Virtual Process Integration (VPI)

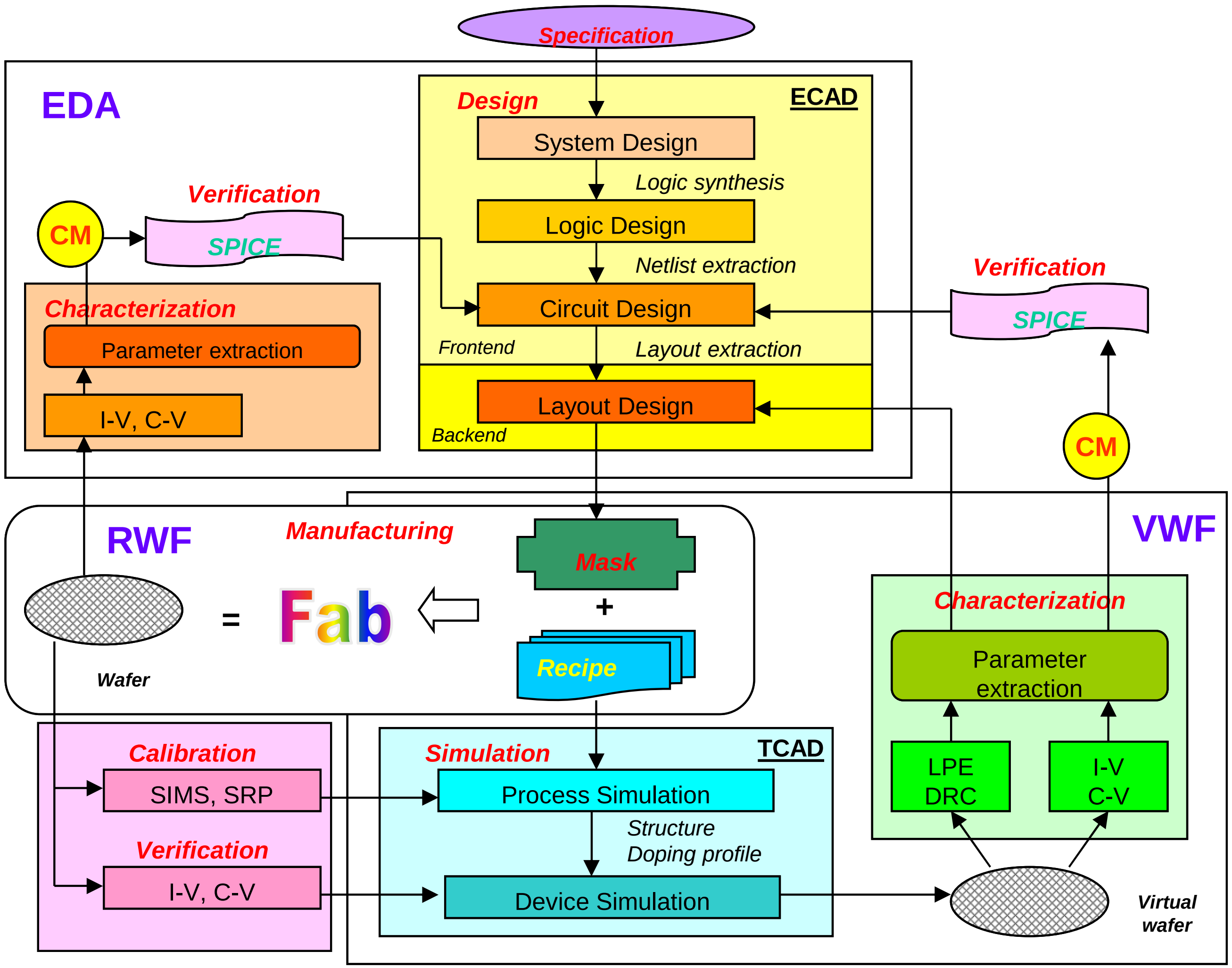
Dr Zhou Xing

Office: S1-B1c-95
Phone: 6790-4532
Email: exzhou@ntu.edu.sg

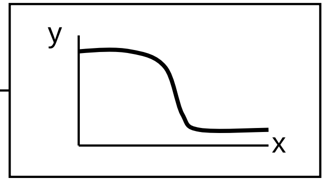
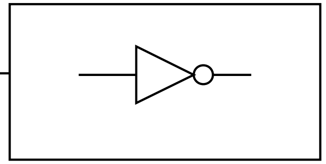
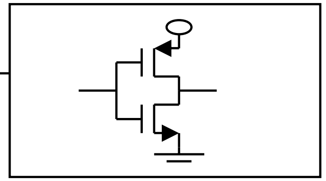
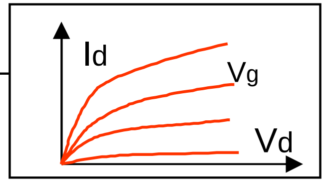
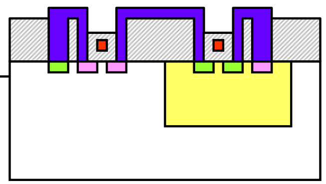
Web: https://www3.ntu.edu.sg/home/exzhou/Teaching/TUM-NM6604/

Overall Picture: Chip Design and Wafer Fabrication

Design–Manufacturing–Characterization–Simulation–Verification

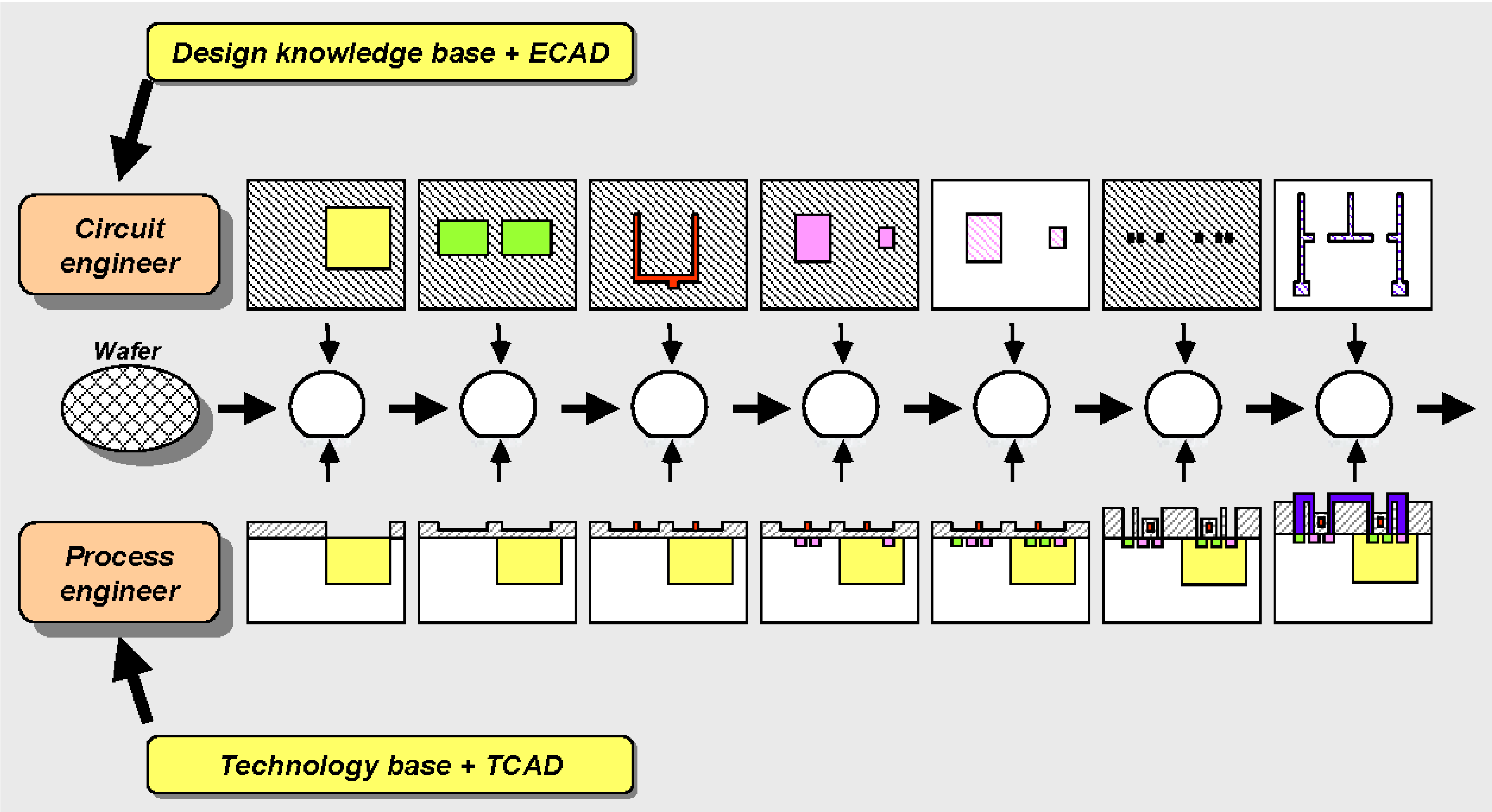


Multi-Level Representation

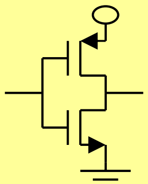
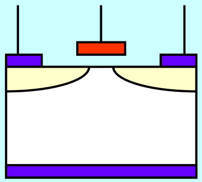
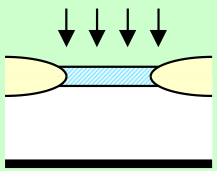
Level	*Representation*	*Function*	*Graphical*
System	**Functional**	$y = f(x)$	
Logic	**Boolean**	x | y: 0 | 1; 1 | 0	
Circuit	**Transistor**	$[I] = [Y][V]$	
Device	**Analytical**	Ids = f(Vgs, Vds, Vbs, W, L, T)	
Technology	**Device**	$\nabla^2\psi = -\rho/\varepsilon$ $\frac{d\rho}{dt} + \nabla\cdot \mathbf{J} = G - U$ $\mathbf{J} = q\mu n\nabla\psi + qD\nabla n$	

System → Logic → Circuit → Device → Technology

Layout + Process = Chip

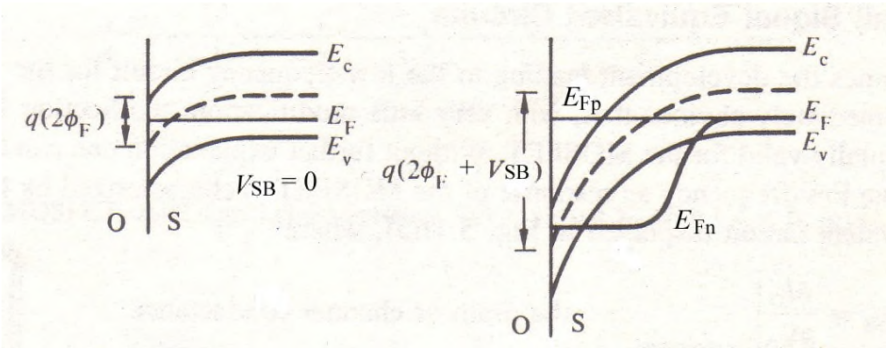


Target–Variable Relationship

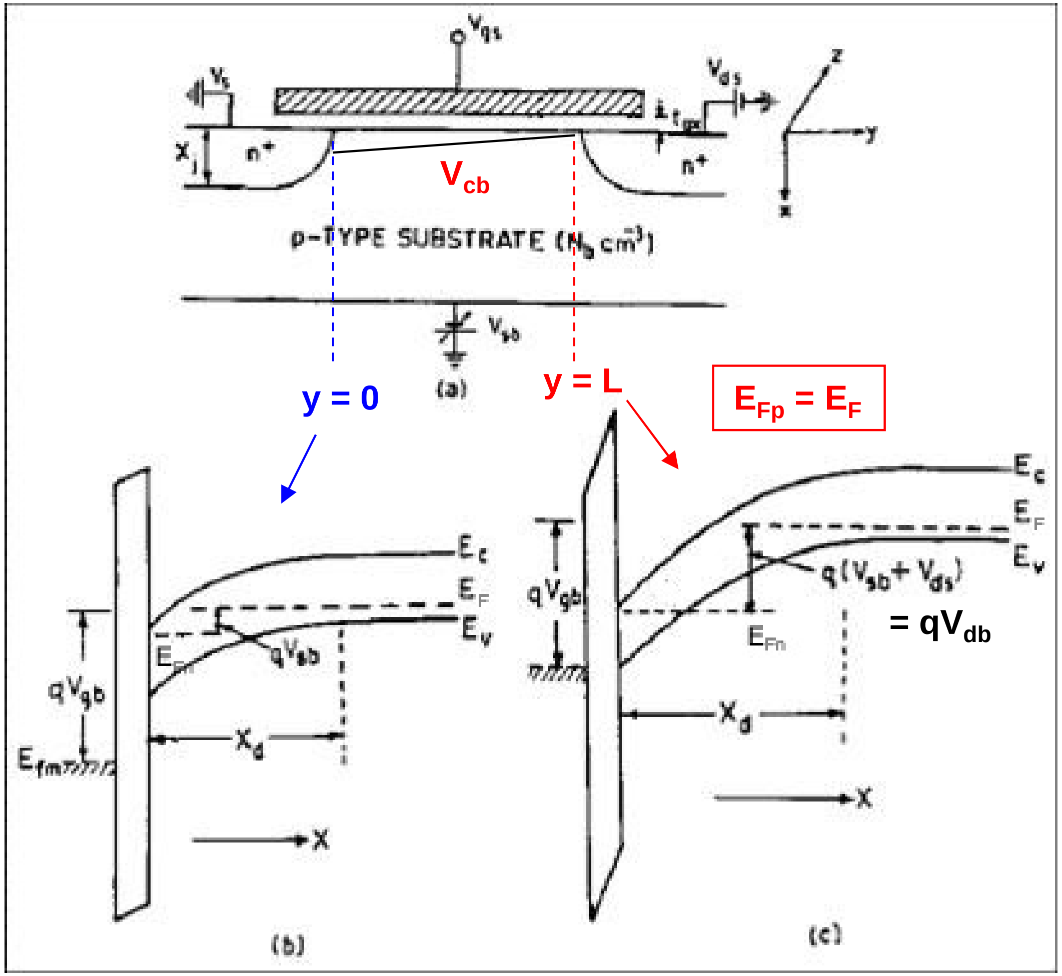
Level	Variables	Targets
Circuit 	Spice: Model parameters, ... Geometrical: Channel length, width, ... Electrical: Supply voltage, substrate bias, ...	Digital: Delay, rise/fall time, drivability, off-state current, noise margin, ... Analog: Voltage gain, cutoff frequency, slew rate, gain-bandwidth, ...
Device 	Structural: Oxide thickness, junction depth, sheet resistance, ... Doping: Peak/surface concentration, ... Electrical: Supply voltage, substrate bias, ...	Electrical: Threshold, transconductance, subthreshold swing, saturation current, punchthrough current, junction capacitance, lifetime, ... Physical: Potential, field, charge, current, carriers, velocity, ...
Process 	Oxidation: Temperature, time, ambient, ... Implantation: Dose, energy, tilt, damage, ... Diffusion: Defect, stress, OED, TED, ...	Layer: Oxide thickness, junction depth, sheet resistance, ... Profile: Peak/surface concentration, projected range/straggle, ...

MOSFET Operation: Due to Inversion Carrier Imref-Split

- MOSFET at equilibrium ($V_{ds} = 0$): no current flow even if channel is created at $V_{gs} = V_t$ ($\psi_s = 2\phi_F$)
- When $V_{sb} \neq 0$ ($V_{SB} > 0$ in NMOS), electron imref will "split" from hole imref with $qV_{sb} = E_{Fn} - E_{Fp}$, so $\psi_s = 2\phi_F + V_{sb}$.



- When $V_{ds} \neq 0$ ($V_{ds} > 0$ in NMOS), holes are still at quasi-equilibrium (since no 'source' nor 'drain'), so we can assume $E_{Fp} = E_F$. However, electron imref will change from V_{sb} at source end to $V_{db} = V_{sb} + V_{ds}$ at drain end relative to E_F, and varying along the channel as $V_{cb}(y)$ ['c' stands for '<u>c</u>hannel'].
- It is the <u>gradient</u> of $V_{cb}(y)$ that drives electrons *drifting/diffusing* from source to drain along y.



Key to understanding MOSFET operation: Band diagram in the x direction along a cutline at (b) source-end (y = 0) and (c) drain-end (y = L).

MOSFET Source-Referenced Threshold Voltage

MOSFET threshold voltage definition (source-referenced)

We define the **threshold voltage** (V_t) to be the gate-to-*source* voltage (V_{gs}) at which source-end surface potential is equal to twice of the bulk Fermi potential ($2\phi_F$) with reference to source–bulk voltage V_{sb}.

Potential balance Gauss law Charge balance Charge-sheet approximation

$$\left(V_{gs}+V_{sb}\right)-V_{FB}=V_{gb}-V_{FB}\equiv V_{gf}=V_{ox}+\psi_s=Q_g/C_{ox}+\psi_s=-Q_{sc}/C_{ox}+\psi_s\approx -Q_b/C_{ox}+\psi_s$$

$$=-\left(-qN_AX_d\right)/C_{ox}+\psi_s=+\sqrt{2q\varepsilon_{Si}N_A\psi_s}/C_{ox}+\psi_s$$

$$V_{gs}=V_{FB}-V_{sb}-Q_b/C_{ox}+\psi_s$$

Full-depletion approximation $X_d=\sqrt{2\varepsilon_{Si}\psi_s/qN_A}$

$$V_t\equiv V_{gs}\Big|_{\psi_s=2\phi_F+V_{sb}}=V_{FB}-V_{sb}+\left[-Q_b\left(\psi_s\right)/C_{ox}+\psi_s\right]\Big|_{\psi_s=2\phi_F+V_{sb}}=V_{FB}-V_{sb}-Q_b\left(\psi_s=2\phi_F+V_{sb}\right)/C_{ox}+\left(2\phi_F+V_{sb}\right)$$

$$\therefore\quad V_t\equiv V_{gs}\Big|_{\psi_s=2\phi_F+V_{sb}}=V_{FB}+\Upsilon\sqrt{2\phi_F+V_{sb}}+2\phi_F$$

where $\Upsilon=\sqrt{2q\varepsilon_{Si}N_A}/C_{ox}$ is the **body factor**.

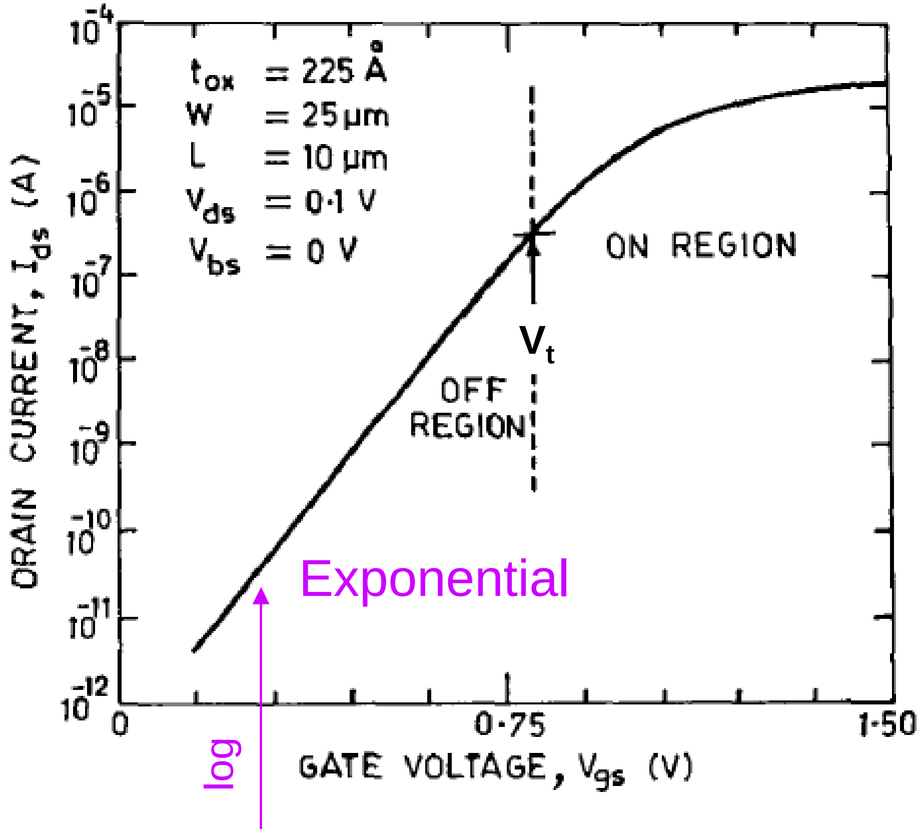
Body effect — Threshold-voltage shift due to non-zero V_{sb}

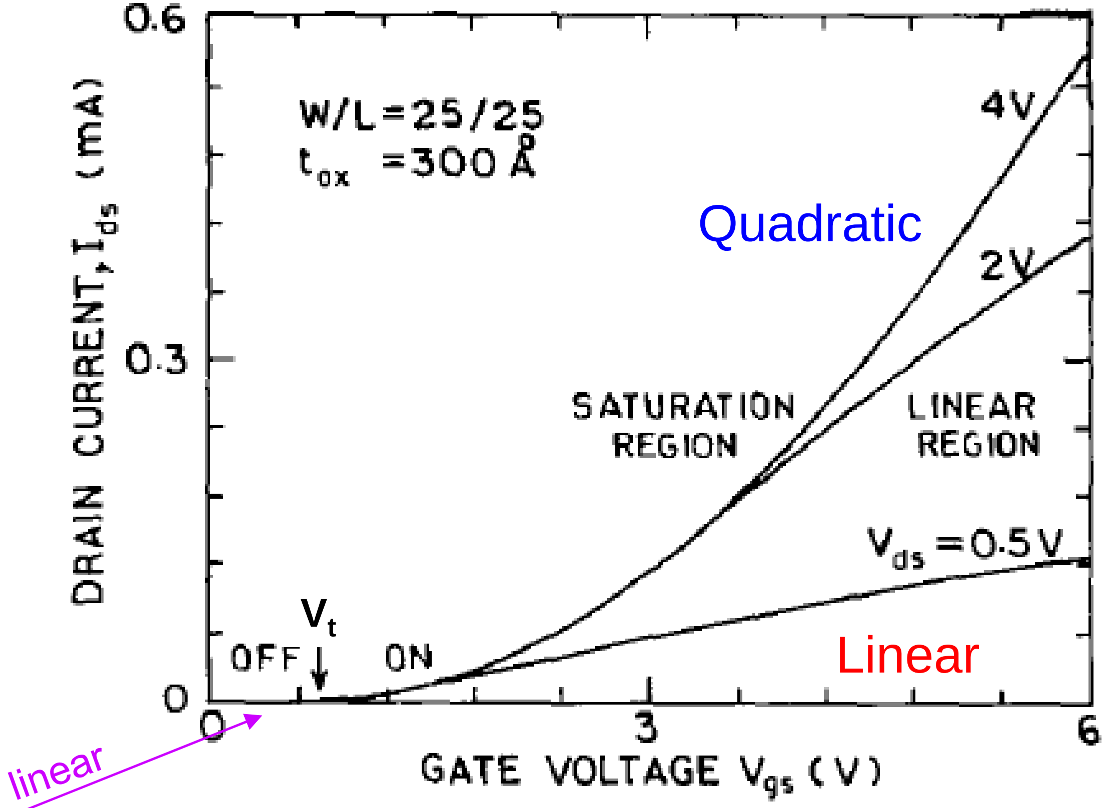
For NMOS, $V_{sb}>0$ so that source/drain-to bulk diodes always reverse biased.

$$V_t\left(V_{sb}\right)=V_{t0}+\Upsilon\left(\sqrt{2\phi_F+V_{sb}}-\sqrt{2\phi_F}\right)$$

$$V_{t0}\equiv V_t\Big|_{V_{sb}=0}=V_{FB}+\Upsilon\sqrt{2\phi_F}+2\phi_F$$

Regions of Operation: Transfer Characteristics





Low gate–source bias ($V_{gs} < V_t$): No inversion layer; diffusion dominant. MOS behaves like a wide-base (long-channel) BJT with $I_{ds} \propto \exp[q(V_{gs} - V_t)/nkT]$, $\mathbf{n = 1 + C_d/C_{ox}}$.

High drain–source bias ($V_{ds} > V_{dsat}$): Drain side “pinched-off”. MOS behaves like a current source.

Low drain–source bias ($V_{ds} < V_{dsat}$): Full channel. MOS behaves like a voltage-controlled resistor.

Current–Voltage in Linear (Triode) Region

- **MOSFET analysis** — major assumptions (NMOS as example)
  - ***"GCA" – Gradual Channel Approximation***: $dE_y/dy \ll dE_x/dx$
  - ***"Unipolar" – hole current can be neglected*** ($E_{Fp} \approx E_F$) in normal region (excluding breakdown)
  - ***Built-in voltages for the source/drain diodes can be ignored*** *(long channel)*
  - ***No recombination/generation*** *and* ***constant mobility***
  - ***Current flows in the y direction only***
- **First-order equation derivation**
  - ***Charge-sheet approximation (CSA)***
  - ***"Pinned" surface potential*** *at strong inversion ($2\phi_F$)*
  - ***Constant bulk charge along channel***
  - ***Drift-current only in linear region***

$$\boxed{\psi_s(y) = \psi_s(0) + V_{cb}(y) = 2\phi_F + V_{sb} + V(y)} \quad (0 \le V \le V_{ds})$$

$$V_{gb} - V_{FB} = V_{ox} + \psi_s = Q_g / C_{ox} + \psi_s = -(Q_b + Q_i)/C_{ox} + \psi_s$$

$$Q_i = -C_{ox}\left(V_{gb} - V_{FB} - \psi_s\right) - Q_b \qquad \boxed{Q_b \approx -\Upsilon C_{ox}\sqrt{2\phi_F + V_{sb}}}$$

$$= -C_{ox}\left[V_{gb} - V_{FB} - 2\phi_F - V_{sb} - V(y) - \Upsilon\sqrt{2\phi_F + V_{sb}}\right]$$

$$= -C_{ox}\left[V_{gs} - V_t - V(y)\right] \qquad \boxed{V_t \equiv V_{FB} + \Upsilon\sqrt{2\phi_F + V_{sb}} + 2\phi_F}$$

$$I_{ds}(y) \approx W\int_0^\infty J_{n,drift}(y)\,dx = W\int_0^\infty qn(x,y)\,\mu_n\left(-d\psi_s/dy\right)dx$$

$$= -W\mu_n Q_i(y)\,dV/dy \qquad \left[Q_i(y) \equiv \int_0^\infty qn(x,y)\,dx\right]$$

$$I_{ds} = \frac{W}{L}\mu_n\int_0^{V_{ds}} -Q_i(y)\,dV \qquad \left(\int_0^L dy \sim \int_{\psi_s(0)}^{\psi_s(L)} d\psi_s = \int_{V_{sb}}^{V_{db}} dV_{cb} = \int_0^{V_{ds}} dV\right)$$

Linear law (I_{ds} is a *linear* function of V_{gs}) ["Sah equation"]:

$$\boxed{I_{ds} = \mu_n C_{ox}\frac{W}{L}\left(V_{gs} - V_t - \frac{1}{2}V_{ds}\right)V_{ds}} \qquad \left(V_{gs} > V_t, V_{gd} > V_t\right)$$

Current–Voltage in Saturation (Pinch-off) Region

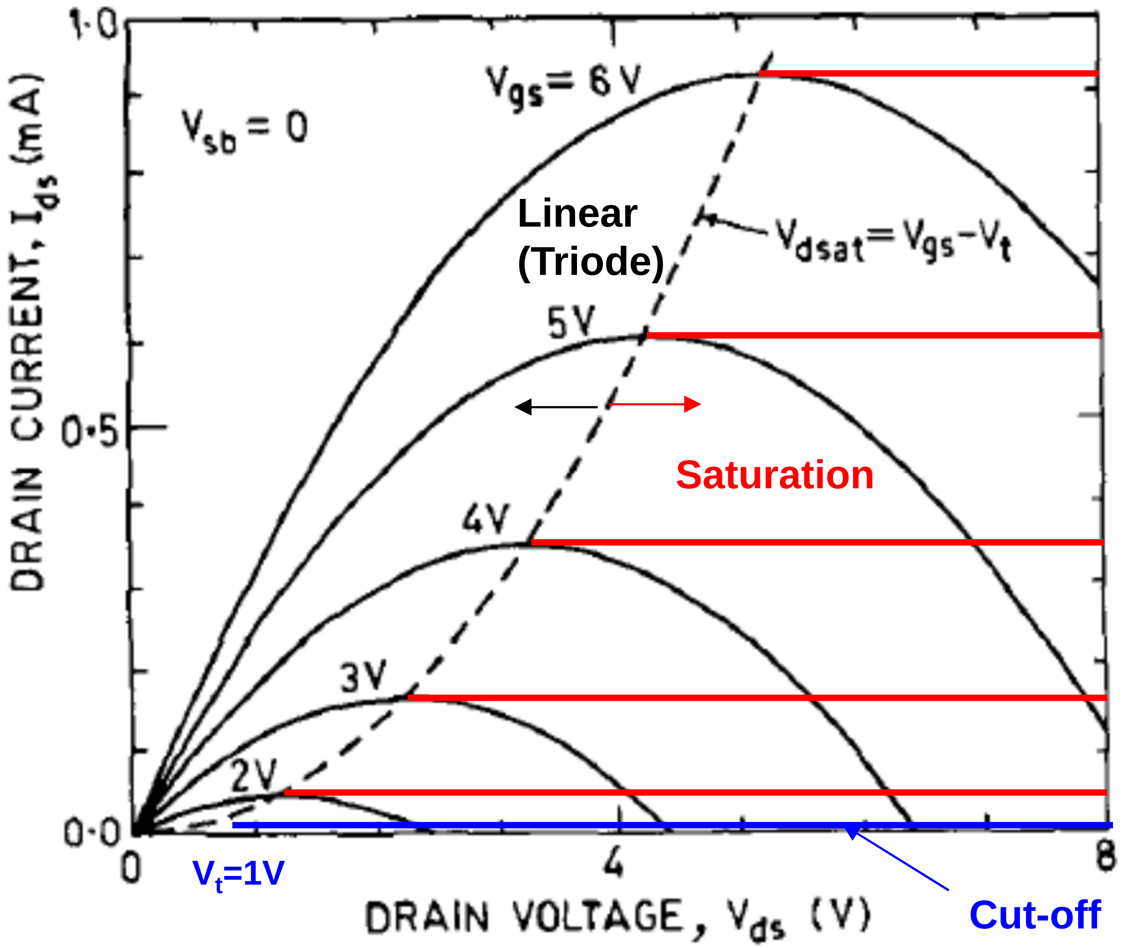
The peak of the linear current is reached when $dI_{ds}/dV_{ds} = \mu_n C_{ox} W/L\left(V_{gs} - V_t - V_{ds}\right) = 0$

For $V_{ds} \geq V_{gs} - V_t \equiv V_{dsat}$, GCA is not valid. Also, $Q_i\left(V = V_{dsat}\right) \approx 0$, channel is said to be "pinched-off." V_{dsat} is called *saturation* or *pinch-off voltage*, and the corresponding current is the *saturation current*:

Square law (I_{ds} is a *quadratic* function of V_{gs}):

$$I_{ds} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{gs} - V_t\right)^2 = I_{dsat} \qquad \begin{pmatrix} V_{gs} > V_t, \\ V_{gd} < V_t \end{pmatrix}$$

The "pinch-off" picture (Q_i = 0 assumption) is not physically correct since it requires the field to be infinite $E_y\left(y\right) = J_{ds}\left(y\right)/\mu_n Q_i\left(y\right)$ at pinch-off and carriers travel with infinite drift velocity. A more correct picture is that Q_i at pinch-off is very small but finite, with carriers drift under the large field in the pinch-off region at a saturated velocity.



MOSFET first-order piece-wise linear/square-law model.

Velocity Saturation and Saturation Current

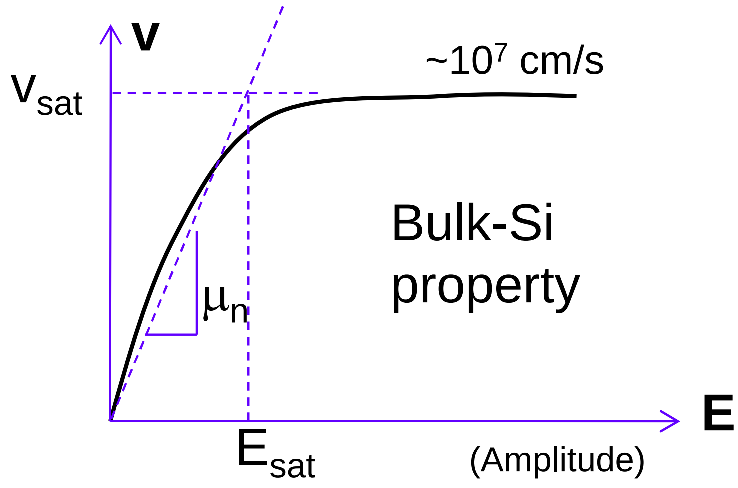
❑ Velocity–field relation — piecewise model

$$v = \begin{cases} \dfrac{\mu_n E}{1+E/E_{sat}} & E < E_{sat} \\ v_{sat} & E \geq E_{sat} \end{cases}$$

$$I_{ds}(y) \approx -WQ_i(y)\frac{\mu_n E}{1+E/E_{sat}}$$

$$I_{ds}(y)\left(1+\frac{1}{E_{sat}}\frac{dV}{dy}\right) = -WQ_i(y)\mu_n\frac{dV}{dy}$$

$$Q_i = -C_{ox}\left(V_{gb} - V_{FB} - (2\phi_F + V_{sb} + V)\right) - Q_b$$



- ***Saturation field***

$$v_{sat} = \frac{\mu_n E_{sat}}{1+E_{sat}/E_{sat}} \rightarrow E_{sat} = \frac{2v_{sat}}{\mu_n}$$

- ***Lateral-field mobility***

$$Q_b \approx -\Upsilon C_{ox}\sqrt{2\phi_F + V_{sb} + V}$$

$$\mu_{eff} = \frac{\mu_n}{1+V_{ds}/\left(E_{sat}L_{eff}\right)} \qquad A_b = 1 + \frac{\Upsilon}{2\sqrt{2\phi_F + V_{sb}}}$$

$$I_{ds} = \mu_{eff}C_{ox}\frac{W}{L}\left(V_{gs} - V_t - \frac{1}{2}A_b V_{ds}\right)V_{ds} \quad (1)$$

❑ Saturation current

$$I_{dsat} = -Wv_{sat}Q_{sat} = Wv_{sat}C_{ox}\left(V_{gs} - V_t - A_b V_{dsat}\right) \quad (2)$$

(1)(V_{dsat}) = (2): $V_{dsat} = \dfrac{E_{sat}L_{eff}\left(V_{gs} - V_t\right)}{V_{gs} - V_t + A_b E_{sat} L_{eff}}$

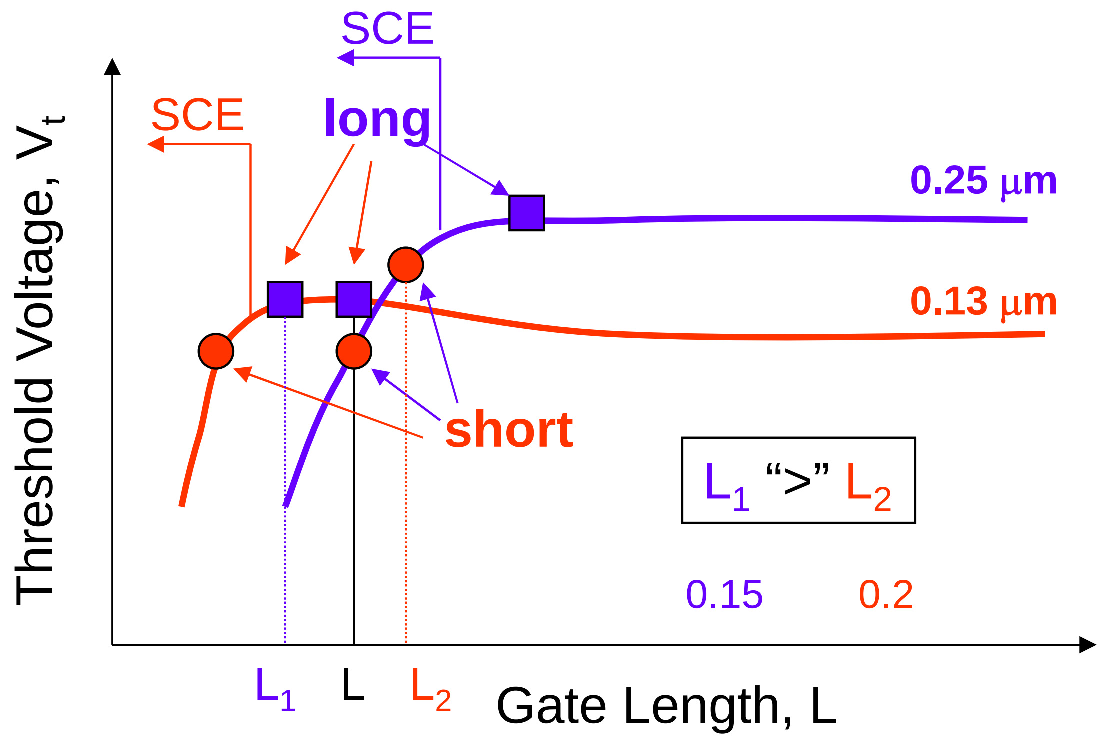
$$I_{dsat} = Wv_{sat}C_{ox}\frac{\left(V_{gs} - V_t\right)^2}{V_{gs} - V_t + A_b E_{sat} L_{eff}} \xrightarrow{L \rightarrow 0} \propto \left(V_{gs} - V_t\right)$$

Linear!

Long-Channel or Short-Channel?

- **Short-channel effect (SCE)** — technology dependent (depends on where the device "sits" on the V_t – L curve, not the actual dimension)

- **Technology scaling** — optimization for each technology node



Charge-Sharing Model: V_t "Roll-Off"

- **Charge-sharing model**
  - ***Without charge-sharing***

$$V_t = V_{FB} - Q_{bm}/C_{ox} + 2\phi_F$$

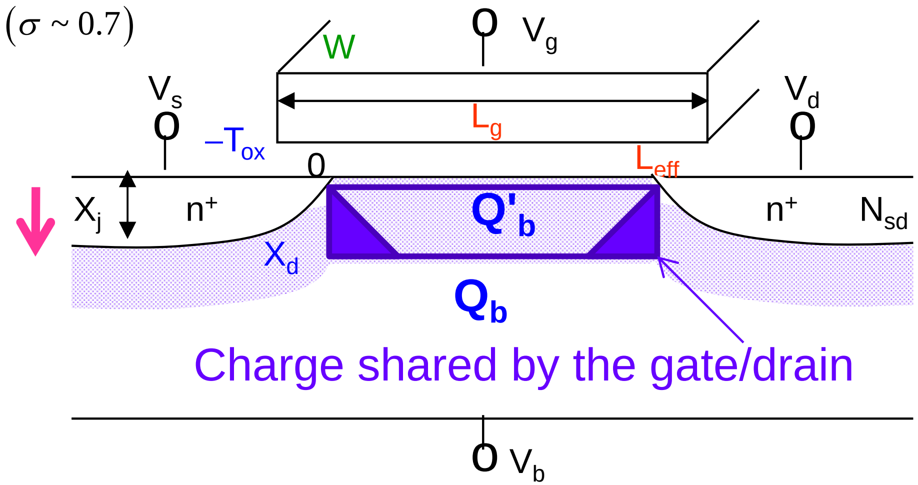
  - ***With charge-charing***

$$V_t' = V_{FB} - Q_{bm}'/C_{ox} + 2\phi_F$$

$$L_{eff} = L_g - 2\sigma X_j \quad (\sigma \sim 0.7)$$

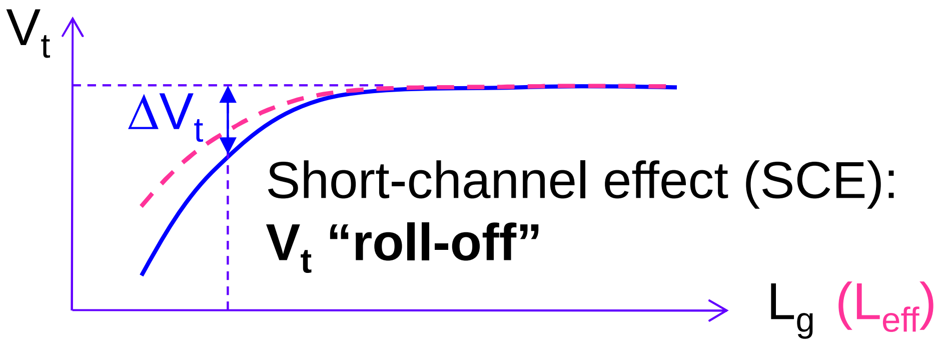
$$C_{ox} = \varepsilon_{ox}/T_{ox} \qquad Q_{bm} = -qN_A X_{dm}$$

$$X_{dm} = \sqrt{2\varepsilon_{Si}\left(2\phi_F + V_{sb}\right)/qN_A}$$



$$\Delta V_t \equiv V_t - V_t' = -\frac{Q_{bm}}{C_{ox}}\left(1 - \frac{Q_{bm}'}{Q_{bm}}\right) = -\frac{Q_{bm}}{C_{ox}}\frac{X_{dm}}{L_{eff}}$$

$$= \frac{qN_A X_{dm}}{\varepsilon_{ox}/T_{ox}}\frac{X_{dm}}{L_{eff}} = \frac{4\varepsilon_{Si}\phi_F}{\varepsilon_{ox}}\frac{T_{ox}}{L_g - 2\sigma X_j}$$



Simple "Triangle" Model

Total bulk charge:

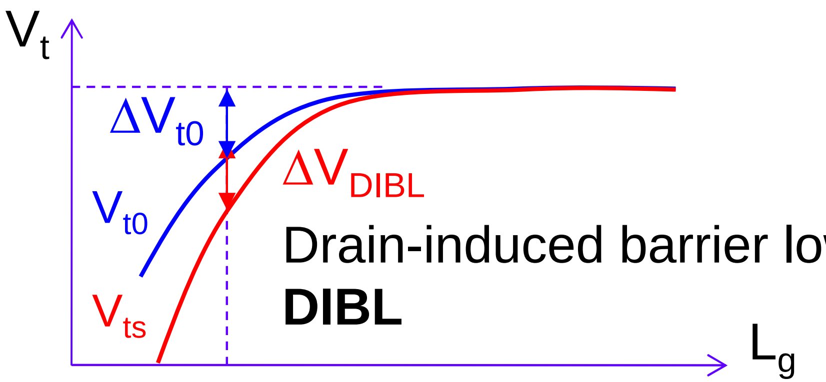
$$Q_B' = -qN_A W\left(L_{eff} X_d - X_d^2\right)$$

$$Q_B = -qN_A W\left(L_{eff} X_d\right)$$

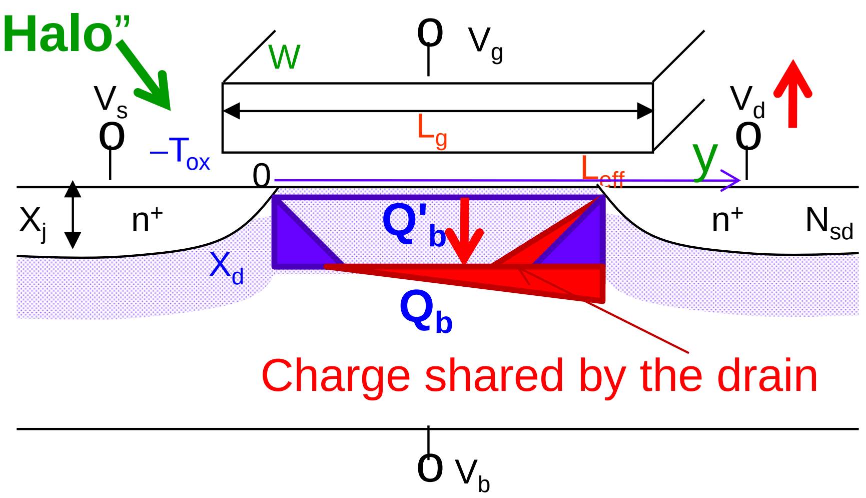
Bulk charge per unit area:

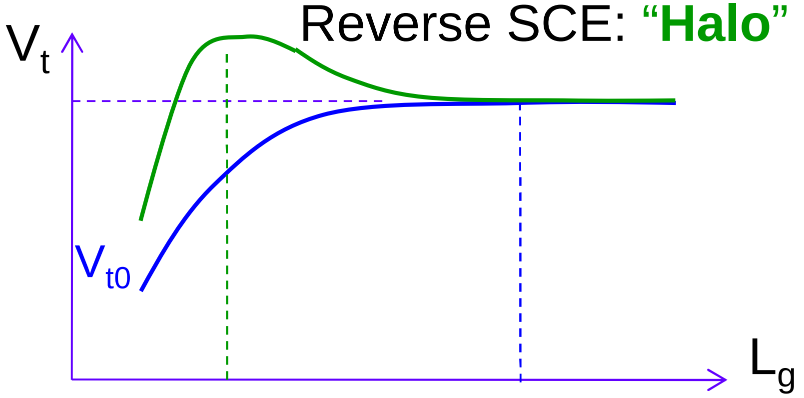
$$\therefore \frac{Q_b'}{Q_b} = \frac{Q_B'}{Q_B} = 1 - \frac{X_d}{L_{eff}}$$

DIBL and Reverse SCE: V_t "Roll-Up"

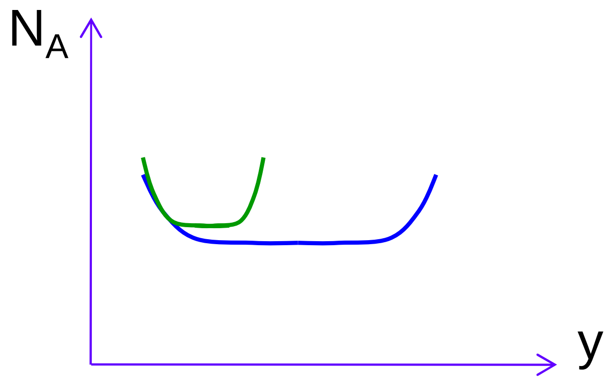


Drain-induced barrier lowering: **DIBL**





Reverse SCE: **"Halo"**



Summary of Important Equations

Threshold voltage

Long-channel (1D theoretical model)

$$V_t \equiv V_{gs}\Big|_{\psi_s = 2\phi_F + V_{sb}} = V_{FB} + \Upsilon\sqrt{2\phi_F + V_{sb}} + 2\phi_F$$

$$\phi_F = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right) \qquad \Upsilon = \sqrt{2q\varepsilon_{Si}N_A}/C_{ox} \qquad C_{ox} = \varepsilon_{ox}/T_{ox}$$

$$V_{FB} \equiv \phi_{MS} - Q_{ox}/C_{ox} = \Phi_M - \left(\chi + E_g/2 + \phi_F\right) - Q_{ox}/C_{ox}$$

Short-channel (triangle charge-sharing model)

$$V_{t0}\left(L_g\right) \equiv V_{t0\_long} - \Delta V_{t0} = V_{t0\_long} - \frac{4\varepsilon_{Si}\phi_F}{\varepsilon_{ox}}\frac{T_{ox}}{L_g - 2\sigma X_j}$$

Short-channel DIBL

$$\Delta V_{DIBL}\left(L_g\right) \equiv V_{t0}\left(L_g\right) - V_{ts}\left(L_g\right)$$

Drain current

Linear

$$I_{ds} = \mu_{eff}C_{ox}\frac{W}{L}\left(V_{gs} - V_t - \frac{1}{2}A_bV_{ds}\right)V_{ds}$$

Subthreshold

$$I_{ds} = \mu_n C_d v_{th}^2 \frac{W}{L} e^{\left(V_{gs}-V_t\right)/\left(nv_{th}\right)}\left(1 - e^{-V_{ds}/v_{th}}\right)$$

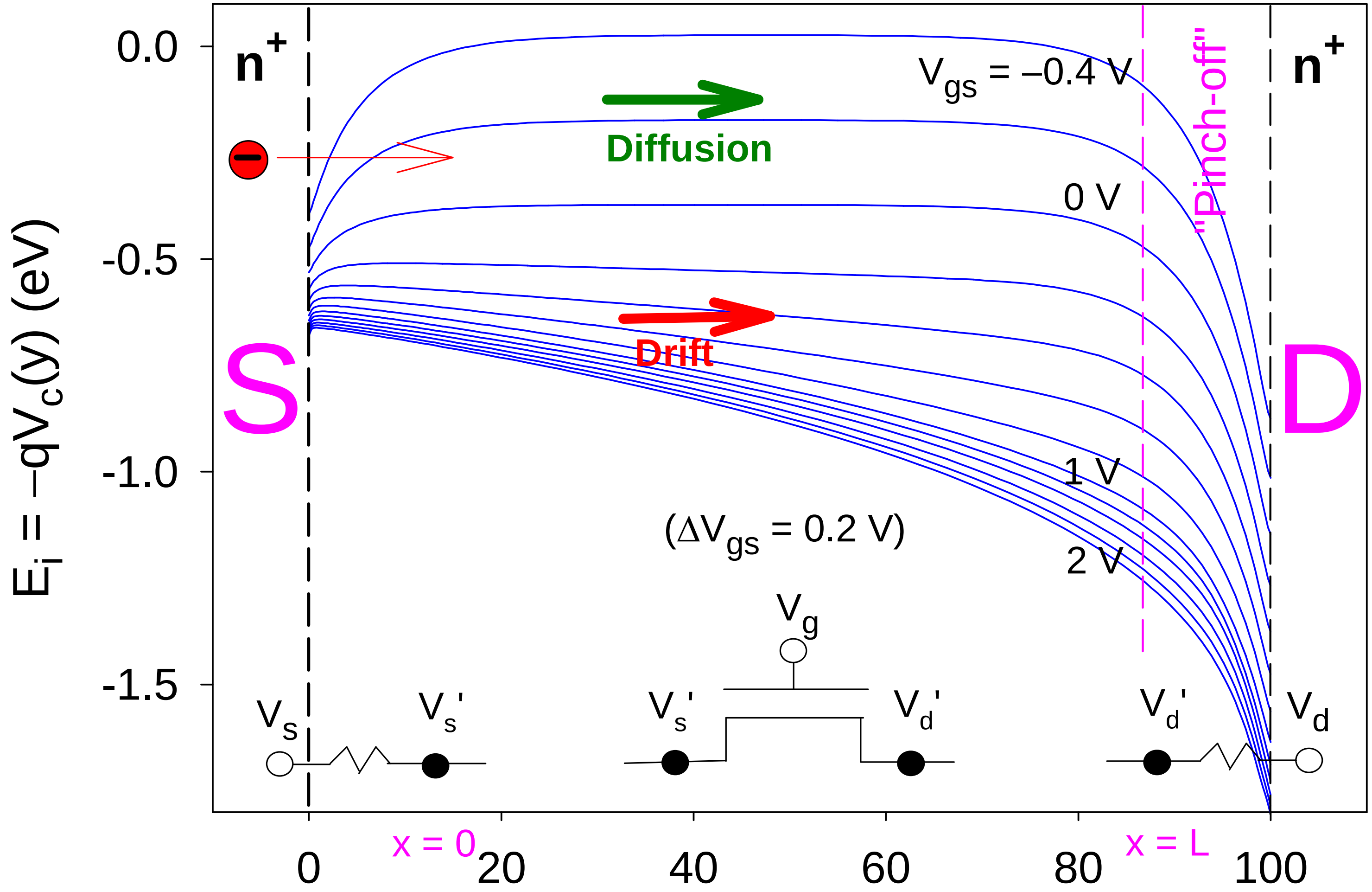
$$n = 1 + C_d/C_{ox}$$

$$C_d = \varepsilon_{Si}/X_{dm} = \frac{\Upsilon C_{ox}}{2\sqrt{2\phi_F + V_{sb}}}$$

Saturation

$$I_{dsat} = Wv_{sat}C_{ox}\frac{\left(V_{gs} - V_t\right)^2}{V_{gs} - V_t + A_bE_{sat}L_{eff}} \Rightarrow \propto \begin{cases} \left(V_{gs} - V_t\right)^2 & \left(L_{eff} \to \infty;\ \text{long-channel: quadratic}\right) \\ \left(V_{gs} - V_t\right) & \left(L_{eff} \to 0;\ \text{short-channel: linear}\right) \end{cases}$$

Gate-Controlled Drift (“ON”) and Diffusion (“OFF”)



Threshold Voltage Definition (I_{crit}@V_{t0})