EE 240B – Spring 2019

Advanced Analog Integrated Circuits

Lecture 2: MOS Transistor Models for Analog Design

Ali Niknejad

Dept. of EECS

Outline

- MOSFET DC I/V Model
  - Square law model
  - Short Channel Effects
  - Weak and moderate inversion
- MOSFET CV Model
- Compact Models
- Next Lecture: MOSFET Small-Signal Model
  - Transconductance
  - Output Resistance
  - Capacitances

Why is modeling important?

- Analog circuits employ transistors in a continuous manner where the precise currents, voltages, and charges need to be correctly calculated. Digital circuits by contrast have a “margin” of error.
- Analog models are our window into the physical device and process. We like to do experiments with “SPICE” rather than pay for actual Si. This is too expensive, time consuming, and often difficult.
- MOS transistor models can be categorized as follows
  - A MOSFET is a switch
  - A MOSFET is a current source and a switch (square law/ regional)
  - Short channel “hand” model
  - A MOS transistor is described by the BSIM equations !

Why not Square Law?

- The square law model is well known and widely used but unfortunately grossly inappropriate for short channel transistors
- For one this model does not address the important transition region of operation, *moderate* inversion, between strong inversion and weak inversion.
- It's good to review the assumptions behind this simple model in order to identify potential problems.

Humble Origin of Square Law Model

- Assume all current flow in transistor is due to drift (as opposed to diffusion). This implicitly assumes that we are in strong inversion.

$$J_n = -Nqv_{drift} = -Nq\mu E = Nq\mu\frac{dV}{dx}$$

- Now assume uniform current flow at the surface (charge-sheet, quasi-static assumption) which implies that

$$W\int_0^L J_n dx = I_{ds}L$$

$$I_{ds} = \frac{W}{L}\int_0^L Nq\mu\frac{dV}{dx}dx = \frac{W}{L}\int_0^{V_{ds}} q\mu N dV$$

Gradual Channel Approximation

- Now we assumed (unstated) that in the MOS transistor the variation in the field is only in the *x*-direction. We have assumed that the vertical field does not affect the *flow* carriers in the channel. This is the gradual channel approximation.
- In practice we know that the concentration of carriers is due to the vertical field. At the source we have

$$N/q = C_{ox}(V_{gs} - V_T(x))$$

- So in a position *x* in the channel

$$N/q = C_{ox}(V_{gs} - V(x) - V_T(x))$$

Constant Mobility and Threshold

- To make life easy, let's assume the threshold voltage $V_T(x)=V_T$ is a constant.
- Also, assume that the mobility μ is a constant along the channel and independent of bias.

$$I_{ds} = \frac{W}{L}\int_0^{V_{ds}} q\mu N dV$$

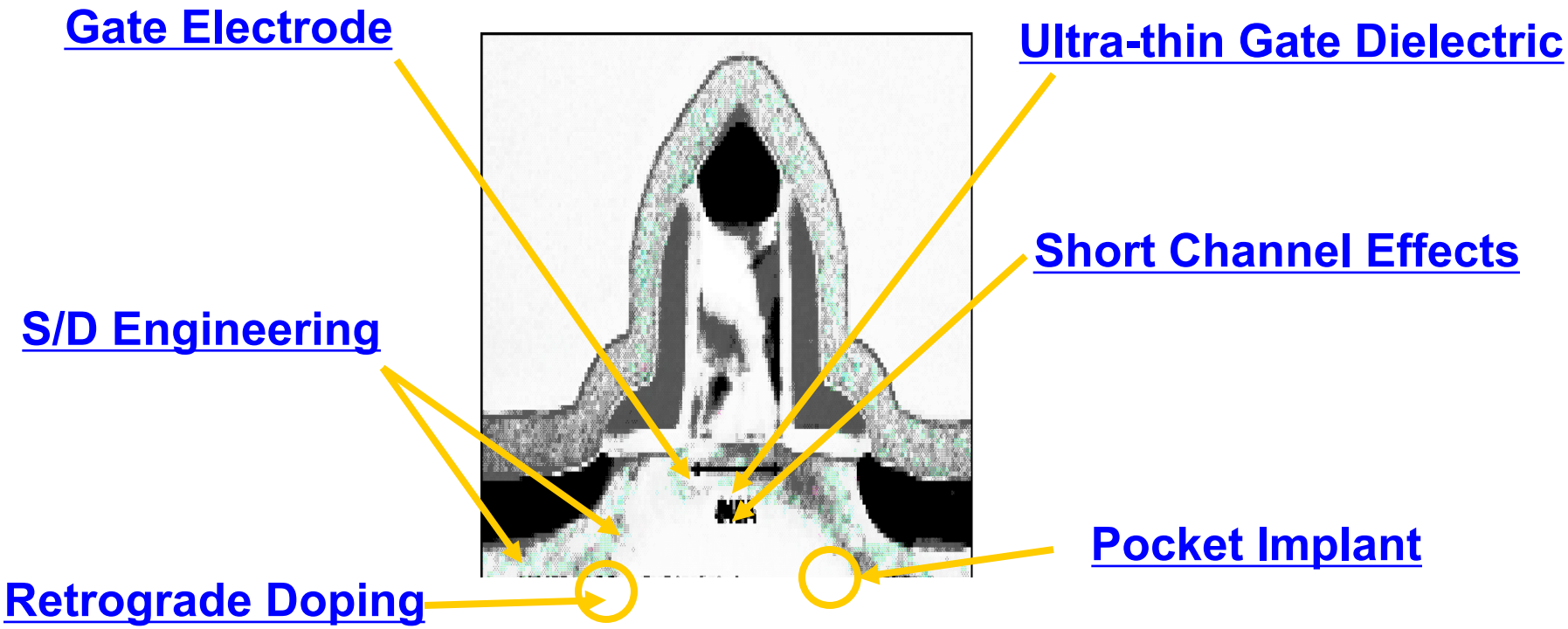
$$= \mu C_{ox}\frac{W}{L}\int_0^{V_{ds}} (V_{gs} - V_T - V)dV$$

$$I_{ds} = \mu C_{ox}\frac{W}{L}\left((V_{gs} - V_T)V_{ds} - V_{ds}^2/2\right)$$

Square Law Summary

- We have derived the simple square law model under the following assumptions:
  - MOSFET is like a non-linear resistor with a continuous channel from source to drain
  - Vertical field determines charge density
  - Lateral field determines drift current
  - Neglect diffusion currents
  - Neglect variation in threshold voltage along channel
  - Assume the mobility is a constant as a function of lateral and vertical fields
  - More assumptions that are too complicated to mention !

A Real Transistor !!!



- Does this look like the “textbook” long-channel transistor?

Doping / High Field Effects

- For deeply scaled devices, the dimensions are small enough such that ~1V exerts a considerable electric field (force) on the carriers. This results in velocity saturation and impact ionization. The surface effective mobility is also considerably lower and a function of the gate bias.
- The drain/channel region is a high field region where the usual 1D and quasi-2D approximations fail to predict the actual influence of the drain and body on the inversion layer.
- In order to provide adequate performance, the doping profile of a modern FET is complicated and leads to complicated geometry variations in the threshold voltage.
- The non-uniform doping also complicates the output resistance of the device.

Saturation?

- Where does saturation come from?
- Well, we know that as we increase the drain voltage, we will eventually "pinch-off" the channel, in other words the density of carriers will be driven to zero (depletion) near the drain end of the transistor. This happens when

$$C_{ox}(V_{gs} - V_T - V_{ds}) = 0$$

- Now the drain cannot "communicate" with the channel and we expect the MOSFET behavior to be independent of the drain voltage. The current therefore will increase and "saturate" at a value of $V_{ds} = V_{gs}$-V_T.

$$I_{ds} = \frac{1}{2}\mu C_{ox}\frac{W}{L}\left((V_{gs} - V_T)^2\right)$$

High Field Region

- So any “extra” drain voltage drop, beyond $V_{dsat} = V_{gs}-V_t$ must be dropped across this depletion region. The growth and of this region with drain voltage gives rise to channel length modulation (output impedance).
- Since the drain depletion region is small, for large values of V_{ds} the lateral electric field in this region can be quite large. Even though there are no mobile carriers in this region, the current flow in the MOSFET does not cease.
- The carriers in fact travel through this region at the velocity saturated speed $\underline{v}_{sat}$. Since $v_{sat} < 1$ there is a small finite density of free carriers in this region.

Bulk Charges

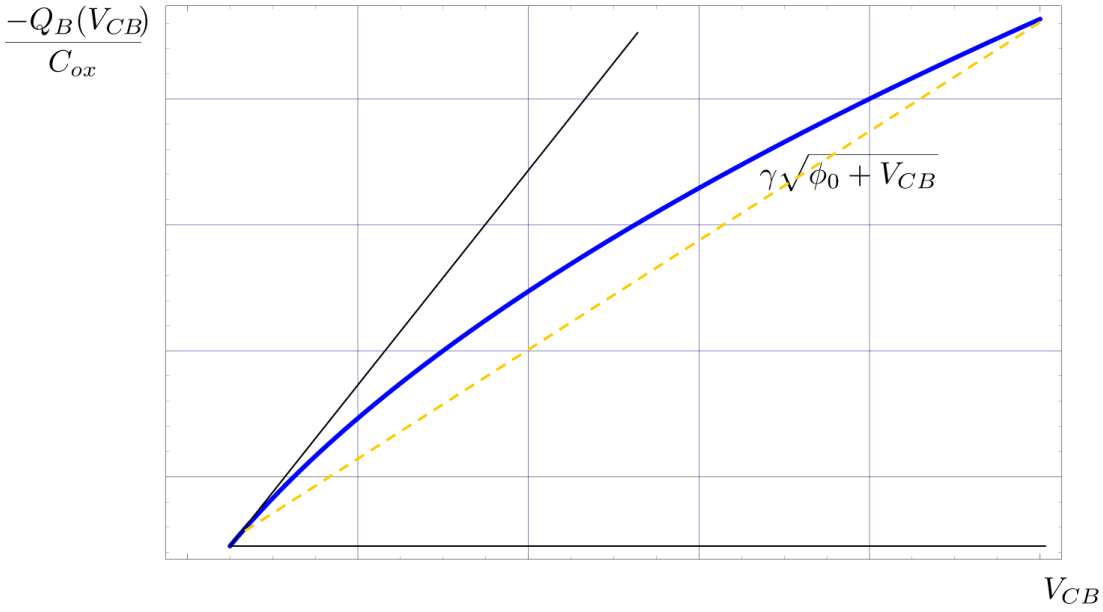
- In our simple derivation, we assumed that the body charge (immobile) is fixed by the value at the source. In reality the background charge increases as we move towards the drain due to the reverse bias. This results in overestimation of the inversion charge and hence current.
- In reality we have an good expression for the body charge

$$Q_B = -\gamma C_{ox}\sqrt{\phi_0 + V_{CB}}$$

- The square root is inconvenient, as it results in 3/2 powers under integration. We need a better approximation (maybe linear) to give a simple saturation current expression.

Approximation of Body Charge

- Previously we assumed that the constant line expression. A slightly more accurate expression is to assume a Taylor series expansion at the source.
- This gives us the more accurate equation. If we use this as a fitting parameter, we get the best model.



$$I_{DS} = \frac{W}{L}\mu C_{ox}\left((V_{GS} - V_T)V_{DS} - \frac{\alpha}{2}V_{DS}^2\right)$$

$$I_{DS,sat} = \frac{W}{L}\mu C_{ox}\frac{(V_{GS} - V_T)^2}{2\alpha}$$

Output Resistance

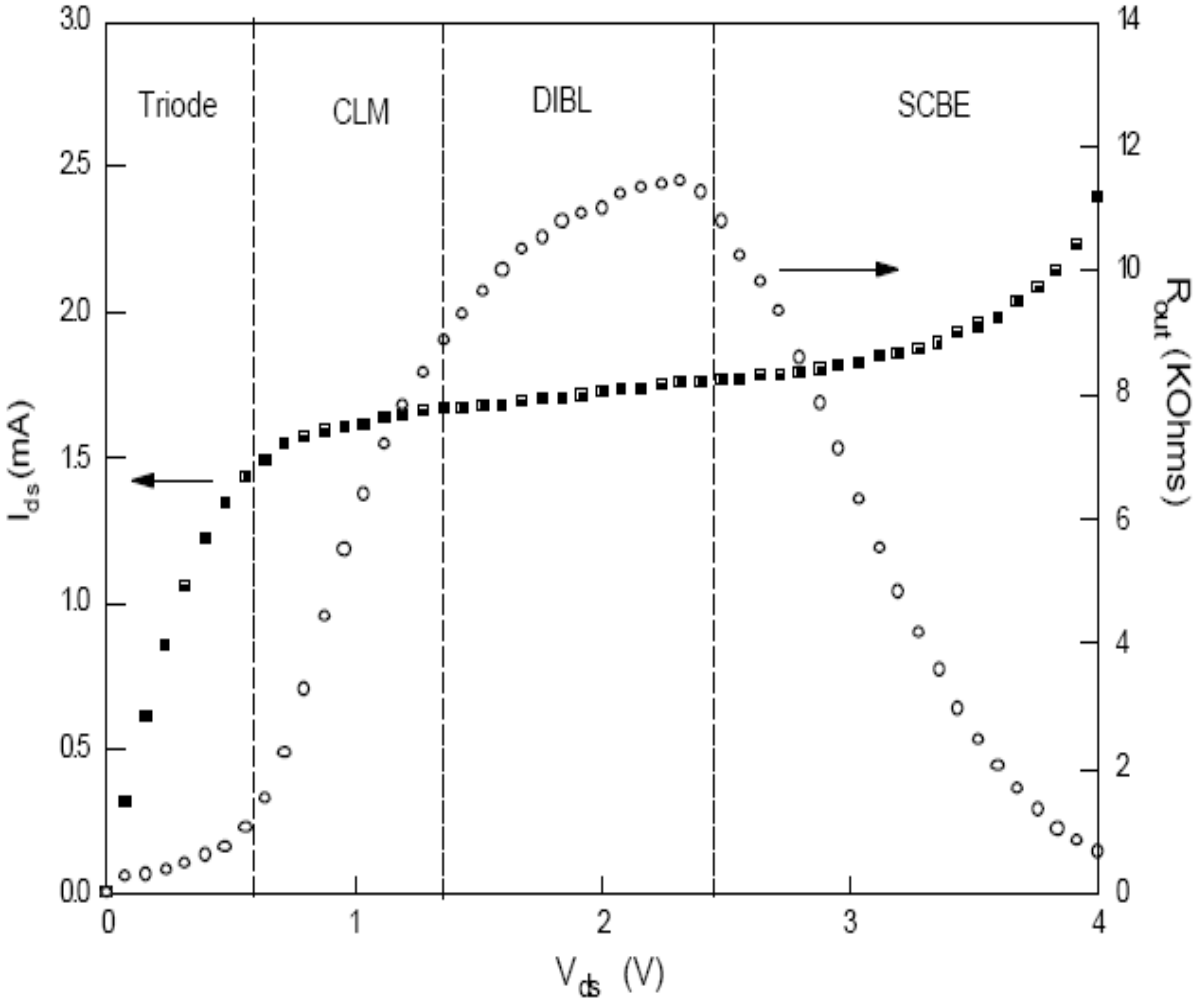
- As noted the variation in the depletion region width at the pinch-off point modulates the channel length

$$I \propto \frac{1}{L - \delta L(V_{ds})} \approx \frac{1}{L}\left(1 + \frac{\delta L(V_{ds})}{L}\right)$$

$$\frac{I_{ds}}{I_{ds0}} = (1 + \lambda V_{ds})$$

- If this effect is a small perturbation, we can assume that it responds linearly to the drain voltage, or

Output Resistance Mechanisms



Source: BSIM3v3 Manual

- In reality all effects active simultaneously
- CLM only for relatively low fields
- DIBL dominates for high fields
- Hot carrier impact ionization dominates for large Vds

CLM/DIBL

- The Channel Length Modulation (CLM) model is derived by assuming a pseudo two-dimensional model for the potential in the drain region
- Drain Induced Barrier Lowering (DIBL) accounts for the "drain" control of the channel. In an ideal transistor, only the front-gate controls the gate. In a real transistor, the channel potential depends additionally on the back-gate and the drain junction.
- Drain control is minimized by using long channel transistors, or by minimizing the drain junction depth.
- In practice, it's convenient to assume that the threshold voltage varies linearly with the drain voltage.

SCBE

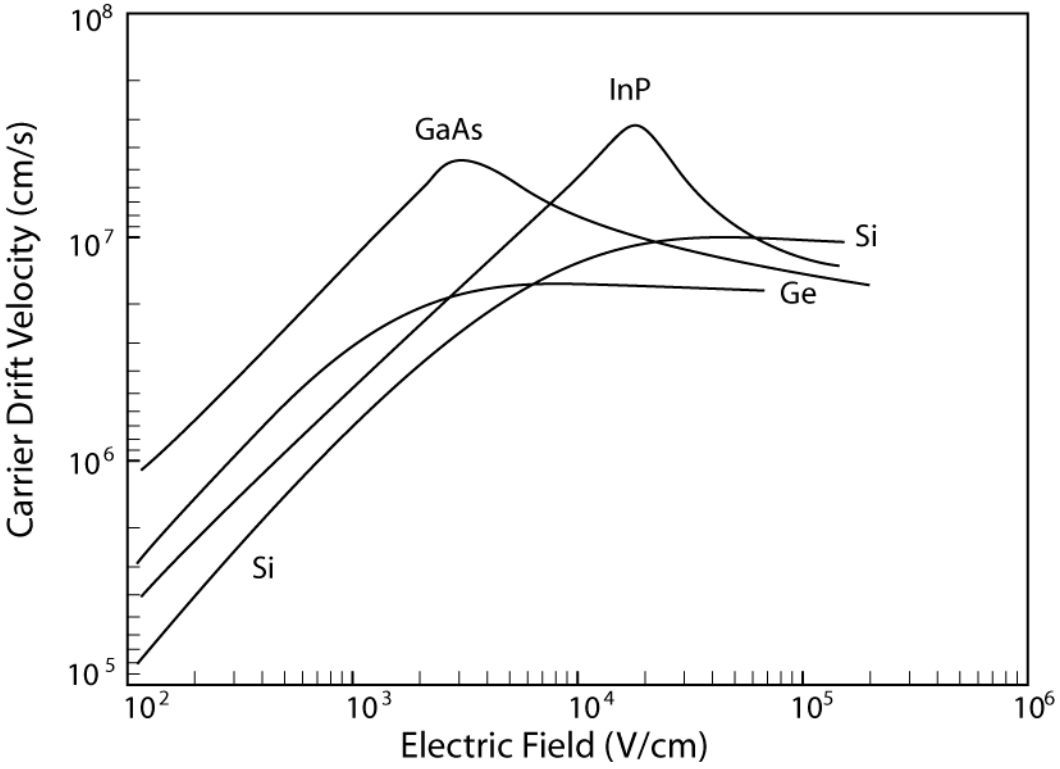
- Substrate Current Body Effect (SCBE) accounts for energetic electrons ("hot" electrons) created near the drain region due to the high electric fields (> 0.1 MV/cm).
- These hot electrons have enough energy so that when they collide with the lattice they knock off electrons from the Si atoms (impact ionization).
- This creates electron/hole pairs leading to a substrate current I_{sub} that flows into the substrate from the drain terminal.

$$I_{ds} = I_{ds0} + I_{sub}$$

$$I_{sub} = \frac{A_i}{B_i} I_{ds}(V_{ds} - V_{dsat}) exp\left(-\frac{B_i l}{V_{ds} - V_{dsta}}\right)$$

Velocity Saturation

- Initially drift velocity increases linearity with field
- For high fields, the velocity saturates
- For some materials there is a peak (not Si), but saturated velocity is best for Si
- For Si, this is modeled by the a fitting equation



Effective Mobility

- Mobility is not constant along the channel. An effective mobility can be used to correct for this

$$\mu_{eff} = \frac{1}{\frac{1}{L}\int_0^L \frac{dx}{\mu}}$$

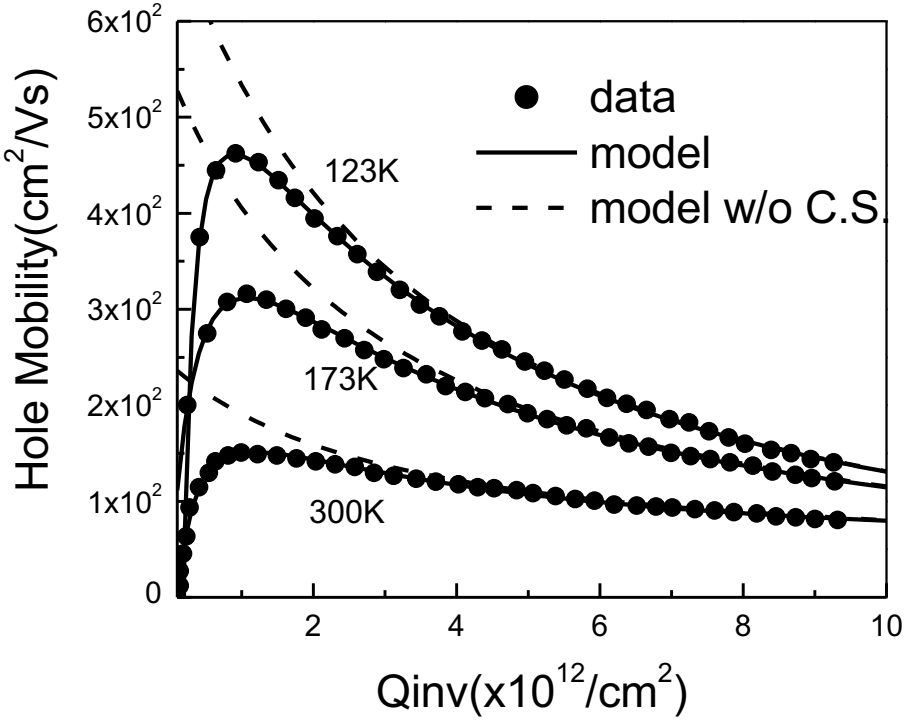
- The mobility varies as a function of the average vertical electric field

$$\mu = \frac{\mu_0}{1 + a_\theta E_{y,ave}}$$

- A strong electric field tends to push carriers close to the surface where enhanced scattering lowers mobility

$$\mu = \frac{\mu_0}{1 + \theta(V_{GS} - V_T) + \theta_B V_{SB}}$$

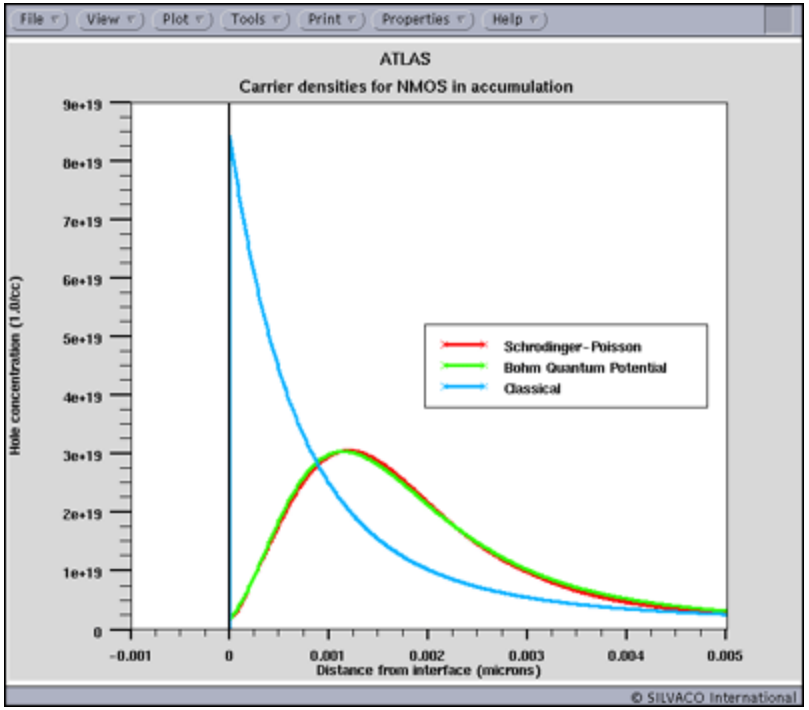
Low Field Mobility



- It is observed that at low fields the mobility drops. The explanation is that the inversion layer "shields" the carriers from the background dopants. Thus there is considerable Coulomb scattering for low fields.

Quantum Effects

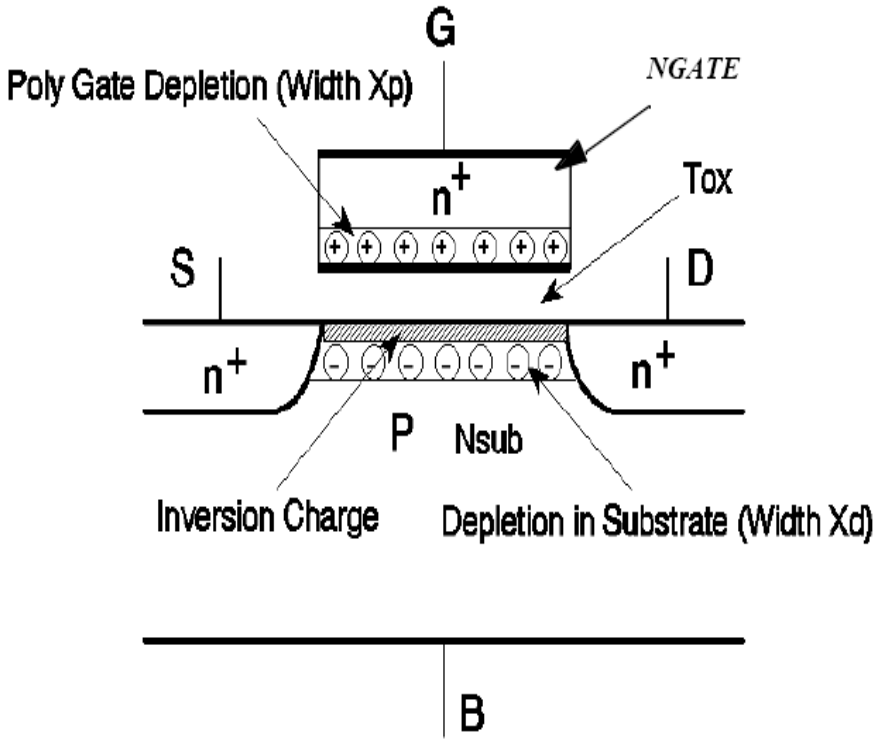
- The inversion charge profile is usually derived by solving Poisson's equation. This results in the peak of the charge at the surface.
- If Schrödinger's equation is solved simultaneously, we find the charge density to peak away from the surface
- The position of the peak varies with applied gate bias.



Source:
http://www.silvaco.com/products/vwf/atlas/quantum3d/quantum_br.html

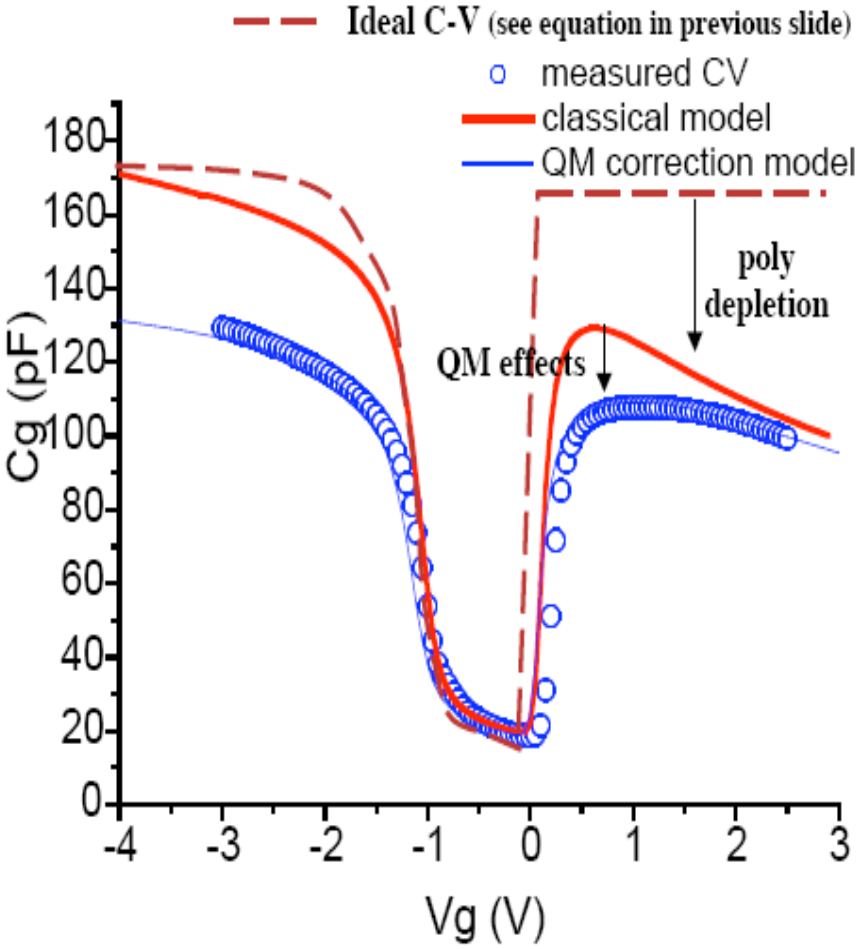
Quantum Polysilicon Depletion

- Since the gate is not a perfect conductor, but also a semiconductor with doping N_{GATE}, we observe a bias dependent depletion region X_p. This is in direct analogy with the surface charge position.
- The effective oxide thickness is thus larger than T_{ox}
- This reduces our gate drive even further.



Quantum Effects in CV

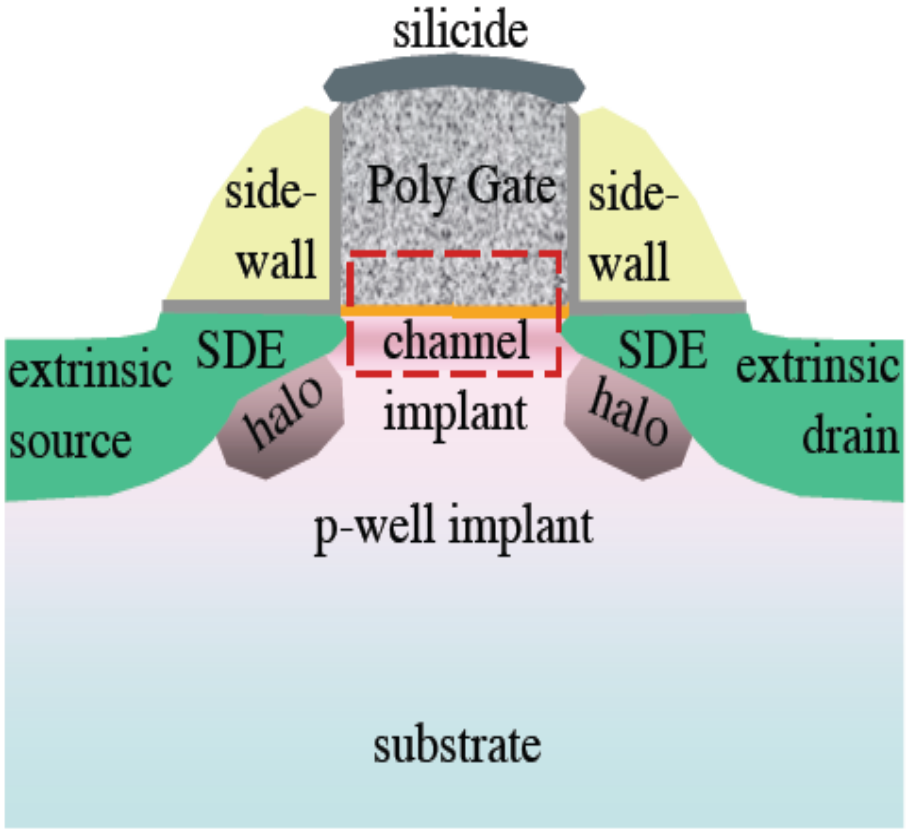
- The actual CV curve seen on the right shows drastically lower oxide capacitance due to the quantum confinement.
- The capacitance varies with bias in the real device.
- For large T_{ox}, these effects are negligible, but for deeply scaled technology it is quite noticeable.



Source: R. Dutton and C.-H. Choi

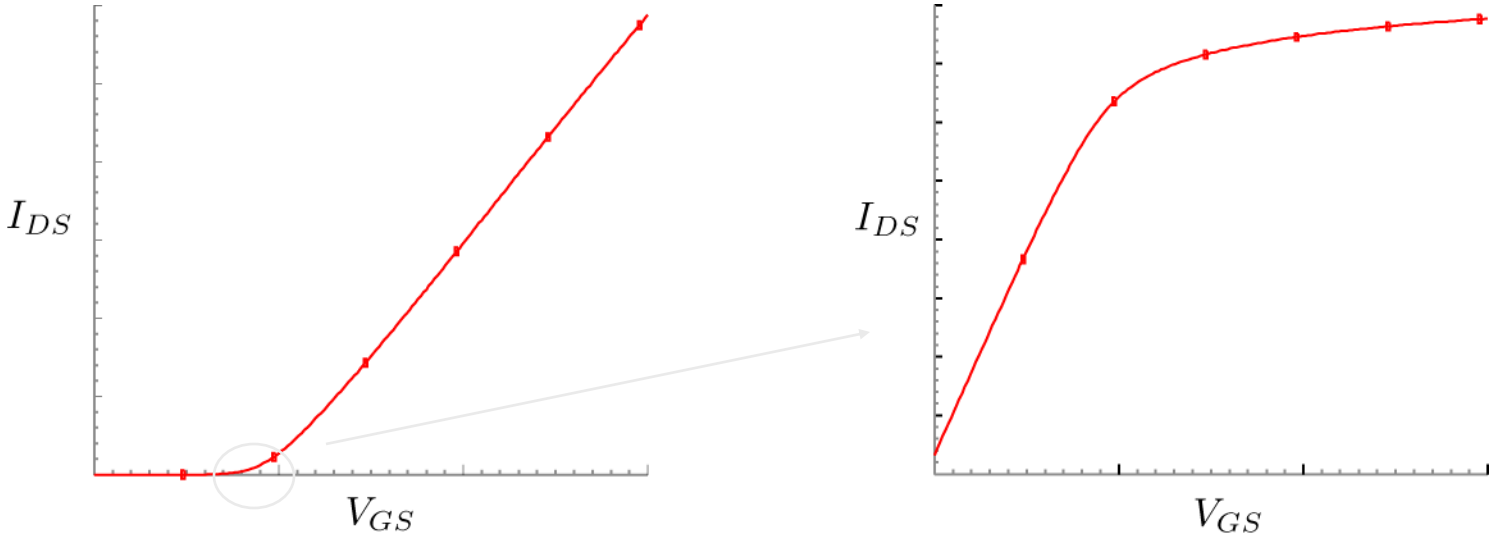
Non-Uniform Doping

- The doping concentration in a modern FET has two important variations.
- The doping is non-uniform in the vertical direction due to “channel” threshold implant.
- The doping is non-uniform in the lateral direction, with higher concentration near the source drain (SDE). In addition, there is often a halo implant.



Source: R. Dutton and C.-H. Choi

Sub-Threshold Region



- The square law model assumes that the current is zero until the threshold and then magically it starts increasing. Of course this is a fantasy. In reality the current flow is observed to increase exponentially for voltages near the threshold voltage. This is easy to explain if we view the MOSFET as a lateral BJT.

“BJT” Weak Inversion Model

- Assume that all the current is due to *diffusion* rather than drift. In other words, the potential along the channel varies negligibly but the carrier density varies linearly.
- Since the source-channel junction is reverse biased, it acts like a diode. By Boltzmann statistics we know that only a small fraction (the “tail” of the distribution) of carriers have sufficient energy to be injected from the source to the channel.
- Once in the channel, carriers will recombine or diffuse either into the drain (or into the substrate).
- If we increase the gate voltage, the channel potential follows almost linearly. This is only true in weak inversion because once we hit strong inversion, the channel potential is “pinned”.
- As the barrier to injection is lowered, an exponential increase in current flow is observed due to the Boltzmann distribution.

Channel Potential in Weak Inversion

- There's no explicit base terminal but the potential of the "base" is controlled indirectly through the capacitive divider formed by C_{ox} and C_{dep}.

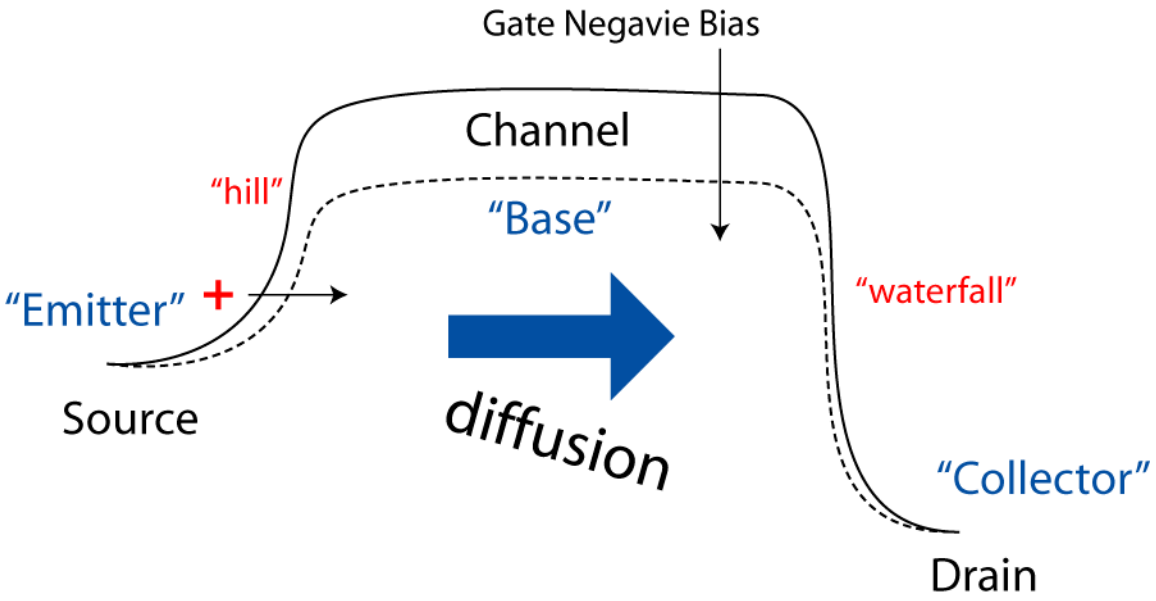
$$\delta V_{ch} \approx \frac{C_{ox}}{C_{dep} + C_{ox}} \delta V_g = \frac{\delta V_g}{n}$$

- Using the capacitive divider, we see that

$$n = 1 + \frac{C_{dep}}{C_{ox}} = 1 + \frac{\epsilon_{dep} t_{ox}}{\epsilon_{ox} t_{dep}}$$

- Note that $n > 1$ is the non-ideality factor of the channel control.

Exponential Current Flow



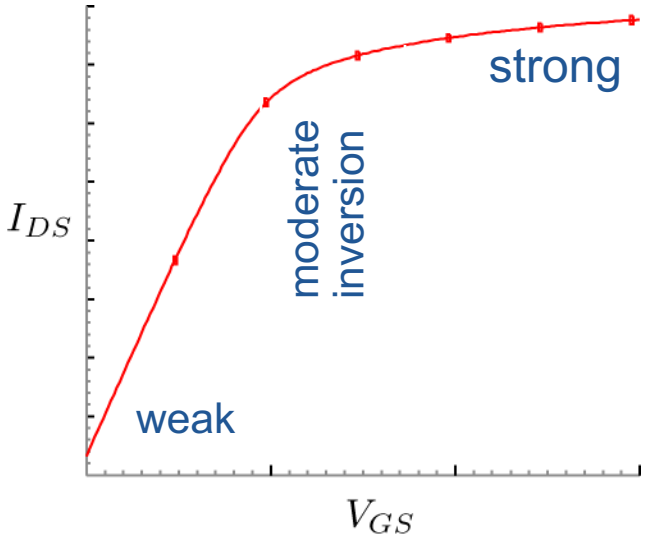
- We can now just borrow our equations from the BJT and write that

$$I_{ds} = \frac{W}{L} I_{ds,0} e^{\frac{q(V_{gs}-V_T)}{nkT}} \left(1 - e^{\frac{-qV_{DS}}{kT}}\right)$$

Weak Inversion

- Weak inversion (sub-threshold) like BJT
- n > 1: base controlled by capacitive divider
  - 0.18μm CMOS: n ~ 1.5
- "slow":
  - "large" C_{GS} for "little" current drive (see later)
- Moderate or weak inversion increasingly common:
  - Low power
  - Submicron L means "high speed" even in weak inversion
- Poor matching:
  - V_{TH} mismatch amplified exponentially
  - Avoid in mirrors

Moderate Inversion



- Weak inversion: current flow is predominately due to *diffusion*
- Strong inversion: current flow is predominately due to *drift*
- Moderate inversion: both *drift* and *diffusion* contribute to the current.
- Closed form equations for this region don't exist.

Patching Models: Smoothing Functions

- We have good models for weak inversion and strong inversion. Why not just interpolate in between?
- Here is an example interpolation model (EKV)

$$I_{DS} = \frac{W}{L}\mu C_{ox}(2n)\left(\frac{kT}{q}\right)^2\left(\left(\ln\left(1+e^{\frac{q(V_{GB}-V_{T0}-nV_{SB})}{2nkT}}\right)\right)^2-\left(\ln\left(1+e^{\frac{q(V_{GB}-V_{T0}-nV_{DB})}{2nkT}}\right)\right)^2\right)$$

- In strong inversion we have:

$$\ln(1+e^x)^2 \approx \ln(e^x)^2 = x^2$$

- In weak inversion we have: $\ln(1+x) \approx x$

Rôle of Compact Models

- Compact models are the interface between the technology and the design.
- A circuit designer learns about a process by experimenting with the compact model (rather than running expensive and time consuming experiments).
- Therefore compact models should be scalable with geometry, accurate across a wide temperature and bias voltage.
- Present day models are not necessarily “compact” (10,000+ lines of C code or 1000’s of lines of VerilogA) but should be fast enough to allow “real time” simulation

Compact Modeling Wish List

- Fast and reasonably accurate
- Good behavior (smooth, symmetric, correct trends, no unphysical features)
- Good RF functionality: Self-consistent AC NQS and TRAN NQS
- Accurate noise modeling including NQS
- Fewer and independent parameters
- Good statistical capability
- Don’t compromise on the following:
  - Efficiency and numerical stability
  - Symmetry and continuity
  - Physical behavior
  - Ease of parameter extraction

BSIM Models

- The *Berkeley Short-channel IGFET Models (BSIMS)* are the industry standard models for modern CMOS devices.
- They include many equations and parameters to model the complications in a real NFET or PFET device. A practical model card has 40-100 parameters and requires advanced software and extraction expertise to extract.
- BSIM3v3 was widely adopted for most foundries and was the workhorse for many years of analog/RF.
- BSIM4 is an enhanced BSIM model that includes the holistic thermal noise model, substrate network, stress, and gate current. It's still around today.

Real World Model Modules

Technology

- Short/Narrow Channel Effects on Threshold Voltage
- Non-Uniform Vertical and Lateral Doping Effects
- Mobility Reduction Due to Vertical Field
- Quantum Mechanic Effective Gate Oxide Thickness Model

Saturation

- Carrier Velocity Saturation
- Channel Length Modulation (CLM)
- Substrate Current Induced Body Effect (SCBE)
- Unified current saturation model(velocity saturation, velocity overshoot, source end velocity limit)

Leakage

- Gate Dielectric Tunneling Current Model
- Gate Induced Drain Current Model (GIDL)
- Trap assisted tunneling and recombination current model

RF

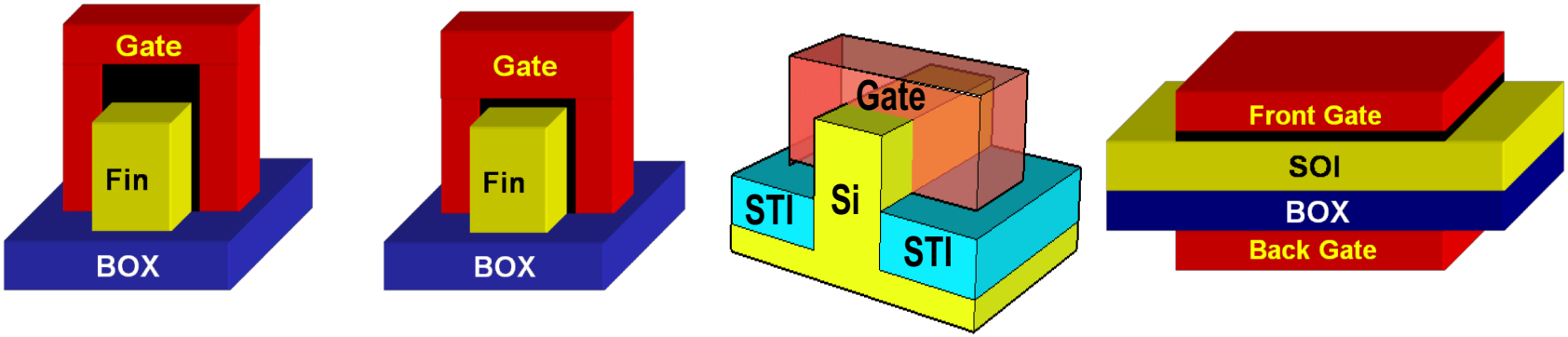
- RF Model (Gate & substrate resistance model)
- Unified Flicker Noise Model
- Holistic Thermal Noise Model

Parasitic

- Asymmetric Layout-dependent Parasitic Model
- Scalable stress effect model

FinFET and SOI Devices

- **Beyond 22nm technology node, bulk CMOS devices are less effective**
- **There are multiple gate structure (double, triple) and technology (bulk or SOI) options for new devices**



Double gate Triple gate Bulk-Si MGFET Independent gate

New Family of BSIM Models

- **BSIM-CMG**, a model for common multi-gate, i.e., FinFET device.
- **BSIM-IMG**, a model for independent multi-gate (usually front and back gate) device.
- **BSIM-SOI**, a model for silicon-on-insulator device.
- **BSIM-BULK**, a charge-based model for bulk MOSFET. (Also known as BSIM6)

BSIM "Hand Calculation" Models

- Requires many (many (many …)) assumptions.
- Assumptions:
  - V_T is given
  - Operate in strong-inversion.
  - Mobility model of mobmod = 2 is used (also applicable to other mobmod with slightly lower accuracy)
  - Bulk charge effect not significant in short channel devices.
  - Channel length modulation is the main contribution to r_{out}.

Effects to be included

- **Mobility degradation**

Define: $u_d = \frac{UA}{t_{ox}}$ mobility degradation coefficient

$u_d \approx 0.5\text{V}^{-1}$ for t_{ox}=10nm

- **Velocity saturation**

Define: $E_C = \frac{2v_{sat}}{U0}$ critical E-field for velocity saturation

$E_C \approx 2\times10^4\,\text{V/cm}$ (typical value)

Strong Inversion Current

$$V_{Dsat} = (V_G - V_T)\left[\frac{1 + u_d(V_G - V_T)}{1 + \left(u_d + 1/E_c L\right)(V_G - V_T)}\right]$$

$$I_{Dlin} = \mu_0 C_{ox}\frac{W}{L}\left(V_G - V_T - \frac{V_D}{2}\right)V_D\left[\frac{1}{1 + u_d(V_G - V_T) + \left(\frac{V_D}{E_C L}\right)}\right] = I_{Dlin(long)}\left[\frac{1}{1 + u_d(V_G - V_T) + \left(\frac{V_D}{E_C L}\right)}\right]$$

$$I_{Dsat} = \mu_0 C_{ox}\frac{W}{2L}\left[\frac{(V_G - V_T)^2}{1 + \left(u_d + 1/E_C L\right)(V_G - V_T)}\right] = I_{Dsat(long)}\left[\frac{1}{1 + \left(u_d + 1/E_C L\right)(V_G - V_T)}\right]$$

Equations of Derivatives

$$g_{msat} = \frac{I_{Dsat}}{(V_G - V_T)}\left[1 + \frac{I_{Dsat}}{I_{Dsat(long)}}\right] = \frac{I_{Dsat}}{(V_G - V_T)}\left[1 + \frac{1}{1 + \left(u_d + 1/E_C L\right)(V_G - V_T)}\right]$$

$$r_{out} = \frac{2\{(V_D - V_{Dsat}) + [1 + u_d(V_G - V_T)](V_G - V_T)\}L^2}{\mu_0 C_{ox} W l P_{CLM}[1 + u_d(V_G - V_T)](V_G - V_T)^2}$$

$$= \frac{\{(V_D - V_{Dsat}) + [1 + u_d(V_G - V_T)](V_G - V_T)\}L}{I_{Dsat(long)} l P_{CLM}[1 + u_d(V_G - V_T)]}$$

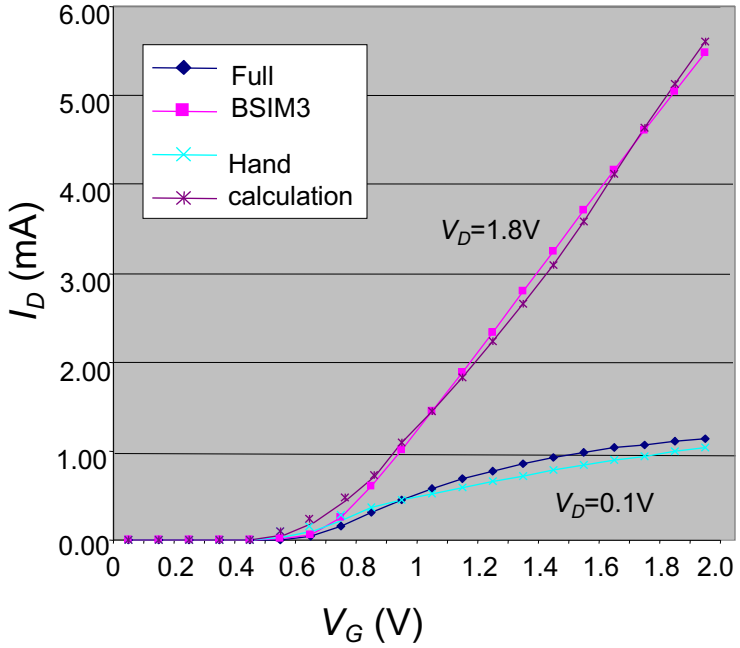
with $l = \sqrt{3 t_{ox} x_j}$

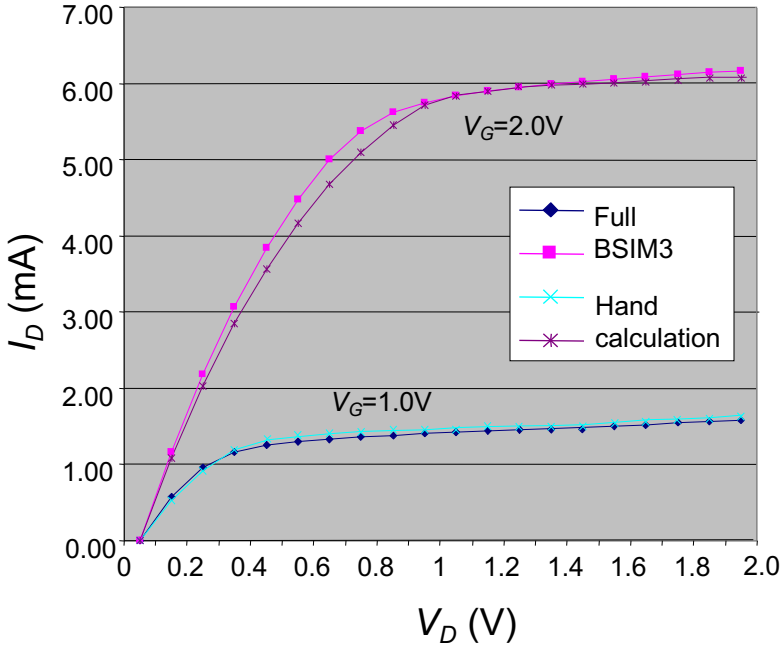
- Required parameters

  W, L, TOX, U0, UA, VSAT, VTH0, PCLM, XJ

Fitting Results

Comparison between full and simplified model

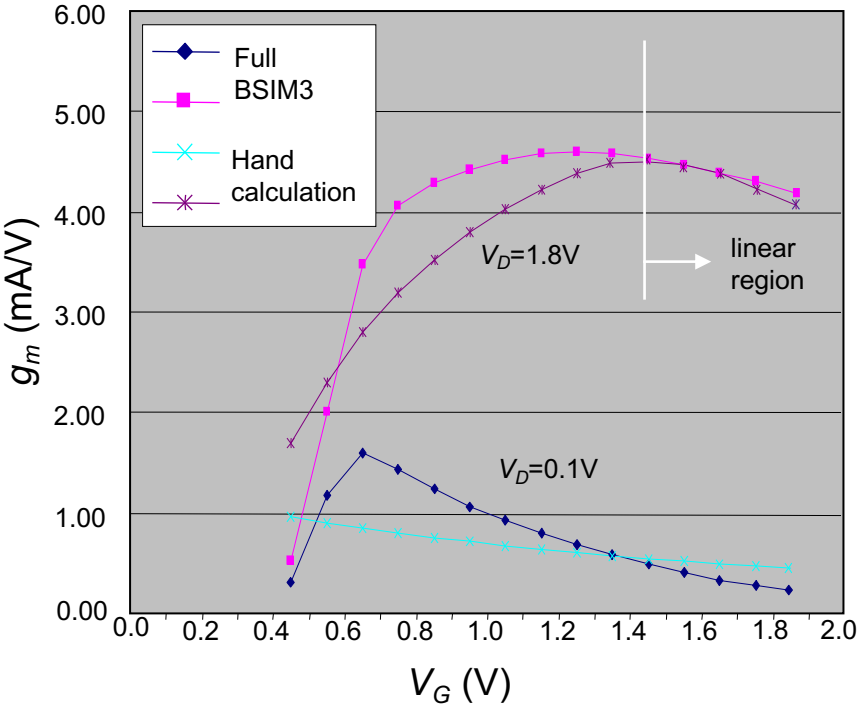


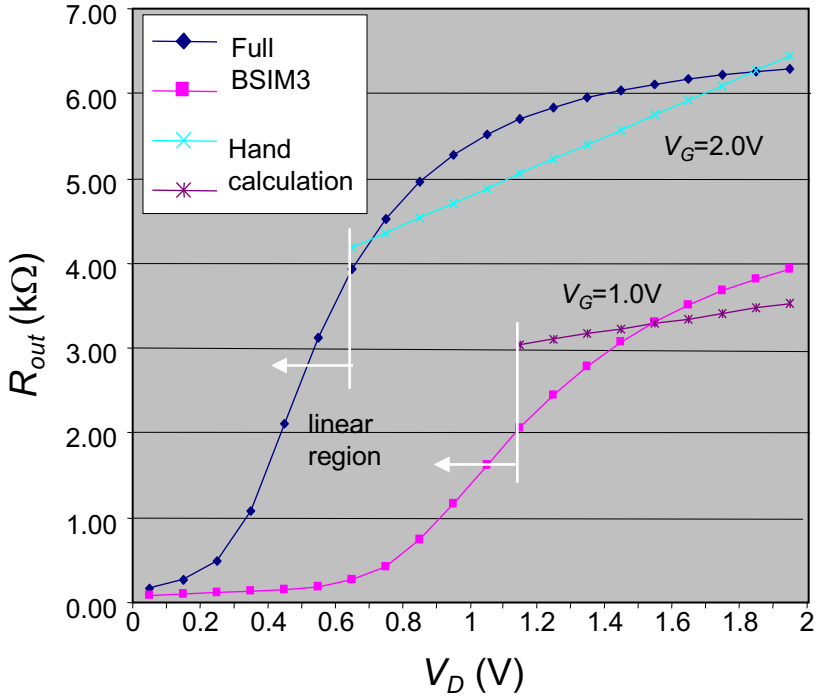


Parameter detail: TSMC 0.18μm process

t_{ox}: 4.1nm, W=10μm, V_{T0}=0.39V

Weakness of Model First Derivatives





“Hand Model” Conclusion

- Not easy to do. Even “simple” model is not convenient.
- Output resistance is key in analog design (for gain) but this is very difficult to model.
- Missing important regions such as moderate inversion.
- Can do better with a smoothing function to include these regions of operation.
- In this class we’ll learn to rely on the simulator in conjunction with some pretty simple small-signal models to do design.

Surface Potential Models

- The BSIM family of models rely on “threshold” voltage to include many important short-channel effects.
- Due to source referencing, BSIM3/4 models suffered from an inherent asymmetry that leads to discontinuity about V_{ds}=0.
- Bulk referenced models (such as EKV) solve this problem.
- Many “Next Generation” models use a surface potential formulation rather than threshold voltage.
- These models are more complicated an implicit by nature but have the advantage that they describe the long channel transistor behavior very well. In fact, many claim this is a “physical” model for this reason.
- BSIM-Bulk (BSIM6), MOS11, HiSIM, and SP (now PSP) are all examples of surface potential or charge based models that address these issues.

Other Important Issues

- **Noise Modeling**
- **Device Extrinsics**
- **Interconnect Models**
- **Matching Models**
- **Variations Modeling**