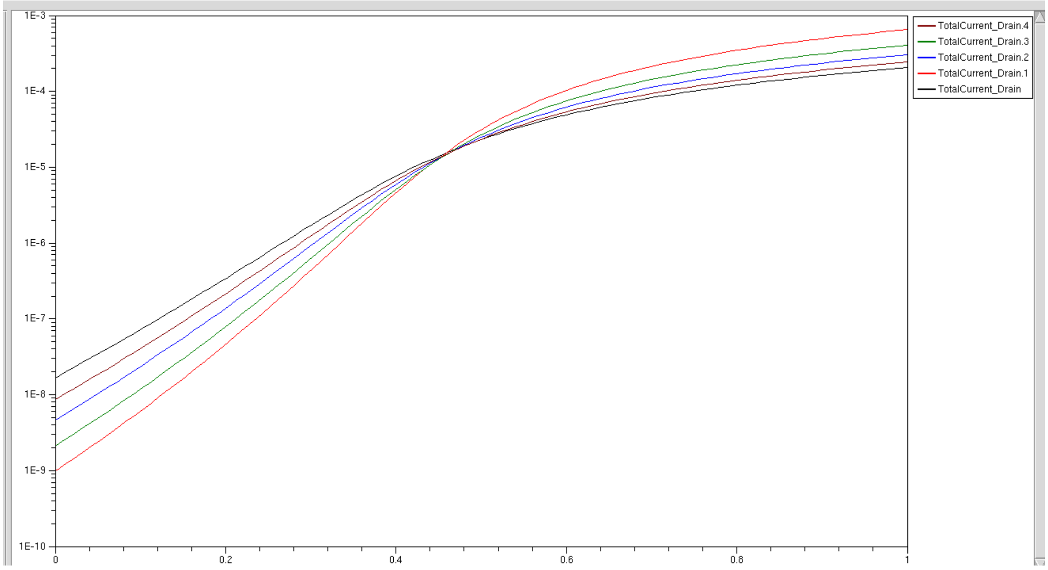


Figure 8: Transfer characteristics at $V_{\mathrm{DS}} = 1\,\mathrm{V}$ for the oxide thickness stepped from 5 nm to 1 nm (five values), $L_{\mathrm{g}} = 100\,\mathrm{nm}$. The curves pivot about a common point near 0.46 V. *To be confirmed by the author: the steepest curve, with the lowest off-current and highest on-current, is the 1 nm oxide.*

Two effects work together as the oxide is thinned. First, the body factor $m = 1 + C_{\mathrm{dep}}/C_{\mathrm{ox}}$ of (10) falls because $C_{\mathrm{ox}} = \varepsilon_{\mathrm{ox}}/t_{\mathrm{ox}}$ grows, so the subthreshold swing steepens and the off-current drops; the steepest curve in the figure therefore belongs to the 1 nm oxide. Second, the same increase in C_{ox} improves the electrostatic integrity directly, since the characteristic length $\ell = \sqrt{3 t_{\mathrm{ox}} X_{\mathrm{dep}}}$ of (12) shrinks with t_{ox}: going from 5 nm to 1 nm reduces ℓ from about 38.7 nm to

$$\ell_{1\,\mathrm{nm}} = \sqrt{3 \times 1\,\mathrm{nm} \times 100\,\mathrm{nm}} \approx 17.3\,\mathrm{nm}, \tag{33}$$

so that at $L_{\mathrm{g}} = 100\,\mathrm{nm}$ the ratio L_{g}/ℓ improves from 2.6 to 5.8 and the device is pushed back from the edge of the short-channel regime toward acceptable behaviour. This is the quantitative statement of why oxide thinning was historically the first lever pulled in every scaling generation:

it improves the swing and the DIBL at the same time. The placeholder of Figure 9 is left for the extracted swing and threshold against t_{ox}.

Plot to be inserted from Inspect / Sentaurus Visual:

SS and V_{th} versus t_{ox} from 5 nm to 1 nm

Extract the swing and the constant-current threshold from each curve of Figure 8 and plot them against oxide thickness; SS should fall toward the 60 mV/dec floor as $t_{\mathrm{ox}} \to 1\,\mathrm{nm}$.

Figure 9: Extracted subthreshold swing and threshold voltage as functions of oxide thickness.

4.6 Effect of substrate doping at $L_g = 20\,\mathrm{nm}$

The most instructive part of the scaling study is the 20 nm device, where the substrate doping was raised from 1×10^{15} to $1 \times 10^{18}/\mathrm{cm}^3$ at $V_{\mathrm{DS}} = 1\,\mathrm{V}$. The transfer characteristics are shown in Figure 10 and the extracted figures of merit in Table 3. The first thing to notice is that the vertical axis never falls below the milliampere range: even at $V_{\mathrm{GS}} = 0$ the device carries more than a milliampere per micrometre, and the gate is barely able to modulate the current at all. The on/off ratio has collapsed from the six or seven decades of the long-channel device to a factor of about two.

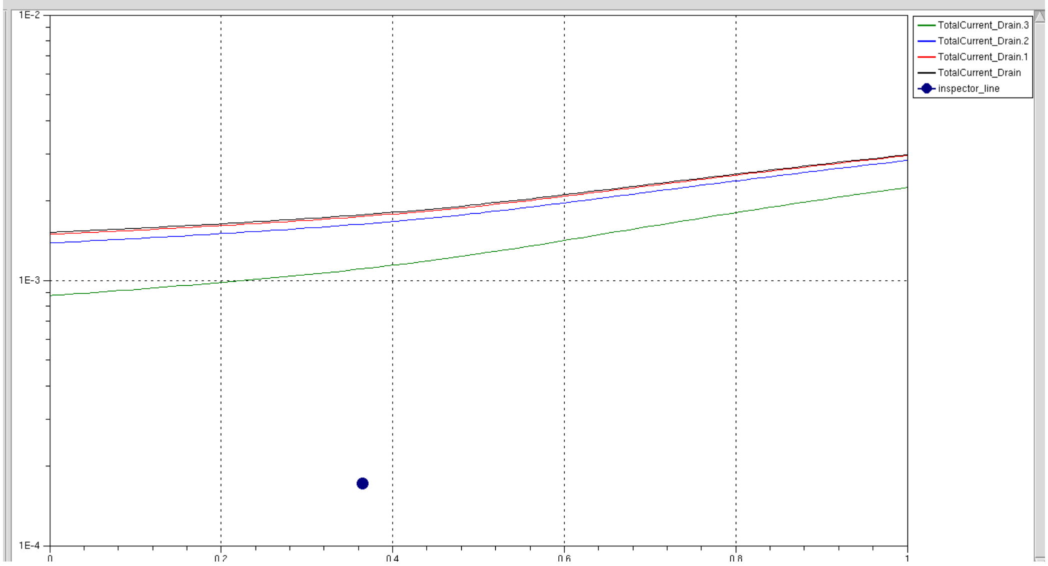


Figure 10: Transfer characteristics at $V_{\mathrm{DS}} = 1\,\mathrm{V}$ for the 20 nm device as the substrate doping is stepped from 1×10^{15} to $1 \times 10^{18}/\mathrm{cm}^3$. The current never drops below the milliampere range and the gate barely modulates it. The lowest (green) curve, with the smallest off-current, corresponds to the highest doping, $1 \times 10^{18}/\mathrm{cm}^3$.

Table 3: Figures of merit extracted in Inspect for the 20 nm device at $V_{\mathrm{DS}} = 1\,\mathrm{V}$, as functions of the substrate acceptor doping. Currents are per micrometre of width.

N_{A} (1/cm^3)	I_{on} (mA)	I_{off} (mA)	SS (mV/dec)
1×10^{15}	2.98	1.54	2890
1×10^{16}	2.98	1.48	2880
1×10^{17}	2.80	1.30	2730
1×10^{18}	2.20	0.87	2480

The trends in Table 3 are weak but consistent, and they say something precise. Raising N_{A} from 1×10^{15} to $1 \times 10^{18}/\mathrm{cm}^3$ lowers the off-current only from 1.54 to 0.87 mA and improves the swing only from 2890 to 2480 mV/dec: a factor of two in current and a fifteen percent improvement in swing, in exchange for a thousandfold change in doping. A swing of 2500 mV/dec is more than forty times the room-temperature limit of (10) and is not a subthreshold slope in any meaningful sense; the device has simply stopped switching. The reason is laid out in Table 4.

Table 4: Electrostatic length scales for the 20 nm device, computed from the analytic models of Section 2.4. The Fermi potential is from (1); the bulk depletion depth X_{dep} from (3), capped by the body thickness $t_{\mathrm{si}} = 100\,\mathrm{nm}$; the characteristic length ℓ from (12); and the drain junction depletion width W_{j} from (15) at $V_{\mathrm{DS}} = 1\,\mathrm{V}$.

N_{A} (1/cm^3)	ϕ_{F} (V)	X_{dep} (nm)	ℓ (nm)	$e^{-L/2\ell}$	W_{j} (nm)
1×10^{15}	0.298	100	38.7	0.77	1080
1×10^{16}	0.357	100	38.7	0.77	351
1×10^{17}	0.417	100	38.7	0.77	161
1×10^{18}	0.477	35	22.9	0.65	51

Two numbers in Table 4 explain the whole picture. First, the characteristic length is 38.7 nm for all dopings up to $1 \times 10^{17}/\mathrm{cm}^3$ (where X_{dep} is capped by the 100 nm body) and only falls to 22.9 nm at $1 \times 10^{18}/\mathrm{cm}^3$. In every case the gate length of 20 nm is smaller than a single characteristic length, so $L_{\mathrm{g}}/\ell < 1$ and the exponential barrier-lowering factor $\exp(-L_{\mathrm{g}}/2\ell)$ is of order 0.7: the drain has as much control over the channel barrier as the gate does. Raising the doping to 1×10^{18} helps a little, by shrinking X_{dep} and hence ℓ, which is exactly why that curve has the lowest off-current and the best swing in Table 3, but it is nowhere near enough. Second, and more seriously, the drain junction depletion width W_{j} is larger than the entire 20 nm channel for every doping up to $1 \times 10^{17}/\mathrm{cm}^3$ and is still half the channel at 1×10^{18}. The source and drain depletion regions therefore touch: the body is fully depleted from junction to junction, a buried current path opens below the surface that the gate cannot reach, and the device is in punch-through. The milliampere-level off-current of Figure 10 is this punch-through current, and no amount of channel doping within the range tried can close it, because closing it would require $W_{\mathrm{j}} < 10\,\mathrm{nm}$, that is a doping well into the $1 \times 10^{19}/\mathrm{cm}^3$ range, at which point the junctions would break down and the mobility would collapse. The lesson of this part of the experiment is the central lesson of device scaling: a 20 nm channel cannot be controlled by tuning a single planar parameter, and this is precisely the electrostatic dead end that drove the industry to thin bodies and multiple gates.

4.7 Effect of temperature

Heating the device from 250 to 400 K acts through the three channels set out in Section 2.5. The threshold falls at roughly $-1\,\mathrm{mV/K}$ to $-2\,\mathrm{mV/K}$ as the Fermi potential of (1) decreases, the subthreshold swing rises in direct proportion to T through (10) (a 60 mV/dec device at 300 K becomes about 80 mV/dec at 400 K), and the on-current falls as the phonon-limited mobility drops like $T^{-2.5}$. The competition between the falling threshold, which raises the current, and the falling mobility, which lowers it, produces a single gate voltage at which the current is temperature independent, the zero-temperature-coefficient point; below it the current has a positive temperature coefficient and above it a negative one.

Plot to be inserted from Inspect / Sentaurus Visual:

Transfer characteristics at $T = 250, 300, 350$ and $400\,\mathrm{K}$

Overlay the four semilog transfer curves; mark the ZTC crossing point, the leftward shift of threshold and the rising off-current floor as T increases.

Figure 11: Temperature dependence of the transfer characteristic, showing the zero-temperature-coefficient crossing.

4.8 Effect of reduced source/drain doping

Lowering the source and drain doping from 1×10^{20} to $1 \times 10^{19}/\mathrm{cm}^3$ acts almost entirely through the parasitic series resistance of the extensions, $R_{\mathrm{SD}} \propto 1/(q\mu_n N_{\mathrm{D}})$. The subthreshold region, governed by the channel barrier rather than by resistance, is essentially unchanged, so the off-current and the swing barely move; the visible effect is a reduction of the on-current at high V_{GS}, where the channel resistance is smallest and the now-larger R_{SD} takes a growing share of the applied bias, flattening the top of the transfer and output curves as described by the extrinsic relation of Section 2.4. There is also a slight reduction of the built-in potential, about 60 mV per decade of N_{D}, which marginally relaxes the short-channel terms, but this is a second-order effect next to the resistance.

Plot to be inserted from Inspect / Sentaurus Visual:

Transfer or output characteristic at $N_{\mathrm{D}} = 1 \times 10^{20}$ and $1 \times 10^{19}/\mathrm{cm}^3$

Overlay the two cases; the curves should coincide in subthreshold and separate at high V_{GS}, the lower-doped device carrying less on-current because of the larger series resistance.

Figure 12: Effect of reduced source/drain doping, acting mainly through parasitic series resistance.

4.9 Effect of body thickness

Finally, thinning the silicon body from 100 nm to 20 nm at the 20 nm gate length attacks the short-channel problem from the one direction the planar parameters could not. When the body is thinner than the depletion depth, X_{dep} in (12) is set by the geometry rather than by the doping, so the characteristic length falls with the body thickness: at $t_{\mathrm{si}} = 20\,\mathrm{nm}$ the depth $X_{\mathrm{dep}} \approx 20\,\mathrm{nm}$ gives $\ell = \sqrt{3 \times 5 \times 20} \approx 17.3\,\mathrm{nm}$, so that L_{g}/ℓ climbs back above one and the buried punch-through path is squeezed shut from below. This is the same physics that makes ultrathin-body silicon-on-insulator and FinFET devices work: removing the neutral body removes the leakage path that the gate cannot reach. The expected result is a recovery of the off-current and a steepening of the swing as t_{si} is reduced, in clear contrast to the near-helplessness of the substrate-doping lever seen in Section 4.6.

Plot to be inserted from Inspect / Sentaurus Visual:

Transfer characteristics for t_si stepped from 100 to 20 nm at $L_\mathrm{g} = 20\,\mathrm{nm}$

Overlay the curves; the off-current should fall and the swing steepen markedly as the body is thinned, recovering part of the gate control lost in Section 4.6.

Figure 13: Effect of body thickness on the 20 nm device, the only single-parameter lever that restores electrostatic integrity.

5 Conclusions

A two-dimensional n-channel MOSFET was designed on the MOS capacitor of the previous experiment and characterised end to end in the Sentaurus drift-diffusion flow, and the same structure was then driven deliberately into the short-channel regime. The long-channel baseline behaved as a textbook switch: a subthreshold swing of about 111 mV/dec was extracted as the reciprocal of the semilog slope, a threshold near 0.35 V was read by the constant-current method, and an on/off ratio of six to seven decades was obtained at a read-out bias of 50 mV. The internal snapshots confirmed the physical meaning of threshold, showing the channel barrier collapsing as the gate was raised through 0 V, 0.5 V and 1 V.

The scaling study then exposed, one lever at a time, why short channels are hard. Shortening the gate raised the off-current, lowered the threshold and degraded the swing, exactly as the quasi-two-dimensional characteristic-length model predicts once L_g falls below about five characteristic lengths; for the baseline body $\ell \approx 38.7\,\mathrm{nm}$, so 100 nm already showed a DIBL of a few hundred millivolts per volt. Thinning the oxide helped on two fronts at once, steepening the swing and shrinking the characteristic length to about 17.3 nm at 1 nm. The substrate-doping study at 20 nm was the decisive negative result: a thousandfold increase in N_A bought only a factor of two in off-current, because the channel was shorter than one characteristic length and the drain depletion region spanned the whole body, putting the device into punch-through that channel doping cannot cure. Only thinning the body, which forces the characteristic length down through geometry rather than doping, was seen to restore gate control. Taken together, the measurements reproduce in miniature the historical argument for abandoning the planar bulk transistor: below roughly 100 nm the electrostatics can no longer be fixed by tuning a single planar parameter, and the device must be re-engineered in three dimensions.

A number of figures in this report are reserved as placeholders for plots that are to be exported from Inspect and Sentaurus Visual, namely the internal band, potential and carrier cuts, the output characteristics, the parameter-versus-length and DIBL-versus-length curves, and the temperature, source/drain doping and body-thickness families. The analytic expectations stated alongside each placeholder provide the reference against which those plots should be checked once inserted.

References

[1] Z. H. Liu, C. Hu, J. H. Huang, T. Y. Chan, M. C. Jeng, P. K. Ko, and Y. C. Cheng, "Threshold voltage model for deep-submicrometer MOSFETs," *IEEE Trans. Electron Devices*, vol. 40, no. 1, pp. 86–95, 1993.

[2] R. R. Troutman, "VLSI limitations from drain-induced barrier lowering," *IEEE J. Solid-State Circuits*, vol. 14, no. 2, pp. 383–391, 1979.

[3] K. K. Young, “Short-channel effect in fully depleted SOI MOSFETs,” *IEEE Trans. Electron Devices*, vol. 36, no. 2, pp. 399–402, 1989.

[4] R. H. Yan, A. Ourmazd, and K. F. Lee, “Scaling the Si MOSFET: from bulk to SOI to bulk,” *IEEE Trans. Electron Devices*, vol. 39, no. 7, pp. 1704–1710, 1992.

[5] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, “Device scaling limits of Si MOSFETs and their application dependencies,” *Proc. IEEE*, vol. 89, no. 3, pp. 259–288, 2001.

[6] G. Masetti, M. Severi, and S. Solmi, “Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon,” *IEEE Trans. Electron Devices*, vol. 30, no. 7, pp. 764–769, 1983.

[7] C. Canali, G. Majni, R. Minder, and G. Ottaviani, “Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature,” *IEEE Trans. Electron Devices*, vol. 22, no. 11, pp. 1045–1047, 1975.

[8] C. Lombardi, S. Manzini, A. Saporito, and M. Vanzi, “A physically based mobility model for numerical simulation of nonplanar devices,” *IEEE Trans. Computer-Aided Design*, vol. 7, no. 11, pp. 1164–1171, 1988.

[9] J. W. Slotboom and H. C. de Graaff, “Measurements of bandgap narrowing in Si bipolar transistors,” *Solid-State Electronics*, vol. 19, no. 10, pp. 857–862, 1976.

[10] J. G. Fossum, “Computer-aided numerical analysis of silicon solar cells,” *Solid-State Electronics*, vol. 19, no. 4, pp. 269–277, 1976.

[11] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[12] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. Hoboken, NJ, USA: Wiley, 2007.

[13] Synopsys, Inc., *Sentaurus Device User Guide*. Mountain View, CA, USA: Synopsys, 2022.

[14] Synopsys, Inc., *Sentaurus Training Manual: MOSFET Simulation (sd module)*, 2022. [Online]. Available: `https://kolegite.com/EE_library/books_and_lectures/Sentaurus_Training/sd/sd_3.html`

A Sentaurus Structure Editor command file

The structure was generated by the following SDE (Scheme) command file. The device parameters appear as Sentaurus Workbench variables (the `@name@` tokens), so a single file produces the whole parametric study: `@Lg@` is the gate length, `@tsi@` the body thickness, `@tox@` the oxide thickness, and `@NA@` the substrate doping. The listing has been re-indented for readability; its logic is unchanged.

```
; ====================================================================
;  SDE : Experiment 4 - MOSFET structure
;  Parameters are Sentaurus Workbench variables (@name@)
; ====================================================================

; -------------------- parameters (microns) -------------------------
(define Lg   @Lg@)        ; gate length, baseline 1.0 um
(define Lsd  0.5)         ; source/drain extension length
(define tsi  @tsi@)       ; silicon body thickness, baseline 0.1 um
(define tox  @tox@)       ; gate oxide thickness, baseline 0.005 um
(define tsd  0.025)       ; source/drain junction depth, 25 nm
(define NA   @NA@)        ; substrate acceptor (boron) doping
(define ND   1e20)        ; source/drain donor (arsenic) doping

; derived x-coordinates of the lateral structure
(define x0   0.0)
(define x1   Lsd)         ; source / channel boundary
(define x2   (+ Lsd Lg))  ; channel / drain boundary
(define x3   (+ (* 2 Lsd) Lg))
(define ytop 0.0)
(define ybot tsi)

; -------------------- silicon body ---------------------------------
(sdegeo:create-rectangle
  (position x0 ytop 0) (position x3 ybot 0)
  "Silicon" "body")

; -------------------- gate oxide over the channel ------------------
(sdegeo:create-rectangle
  (position x1 (- ytop tox) 0) (position x2 ytop 0)
  "SiO2" "gateox")

; -------------------- constant doping profiles ---------------------
; uniform p-type substrate
(sdedr:define-constant-profile
  "substrate_doping" "BoronActiveConcentration" NA)
(sdedr:define-constant-profile-region
  "substrate_region" "substrate_doping" "body")

; n+ source well (top 25 nm of the source extension)
(sdedr:define-refinement-window
  "win_src" "Rectangle"
  (position x0 ytop 0) (position x1 tsd 0))
(sdedr:define-constant-profile
  "src_doping" "ArsenicActiveConcentration" ND)
(sdedr:define-constant-profile-placement
  "src_place" "src_doping" "win_src")

; n+ drain well (top 25 nm of the drain extension)
(sdedr:define-refinement-window
  "win_drn" "Rectangle"
  (position x2 ytop 0) (position x3 tsd 0))
(sdedr:define-constant-profile
  "drn_doping" "ArsenicActiveConcentration" ND)
(sdedr:define-constant-profile-placement
  "drn_place" "drn_doping" "win_drn")

; -------------------- mesh refinement windows ----------------------
; background mesh of the bulk: 20 nm x 5 nm
(sdedr:define-refinement-size "mesh_si" 0.020 0.005 0.020 0.005)
(sdedr:define-refinement-region "ref_si" "mesh_si" "body")

; channel surface: top 30 nm, 10 nm x 1 nm to resolve inversion layer
(sdedr:define-refinement-window
  "ch_win" "Rectangle"
```

```
  (position x1 ytop 0) (position x2 0.030 0))
(sdedr:define-refinement-size "mesh_ch" 0.010 0.001 0.010 0.001)
(sdedr:define-refinement-placement "ref_ch" "mesh_ch" "ch_win")

; oxide: 10 nm x 1 nm
(sdedr:define-refinement-size "mesh_ox" 0.010 0.001 0.010 0.001)
(sdedr:define-refinement-region "ref_ox" "mesh_ox" "gateox")

; junction sidewalls: keep doping change below one decade per element
(sdedr:define-refinement-function "mesh_si" "MaxTransDiff" 1)
(sdedr:define-refinement-function "mesh_ch" "MaxTransDiff" 1)

; -------------------- contacts -------------------------------------
(sdegeo:define-contact-set "Source"    4 (color:rgb 1 0 0) "##")
(sdegeo:define-contact-set "Drain"     4 (color:rgb 0 0 1) "##")
(sdegeo:define-contact-set "Gate"      4 (color:rgb 0 1 0) "##")
(sdegeo:define-contact-set "Substrate" 4 (color:rgb 1 1 0) "##")

(sdegeo:set-current-contact-set "Source")
(sdegeo:define-2d-contact
  (find-edge-id (position (/ x1 2) ytop 0)) "Source")
(sdegeo:set-current-contact-set "Drain")
(sdegeo:define-2d-contact
  (find-edge-id (position (+ x2 (/ Lsd 2)) ytop 0)) "Drain")
(sdegeo:set-current-contact-set "Gate")
(sdegeo:define-2d-contact
  (find-edge-id (position (/ (+ x1 x2) 2) (- ytop tox) 0)) "Gate")
(sdegeo:set-current-contact-set "Substrate")
(sdegeo:define-2d-contact
  (find-edge-id (position (/ x3 2) ybot 0)) "Substrate")

; -------------------- build the mesh ------------------------------
(sde:build-mesh "snmesh" "" "n@node@_msh")
```

B Sentaurus Device command file

The electrical simulation was driven by the following SDevice command file. The `Physics` section selects the models discussed in Section 3.3, and the `Solve` block executes the four-stage bias sequence of Section 3.4. The listing is reproduced with corrected indentation; comments that previously used long dashes have been rewritten with colons.

```
#===================== SDEVICE : Experiment 4 - MOSFET ==============
File {
  Grid      = "@tdr@"
  Plot      = "@tdrdat@"
  Current   = "@plot@"
  Output    = "@log@"
  Parameter = "@parameter@"
}

#-------------------- Electrodes ----------------------------------
Electrode {
  { Name="Source"    Voltage=0.0 }
  { Name="Drain"     Voltage=0.0 }
  { Name="Gate"      Voltage=0.0 }
  { Name="Substrate" Voltage=0.0 }
}

#-------------------- Physics Models ------------------------------
Physics {
  Mobility(
    DopingDependence
    eHighFieldSaturation
    Enormal
  )
  Recombination( SRH( DopingDependence ) )
  EffectiveIntrinsicDensity( BandGapNarrowing(OldSlotboom) )
  Fermi
  Temperature = @temp@
}
```

```

#-------------------- Plot Quantities -------------------------------
Plot {
  ConductionBandEnergy   ValenceBandEnergy
  ElectrostaticPotential ElectricField
  eDensity    hDensity
  eCurrent    hCurrent
  eMobility   hMobility
  eVelocity   hVelocity
  eQuasiFermi hQuasiFermi
  Doping  DonorConcentration  AcceptorConcentration
  SpaceCharge
}

#-------------------- Numerical Settings ----------------------------
Math {
  Extrapolate
  Derivatives
  RelErrControl
  Digits=5
  Iterations=100
  NotDamped=50
  ErRef(electron)=1e10
  ErRef(hole)=1e10
  DirectCurrent
  Method=ParDiSo
  Number_of_Threads=4
}

#====================== Solve Block =================================
Solve {
  * Step 1: Equilibrium
  NewCurrentPrefix="init_"
  Coupled(Iterations=100) { Poisson }
  Coupled { Poisson Electron Hole }
  Save(FilePrefix="n@node@_eq")

  * Step 2: Ramp VDS to target (50 mV), silent prefix
  NewCurrentPrefix="vds_ramp_"
  QuasiStationary(
    InitialStep=0.01
    Increment=1.5
    MinStep=1e-5
    MaxStep=0.1
    Goal { Name="Drain" Voltage=@Vds@ }
  ) { Coupled { Poisson Electron Hole } }
  Save(FilePrefix="n@node@_vds_ready")

  * Step 3: IDS-VGS sweep. VGS ramps 0 to Vgs_max, VDS fixed.
  * This is the ONLY step writing to the main current file.
  NewCurrentPrefix=""
  QuasiStationary(
    InitialStep=0.005
    Increment=1.5
    MinStep=1e-5
    MaxStep=0.01
    Goal { Name="Gate" Voltage=@Vgs_max@ }
  ) {
    Coupled { Poisson Electron Hole }
    Plot(
      FilePrefix = "n@node@_banddiag"
      Time       = (0.0; 0.5; 1.0)
      NoOverwrite
    )
  }

  * Step 4: Reload equilibrium for IDS-VDS sweep
  Load(FilePrefix="n@node@_eq")

  * Step 4a: Ramp VGS to fixed overdrive, separate prefix, not plotted
  NewCurrentPrefix="IdVd_setup_"
  QuasiStationary(
    InitialStep=0.01
    Increment=1.5
```

```
    MinStep=1e-5
    MaxStep=0.05
    Goal { Name="Gate" Voltage=@Vgs_fixed@ }
  ) { Coupled { Poisson Electron Hole } }

  * Step 4b: VDS sweep. This is the IDS-VDS output.
  NewCurrentPrefix="IdVd_"
  QuasiStationary(
    InitialStep=0.01
    Increment=1.5
    MinStep=1e-5
    MaxStep=0.02
    Goal { Name="Drain" Voltage=1.0 }
  ) { Coupled { Poisson Electron Hole } }
}
```