- Four structural masks: Field, Gate, Contact, Metal.
- Reverse doping polarities for pMOSFET in N-well.
N-MOSFET Schematic

- Source terminal: Ground potential.
- Gate voltage: $V_{gs}$
- Drain voltage: $V_{ds}$
- Substrate bias voltage: $-V_{bs}$

$\psi(x,y)$: Band bending at any point $(x,y)$.  
$V(y)$: Quasi-Fermi potential along the channel.  
Boundary conditions: $V(y=0) = 0$, $V(y=L) = V_{ds}$. 

Do you remember what is quasi-Fermi level?
Long Channel Behavior

- The electric field in the channel is essentially one-dimensional (normal to the semiconductor surface)

- Mathematically: $E_x \gg E_y$
Drain Current Model

Electron concentration:

\[ n(x) = \frac{n_i^2}{N_a} e^{q(\psi - V) / kT} \]

Electric field:

\[ E^2(x, y) = \left( \frac{d\psi}{dx} \right)^2 = \frac{2kTN_a}{\varepsilon_{si}} \left[ \left( e^{-q\psi / kT} + \frac{q\psi}{kT} - 1 \right) + \frac{n_i^2}{N_a^2} \left( e^{-qV / kT} \left( e^{q\psi / kT} - 1 \right) - \frac{q\psi}{kT} \right) \right] \]

Condition for surface inversion:

\[ \psi(0, y) = V(y) + 2\psi_B \]

Maximum depletion layer width at inversion:

\[ W_{dm}(y) = \sqrt{\frac{2\varepsilon_{si} [V(y) + 2\psi_B]}{qN_a}} \]

\[ V(y) \] plays the role of the reverse bias in a MOS capacitor under non-equilibrium.
Current Density Equation

\[ J_n = q \mu_n nE + qD_n \frac{dn}{dy} \]

\[ n = \frac{n_i^2}{N_A} e^{\frac{q(\psi - V)}{kT}} \]

\[ \frac{dn}{dy} = \frac{q}{kT} n \left( \frac{d\psi}{dy} - \frac{dV}{dy} \right) \]

Current density equation (both drift and diffusion):

\[ J_n(x, y) = -q \mu_n n(x, y) \frac{dV(y)}{dy} \]

Quasi-Fermi level
Gradual Channel Approximation

Assumes that vertical field \( E_x \) is stronger than lateral field \( E_y \) in the channel region, thus 2-D Poisson’s equation can be solved in terms of 1-D vertical slices.

Current density equation (both drift and diffusion):

\[
J_n(x, y) = -q\mu_n n(x, y) \frac{dV(y)}{dy}
\]

Integrate in \( x \)- and \( z \)-directions,

\[
I_{ds}(y) = -\mu_{\text{eff}} W \frac{dV}{dy} \int Q_i(y) = -\mu_{\text{eff}} W \frac{dV}{dy} Q_i(V)
\]

where \( Q_i(y) = -q \int_0^{x_i} n(x, y) dx \) is the inversion charge per unit area.

Current continuity requires \( I_{ds} \) independent of \( y \), integration with respect to \( y \) from 0 to \( L \) yields

\[
I_{ds} = \mu_{\text{eff}} \frac{W}{L} \int_{V_{ds}}^{0} (-Q_i(V)) dV
\]
Pao-Sah's Double Integral


Change variable from \((x,y)\) to \((\psi,V)\),

\[ n(x, y) = n(\psi, V) = \frac{n_i^2}{N_a} e^{q(\psi - V)/kT} \]

\[ Q_i(V) = -q \int_{\psi_s}^{\psi_B} n(\psi, V) \frac{dx}{d\psi} d\psi = -q \int_{\psi_s}^{\psi_B} \frac{(n_i^2 / N_a) e^{q(\psi - V)/kT}}{E(\psi, V)} d\psi \]

Substituting into the current expression,

\[ I_{ds} = q\mu_{eff} \frac{W}{L} \int_0^{V_{ds}} \left[ \int_{\psi_s}^{\psi_B} \frac{(n_i^2 / N_a) e^{q(\psi - V)/kT}}{E(\psi, V)} d\psi \right] dV \]

where \(\psi_s(V)\) is solved by the gate voltage eq. for a vertical slice of the MOSFET:

\[ V_g = V_{fb} + \psi_s - \frac{Q_s}{C_{ox}} = V_{fb} + \psi_s + \sqrt{2\varepsilon_{si} kTN_a \left[ \frac{q\psi_s}{kT} + \frac{n_i^2}{N_a} e^{q(\psi_s - V)/kT} \right]} \]
Example of $\psi_s$ vs $V_G$ Relationship

$N_A = 8 \times 10^{16} \text{ cm}^{-3}$
$t_{ox} = 15 \text{ nm}$

Charge Sheet Approximation

Assumes that all the inversion charges are located at the silicon surface like a sheet of charge and that there is no potential drop across the inversion layer.

After the onset of inversion, the surface potential is pinned at \( \psi_s = 2\psi_B + V(y) \).

- **Depletion charge:** \( Q_d = -qN_aW_{d\text{max}} = -\sqrt{2\varepsilon_s q N_a (2\psi_B + V)} \)
- **Total charge:** \( Q_s = -C_{ox}(V_g - V_{fb} - \psi_s) = -C_{ox}(V_g - V_{fb} - 2\psi_B - V) \)
- **Inv. charge:** \( Q_i = Q_s - Q_d = -C_{ox}(V_g - V_{fb} - 2\psi_B - V) + \sqrt{2\varepsilon_s q N_a (2\psi_B + V)} \)

Substituting in \( I_{ds} = \mu_{eff} \frac{W}{L} \int_0^{V_{ds}} (-Q_i(V))dV \) and integrate:

\[
I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} \left\{ \left(V_g - V_{fb} - 2\psi_B - \frac{V_{ds}}{2}\right)V_{ds} - \frac{2\sqrt{2\varepsilon_s q N_a}}{3C_{ox}} \left[ (2\psi_B + V_{ds})^{3/2} - (2\psi_B)^{3/2} \right] \right\}
\]

Linear Region I-V Characteristics

For $V_{ds} \ll V_g$, 

$$I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} \left( V_g - V_{fb} - 2\psi_B - \frac{\sqrt{4\varepsilon_{si} q N_a \psi_B}}{C_{ox}} \right) V_{ds} = \mu_{eff} C_{ox} \frac{W}{L} (V_g - V_t) V_{ds}$$

where $V_t = V_{fb} + 2\psi_B + \frac{\sqrt{4\varepsilon_{si} q N_a \psi_B}}{C_{ox}}$ is the MOSFET threshold voltage.

Figure 6: Sub-threshold curves for 35nm L-gate devices

Experimental Determination of the Threshold Voltage

- Linear extrapolation (LE) at the maximum $G_m$ point
- Constant current (CC) method
- Transconductance change (TC) method

Example $I_d$ (and $G_m$) vs $V_{GS}$ curves

Example $\log(I_d) \text{ vs } V_{GS}$ curves

- $V_{SB} = 0.0 \text{ V}$
- $V_{SB} = 0.9 \text{ V}$
- $V_{SB} = 1.8 \text{ V}$

$W/L = 10/0.26 \text{ \mu m}$
$V_{DS} = 50 \text{ mV}$

Threshold Voltage Extraction Method Illustration


Fig. 9. Schematic illustration of determining threshold voltage by the transconductance change (TC) method. $V_{TH} = 5$ mV. The gate to source voltage at which the derivative of the transconductance is a maximum locates the threshold voltage determined by the TC method ($V_{TH}^{(TC)}$). The threshold voltage of the classical "2$\phi_F$ point" definition is labeled $V_{TH}^{(2\phi_F)}$. The linearly extrapolated threshold voltage is labeled $V_{TH}^{(LE)}$. The extrapolation line is drawn through the point of maximum slope (transconductance) of the $I_D$ curve with slope equal to the peak transconductance. The "knee" point of the band-bending $\phi$, $I_D$ corresponds to the point where the derivative of the transconductance goes through a maximum. This figure also shows the gate-channel capacitance $C$, the drain current $I_D$, and the linear extrapolation line $C_{GC}$ for reference.
Example $\frac{\partial I_d}{\partial V_{GS}}$ and $\frac{\partial^2 I_d}{\partial V_{GS}^2}$ vs $V_{GS}$ curves

Saturation Region I-V Characteristics

\[ I_{ds} = \mu_{\text{eff}} C_{ox} \frac{W}{L} \left\{ V_g - V_{fb} - 2\psi_B - \frac{V_{ds}}{2} \right\} V_{ds} - \frac{2\sqrt{2\varepsilon_{si} q N_a}}{3C_{ox}} \left[ (2\psi_B + V_{ds})^{3/2} - (2\psi_B)^{3/2} \right] \]

Keeping the 2nd order terms in \( V_{ds} \):

\[ I_{ds} = \mu_{\text{eff}} C_{ox} \frac{W}{L} \left[ (V_g - V_t) V_{ds} - \frac{m}{2} V_{ds}^2 \right] \]

where

\[ m = 1 + \frac{\varepsilon_{si} q N_a}{4\psi_B} \frac{1}{C_{ox}} = 1 + \frac{C_{dm}}{C_{ox}} = 1 + \frac{3t_{ox}}{W_{dm}} \]

is the body-effect coefficient

\[ I_{ds} = I_{dsat} = \mu_{\text{eff}} C_{ox} \frac{W}{L} \frac{(V_g - V_t)^2}{2m} \]

when

\[ V_{ds} = V_{dsat} = \frac{(V_g - V_t)}{m} \]

Typically, \( m \approx 1.2 \)
Pinch-off Condition

From inversion charge density point of view,

\[ Q_i(V) = -C_{ox}(V_g - V_t - mV) \]

while

\[ I_{ds} = \mu_{eff} \frac{W}{L} \int_{0}^{V_{ds}} (-Q_i(V))dV \]

At \( V_{ds} = V_{dsat} = (V_g - V_t)/m \), \( Q_i = 0 \) and \( I_{ds} = \text{max.} \)
Beyond Pinch-off

Channel length modulation
Saturation Characteristics – Experimental Example

- PMOS
- NMOS

65 nm technology

VGS = 1.2V

VGS = 1.1V

VGS = -1.2V

VGS = -1.1V

VG steps of 0.1V

Slope due to:
- Channel length modulation (CLM)
- Drain induced barrier lowering (DIBL) – to be discussed later

Subthreshold Region

\[ J_n = q\mu_n nE + qD_n \frac{dn}{dy} \]
Subthreshold Currents

\[-Q_s = \varepsilon_{si} E_s = \sqrt{2 \varepsilon_{si} kT N_a} \left[ \frac{q \psi_s}{kT} + \frac{n_i}{N} e^{q(\psi_s - V)/kT} \right]^{1/2} \]

Power series expansion: 1st term \(Q_d\), 2nd term \(Q_i\),

\[-Q_i = \frac{\varepsilon_{si} q N_a}{2 \psi_s} \left( \frac{kT}{q} \right) \left( \frac{n_i}{N} \right)^2 e^{q(\psi_s - V)/kT} \]

\[I_{ds} = \mu_{eff} \frac{W}{L} \int_0^{V_{ds}} (-Q_i(V)) dV \]

\[\Rightarrow I_{ds} = \mu_{eff} \frac{W}{L} \sqrt{\frac{\varepsilon_{si} q N_a}{2 \psi_s} \left( \frac{kT}{q} \right) \left( \frac{n_i}{N} \right)^2 e^{q(\psi_s - V)/kT} \left( 1 - e^{-qV_{ds}/kT} \right)} \]

Solving for \(\psi_s\) using

\[V_g = V_{fb} + \psi_s + \frac{\sqrt{2 \varepsilon_{si} kT N_a}}{C_{ox}} \left[ \frac{q \psi_s}{kT} \right]^{1/2} \]

or,

\[I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} (m - 1) \left( \frac{kT}{q} \right)^2 e^{q(V_g - V_i)/mkT} \left( 1 - e^{-qV_{ds}/kT} \right) \]

Inverse subthreshold slope:

\[S = \left( \frac{d(\log I_{ds})}{dV_g} \right)^{-1} = 2.3 \frac{mkT}{q} = 2.3 \frac{kT}{q} \left( 1 + \frac{C_{dm}}{C_{ox}} \right) \ln(10) \]

\[\left( \ln(10) kT / q \right)^{-1} \approx 60 \text{ mV / dec} \]

at 300K
Subthreshold Slope ($S$)

\[ S = \left( \frac{d \log I_D}{dV_G} \right)^{-1} = \frac{\partial V_G}{\partial \Psi_S} \frac{\partial \Psi_S}{\partial \log I_D} = \left( 1 + \frac{C_{dm}}{C_{ox}} \right) \frac{kT}{q} \ln(10) \]

1 decade

Gate to channel potential coupling: >1 in MOSFETs

60 mV/dec in MOSFETs due to Fermi-Dirac distribution
Questions?
Body Effect: Dependence of Threshold Voltage on Substrate Bias

You can either

1. Start with the Poisson’s equation solution for $Q_s$, $Q_d$, $Q_i$ with the quasi-Fermi levels of the holes and electrons separated by the substrate bias $V_{bs}$, or

2. Keep the substrate at zero (as the reference) and shift the source, drain, and gate biases by $V_{bs}$
Body Effect: Dependence of Threshold Voltage on Substrate Bias

You can either
1. Start with the Poisson’s equation solution for $Q_s$, $Q_d$, $Q_i$ with the quasi-Fermi levels of the holes and electrons separated by the substrate bias $V_{bs}$, or
2. Keep the substrate at zero (as the reference) and shift the source, drain, and gate biases by $V_{bs}$.

$$I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} \left( V_g - V_{fb} - 2\psi_B - \frac{V_{ds}}{2} \right) V_{ds} - \frac{2\sqrt{2\varepsilon_s q N_a}}{3C_{ox}} \left( \sqrt{2\varepsilon_s q N_a} (2\psi_B + V_{bs}) \right) \left[ (2\psi_B + V_{bs} + V_{ds})^{3/2} - (2\psi_B + V_{bs})^{3/2} \right]$$

$$V_t = V_{fb} + 2\psi_B + \frac{\sqrt{2\varepsilon_s q N_a (2\psi_B + V_{bs})}}{C_{ox}}$$

$$dV_t \over dV_{bs} = \frac{\sqrt{\varepsilon_s q N_a / 2(2\psi_B + V_{bs})}}{C_{ox}}$$

$$m = 1 + \frac{\sqrt{\varepsilon_s q N_a / 4\psi_B}}{C_{ox}} = 1 + \frac{C_{dm}}{C_{ox}} = 1 + \frac{3\varepsilon_{ox}}{W_{dm}}$$

$t_{ox}=200\ \text{Å} \quad \text{Applied bias} = -V_{bs}$

$V_{fb}=0$

$N_a=10^{16}\ \text{cm}^{-3}$

$N_a=3\times10^{15}\ \text{cm}^{-3}$
Efficiency of Body Biasing in 90-nm CMOS for Low-Power Digital Circuits

Klaus von Arnim, Eduardo Borinski, Peter Seegebrecht, Member, IEEE, Horst Fiedler, Member, IEEE, Ralf Brederlow, Roland Thewes, Member, IEEE, Jörg Berthold, and Christian Pacha, Member, IEEE

Abstract — The efficiency of body biasing for leakage reduction and performance improvement in a 90-nm CMOS low-power technology with triple-well option is evaluated. Static measurements of single devices and dynamic measurements of ring oscillators and 32-b parallel prefix adders are presented. Whereas forward biasing still provides a significant performance improvement of up to 37% for low-leakage devices with 2.2-nm gate oxide thickness, the application of reverse biasing to reduce subthreshold leakage currents is inefficient due to additional leakage currents such as gate leakage and gate-induced drain leakage. Experimental results confirm that, in 90-nm CMOS circuits, the efficiency of body biasing strongly depends on the device type and operating temperature. Moreover, the impact of the zero-temperature coefficient point on static device and dynamic circuit performance is investigated.

Index Terms—Body biasing, CMOS digital integrated circuits, zero-temperature coefficient point.

I. INTRODUCTION

Fig. 1. Classifications of different circuit applications in the 90-nm CMOS system-on-chip technology.
Application of Body Bias for Controlling Variations

Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-Die Parameter Variations on Microprocessor Frequency and Leakage

James W. Tschanz, Member, IEEE, James T. Kao, Member, IEEE, Siva G. Narendra, Member, IEEE, Raj Nair, Member, IEEE, Dimitri A. Antoniadis, Fellow, IEEE, Anantha P. Chandrakasan, Senior Member, IEEE, and Vivek De, Member, IEEE

Abstract—Bidirectional adaptive body bias (ABB) is used to compensate for die-to-die parameter variations by applying an optimum pMOS and nMOS body bias voltage to each die which maximizes the die frequency subject to a power constraint. Measurements on a 150-nm CMOS testchip which incorporates on-chip ABB, show that ABB reduces variation in die frequency by a factor of seven, while improving the die acceptance rate. An enhancement of this technique, that compensates for within-die parameter variations as well, increases the number of dies accepted in the highest frequency bin. ABB is therefore shown to provide bin split improvement in the presence of increasing process parameter variations.

Index Terms—Body bias, CMOS digital integrated circuits, forward bias, low-power circuits, microprocessors, parameter variations, substrate bias, within die variation.

Fig. 1. Measured leakage power and frequency for 62 dies.
Dependence of Threshold Voltage on Temperature

For n\(^+\) poly gated nMOSFET, \(V_{fb} = -(E_g/2q) - \psi_B\)

\[
\Rightarrow V_t = -\frac{E_g}{2q} + \psi_B + \frac{\sqrt{4\varepsilon_{si}qN_a\psi_B}}{C_{ox}}
\]

\[
\frac{dV_t}{dT} = -\frac{1}{2q} \frac{dE_g}{dT} + \left(1 + \frac{\sqrt{\varepsilon_{si}qN_a/\psi_B}}{C_{ox}}\right) \frac{d\psi_B}{dT} = -\frac{1}{2q} \frac{dE_g}{dT} + (2m-1)\frac{d\psi_B}{dT}
\]

\[
\Rightarrow \frac{dV_t}{dT} = -(2m-1)\frac{k}{q} \ln\left(\frac{\sqrt{N_cN_v}}{N_a}\right) + \frac{3}{2} + \frac{m-1}{q} \frac{dE_g}{dT}
\]

From Table 2.1, \(dE_g/dT \approx -2.7 \times 10^{-4}\) eV/K and \((N_cN_v)^{1/2} \approx 2.4 \times 10^{19}\) cm\(^{-3}\).

For \(N_a \sim 10^{16}\) cm\(^{-3}\) and \(m \sim 1.1\), \(dV_t/dT\) is typically \(-1\) mV/K.

Note: Operating temperature is specified at 85 °C for microprocessors and 150 °C for automotive applications.
Carrier Transport and Gate Capacitance

Linear Region:

\[ I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} (V_g - V_t)V_{ds} \]

Saturation Region:

\[ I_{ds} = I_{dsat} = \mu_{eff} C_{ox} \frac{W}{L} \frac{(V_g - V_t)^2}{2m} \]

Will come back to a more elaborate discussion later in the course about carrier transport.

Let’s first digress briefly about the gate capacitance \( C_{ox} \) and the effective mobility \( \mu_{eff} \) right now (we will return to them later again)
MOSFET Channel Mobility

\[
\mu_{\text{eff}} = \frac{\int_{0}^{x_i} \mu_n n(x) \, dx}{\int_{0}^{x_i} n(x) \, dx}
\]

It was empirically found that when \( \mu_{\text{eff}} \) is plotted against an effective normal field \( E_{\text{eff}} \), there exists a “universal relationship” independent of the substrate bias, doping concentration, and gate oxide thickness (Sabnis and Clemens, IEDM 1979).

Here

\[
E_{\text{eff}} = \frac{1}{\varepsilon_{si}} \left( |Q_d| + \frac{1}{2} |Q_i| \right)
\]

Since \(|Q_d| = \sqrt{4 \varepsilon_{si} q N_a \psi_B} = C_{ox} (V_t - V_{fb} - 2 \psi_B) \) and \(|Q_i| \approx C_{ox} (V_g - V_t)\),

\[
\Rightarrow E_{\text{eff}} = \frac{V_t - V_{fb} - 2\psi_B}{3t_{ox}} + \frac{V_g - V_t}{6t_{ox}}
\]

For n\(^+\) poly gated nMOSFET,

\[
E_{\text{eff}} = \frac{V_t + 0.2}{3t_{ox}} + \frac{V_g - V_t}{6t_{ox}}
\]
Electron Mobility

- Low field region (low electron density): Limited by impurity or Coulomb scattering (screened at high electron densities).

- Intermediate field region: Limited by phonon scattering,

\[ \mu_{\text{eff}} \approx 32500 \times E^{-1/3} \]

- High field region (> 1 MV/cm): Limited by surface roughness scattering (less temp. dependence).
Temperature Dependence of MOSFET Current

Note: Operating temperature is specified at 85 °C for microprocessors and 150 °C for automotive applications.
In general, pMOSFET mobility does not exhibit “universal” behavior as well as nMOSFET.
Intrinsic MOSFET Capacitance

- **Subthreshold region:** \( C_g = W L \left( \frac{1}{C_{ox}} + \frac{1}{C_d} \right)^{-1} \approx W L C_d \)

- **Linear region:** \( C_g = W L C_{ox} \)

- **Saturation region:**

  \[
  Q_i(y) = -C_{ox}(V_g - V_t) \sqrt{1 - \frac{y}{L}}
  \]

\[
\Rightarrow C_g = \frac{2}{3} W L C_{ox}
\]

See Taur & Ning p. 131 for derivation steps.