Next-Generation Compact Modeling*

Robert W. Dutton and Chang-Hoon Choi

Stanford University

Motivation*--Moore's Law scaling has led to ultra-short channel length devices that, while giving multi-GHz performance, present a host of new modeling challenges, especially for analog devices in SoC integration.*

▪There are a range of "**other**" issues that face compact modeling of nano-meter scale technology, including **parasitic effects** related to:

- ✓gate leakage,
- ✓substrate coupling
- ✓thermal limitations

▪There are also issues of **intrinsic device scaling**; high-level language (HLL) specifications for models that facilitate **model portability** is one example.

***Support from SRC(TI); CIS(Infineon, Philips); MARCO(MSD)**

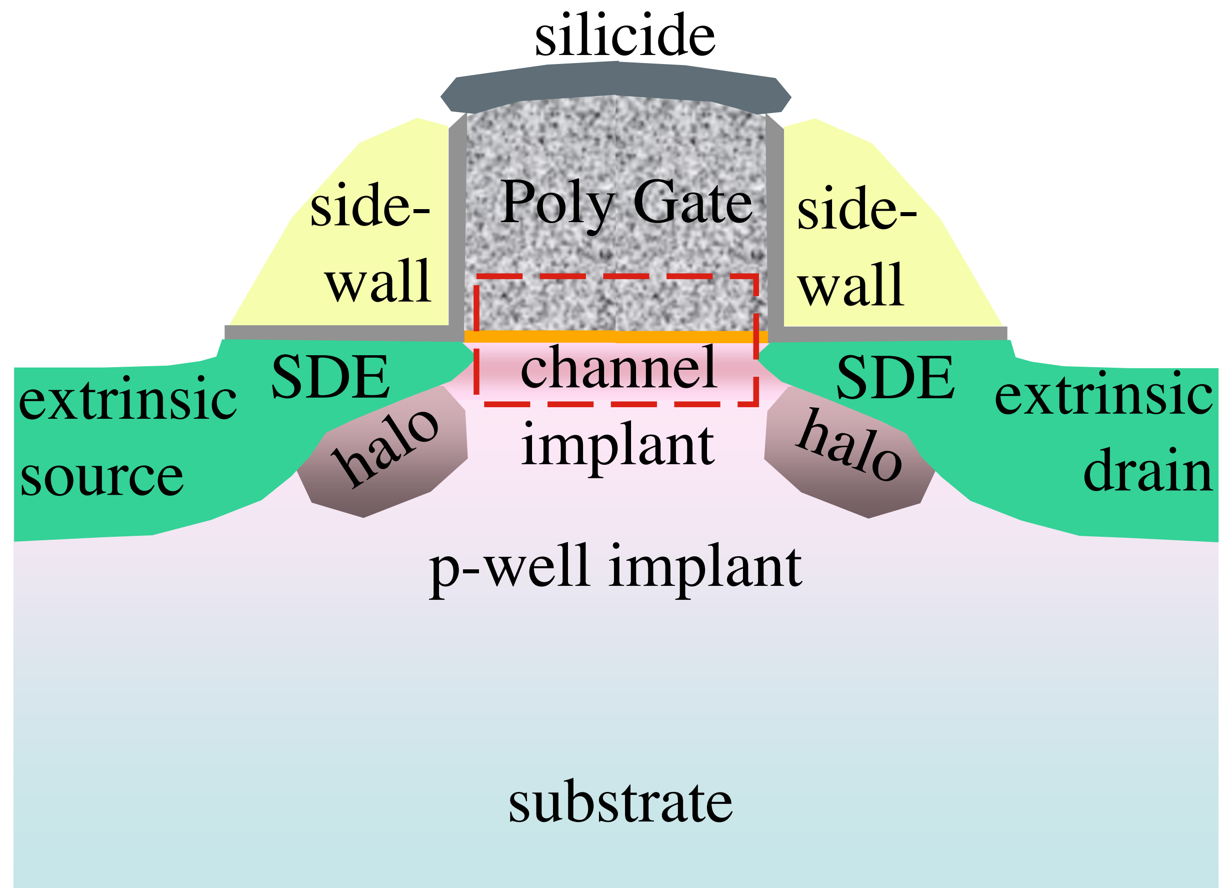
Outline*

- **Intrinsic Devices:**
  - Challenges of MOS Scaling
    - Gate Current, QM Effects…
      (**C.-H. Choi**, PhD 2002)
  - RF Modeling
    - Non-Quasi-Static & Substrate Effects
      (**J. Jang** PhD 2004)
    - Thermal Noise (**T. Oh**, PhD 2004)
  - Thermal Modeling
    - Self Heating in Nano-Devices (**E. Pop**)
    - RF Power MOS (**C. Ito**)
- **Model Portability**
- **Summary and Discussion**

---

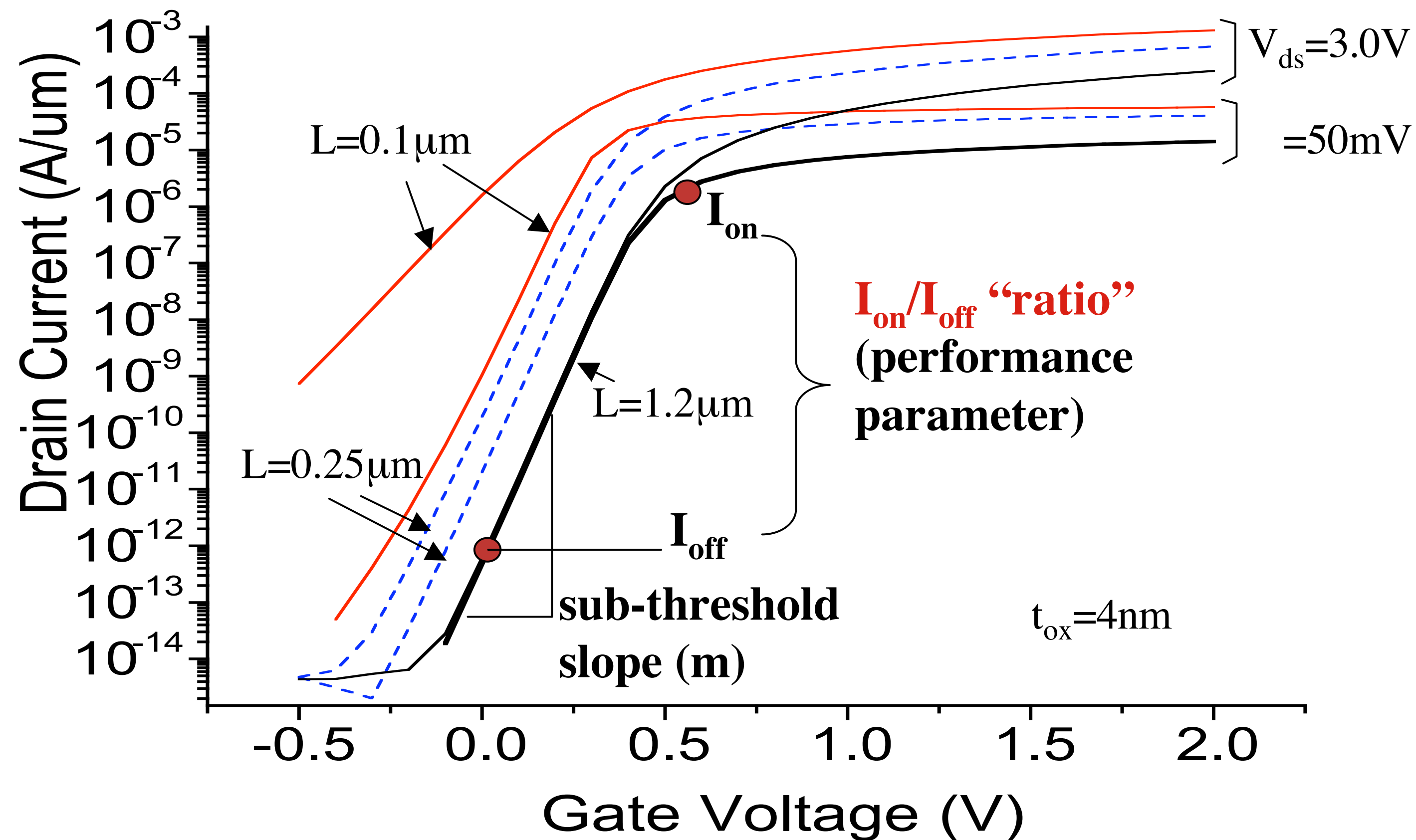
*PhD Theses and related publications: **www-tcad.stanford.edu**

Scaling Challenges of CMOS



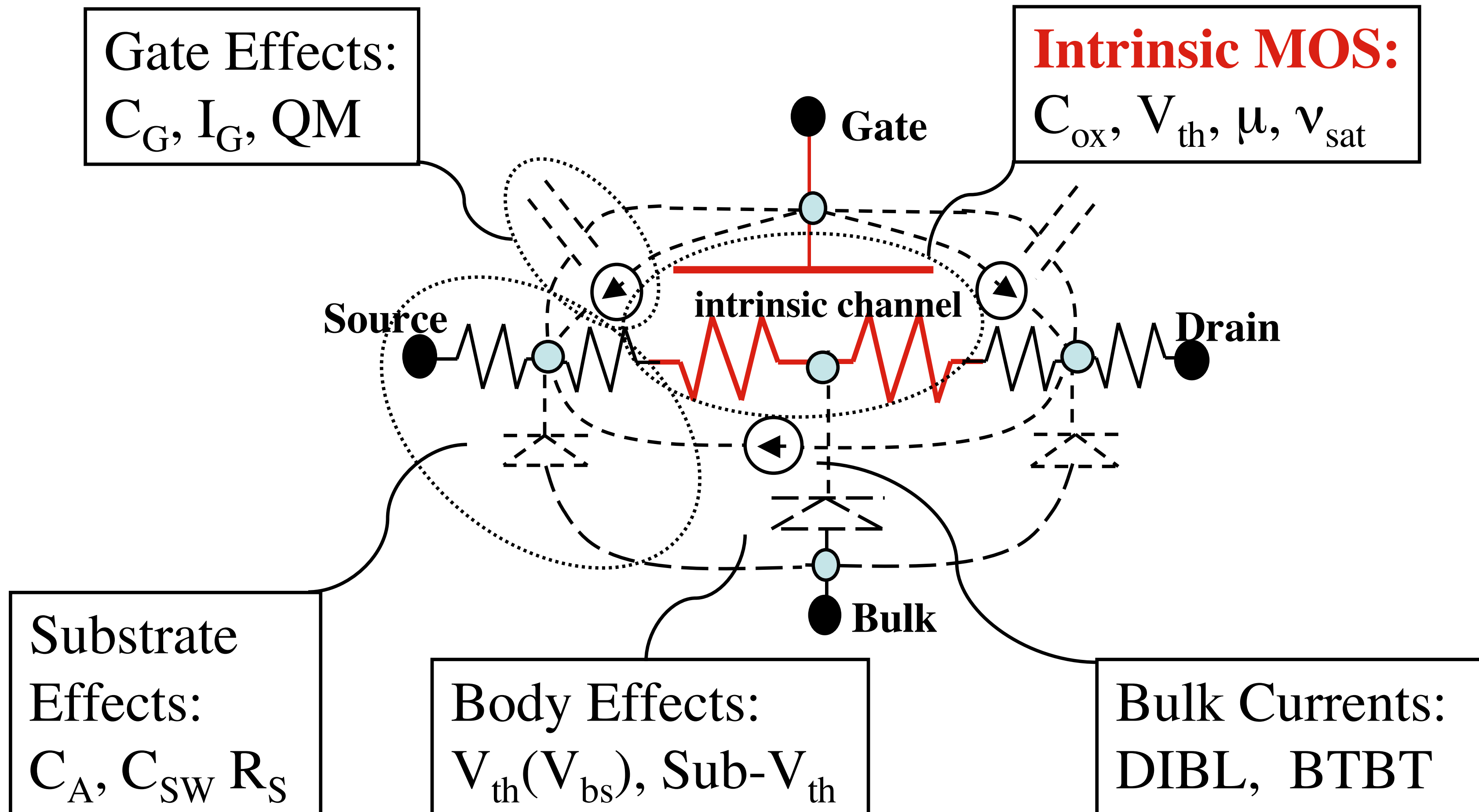
Cross-section of a 90nm N-channel MOS transistor, including details of gate, sidewalls and substrate doping profiles (SDE, channel/well implants, halo doping)

Threshold, Sub-Threshold & I_{on}/I_{off}



Semi-Log plot of Drain Current (I_{ds}) versus Gate Voltage (V_{gs}) with Drain Voltage (V_{ds}) as a parameter for three channel length NMOS devices.

Mapping Device Physics to Compact Models



Intrinsic MOS, Gate, Body/Bulk and Parasitic Substrate Capacitance/Resistance Effects

Channel Charge and Carrier Mobility!

$$I_{channel} \propto \overbrace{charge}^{} * \overbrace{carrier\ velocity}^{}$$

$$\propto \underbrace{C_{ox}(V_{gs} - V_{th})}_{charge} * \underbrace{E_{lateral}\ \mu_{channel}}_{carrier\ velocity}$$

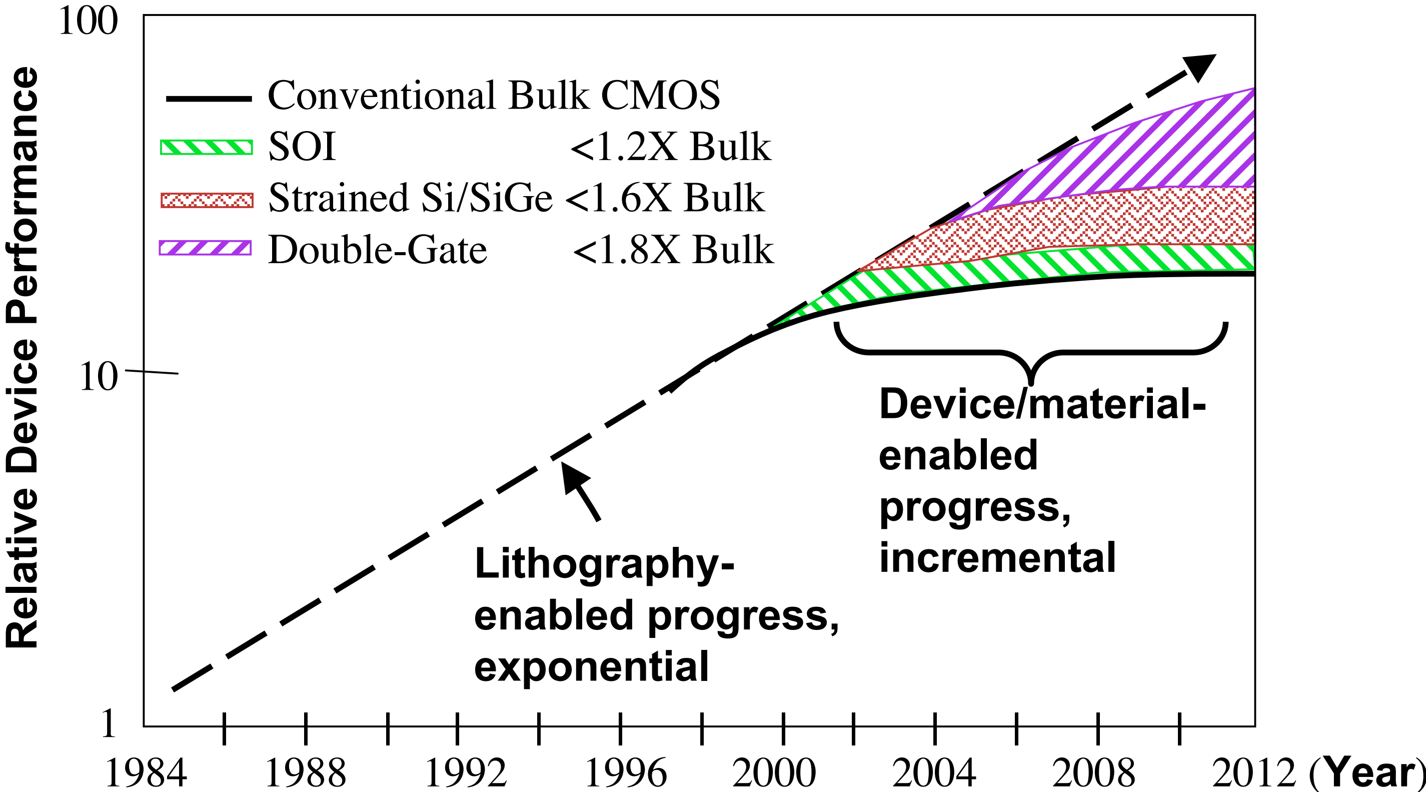
$$C_{ox} = \frac{\varepsilon_{Gate}}{x_{Gate}}$$

$$\frac{1}{C_G(V_{gs})} = \frac{1}{C_{ox}} + \frac{x_{Depletion}(V_{gs})}{\varepsilon_{Silicon}} = \frac{1}{C_{ox}} + \frac{1}{C_{Depletion}(V_{gs})}$$

$$\frac{1}{\mu_{channel}} = \sum_{\substack{i_{th} scattering \\ mechanisms}} \frac{1}{\mu_i}$$

First-order model of how MOS inversion layer channel current ($I_{channel}$) depends of gate-induced charge and carrier velocity. The physical parameters are C_G and $\mu_{channel}$.

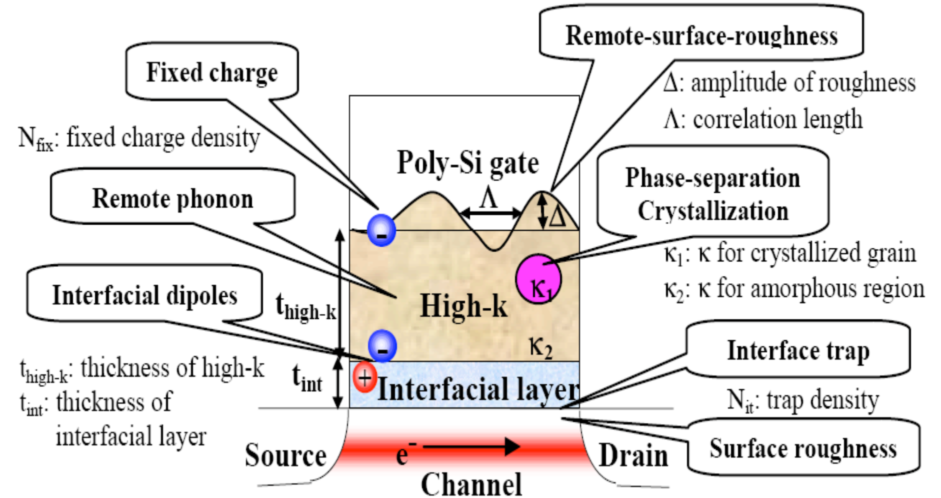
Scaling of MOS Performance

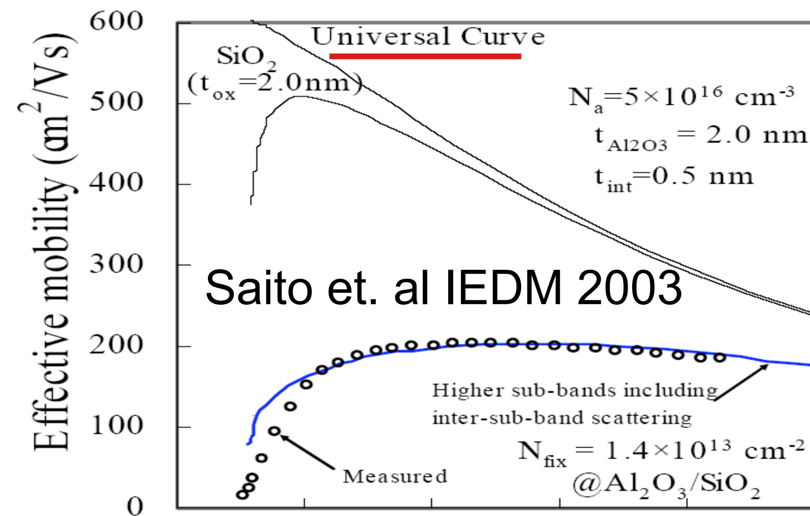


Mobility Degradation for High-k Gate Stacks

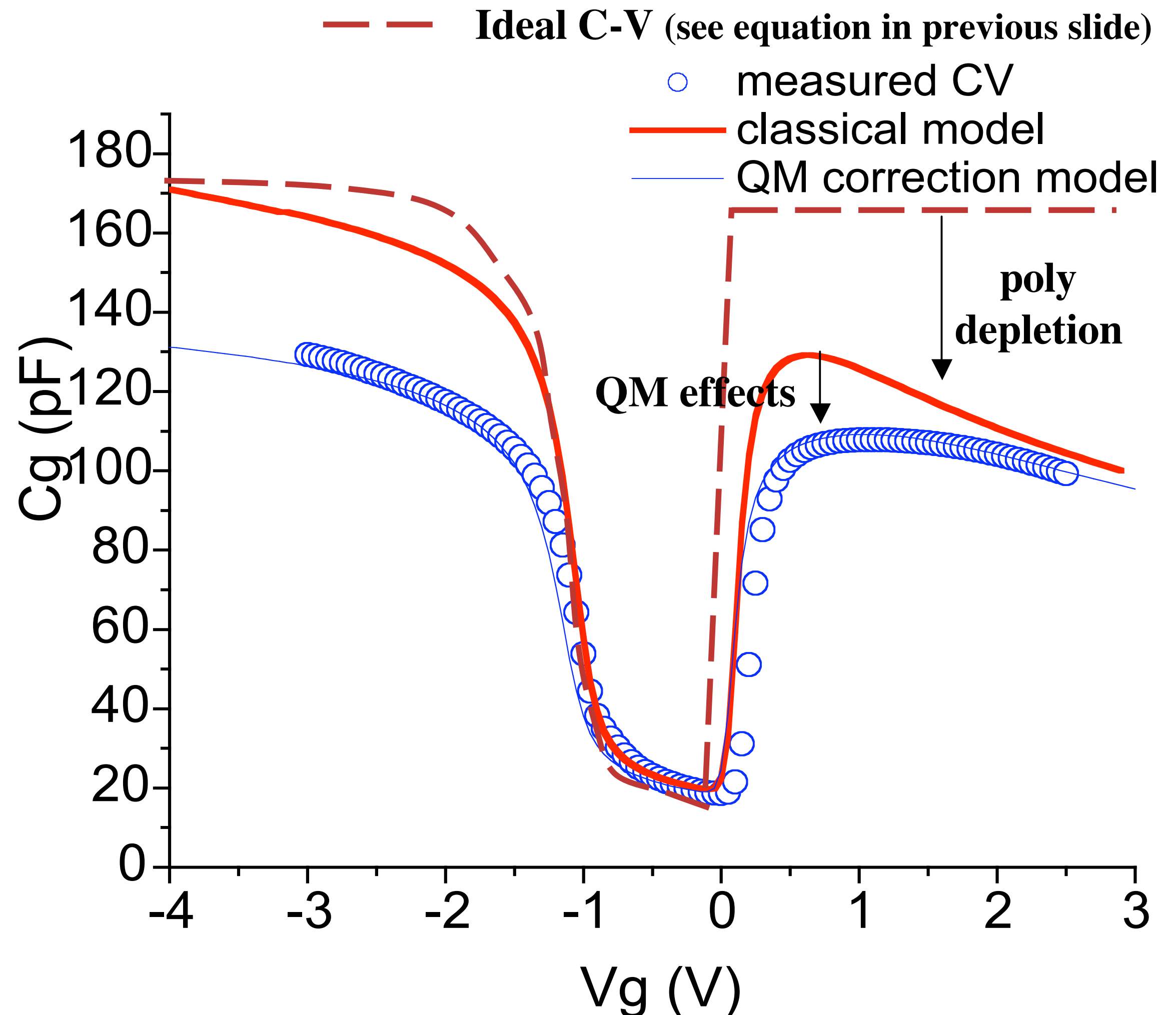
- **Various mechanisms responsible for degradation:**
  - Remote polar optical phonon scattering
  - Remote charge scattering
  - Remote surface roughness
  - Phase separation
- **Universal mobility curve**
  - Effective mob. dep. on effective field; good for acoustic phonon scatt. and surface roughness scatt.
  - Deviation at low inversion due to Coulomb scatt. for SiO_2
  - More severe deviation for high-k
    - Aggravated surface roughness
    - Stronger Coulombic scatt. at low and medium inversion

+ Gate Tunneling…



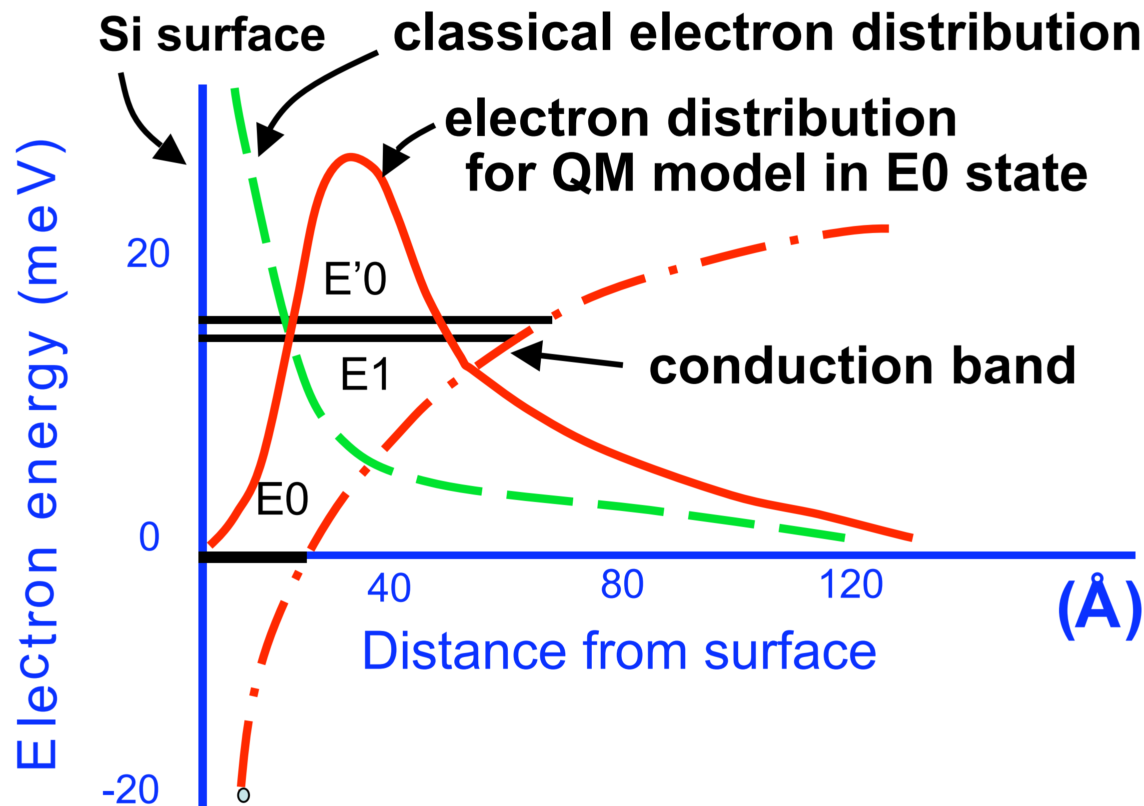


Poly-Gate and QM Effects



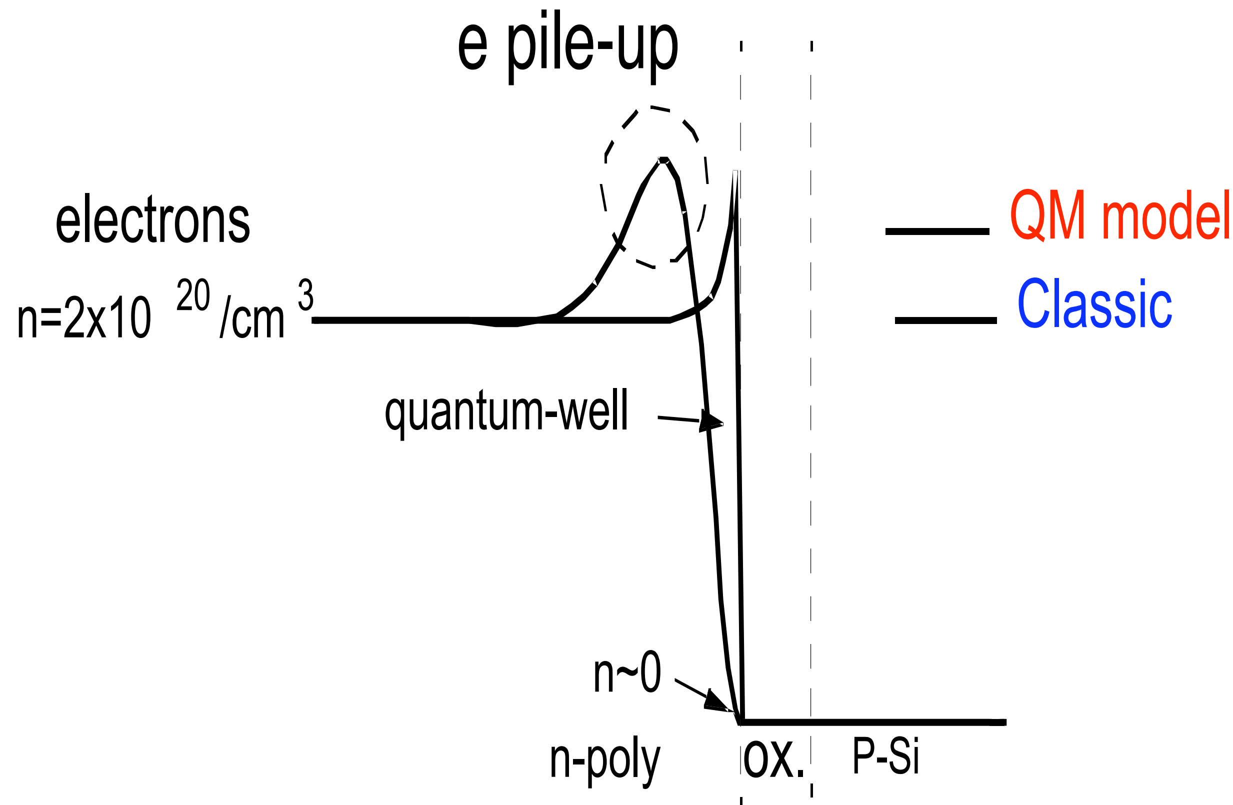
Measured and simulated curves of MOS gate capacitance vs. gate voltage for NMOS, poly-silicon gate, 2nm oxide thickness. Curves compare impact of QM effects

QM Effects on Channel Charge



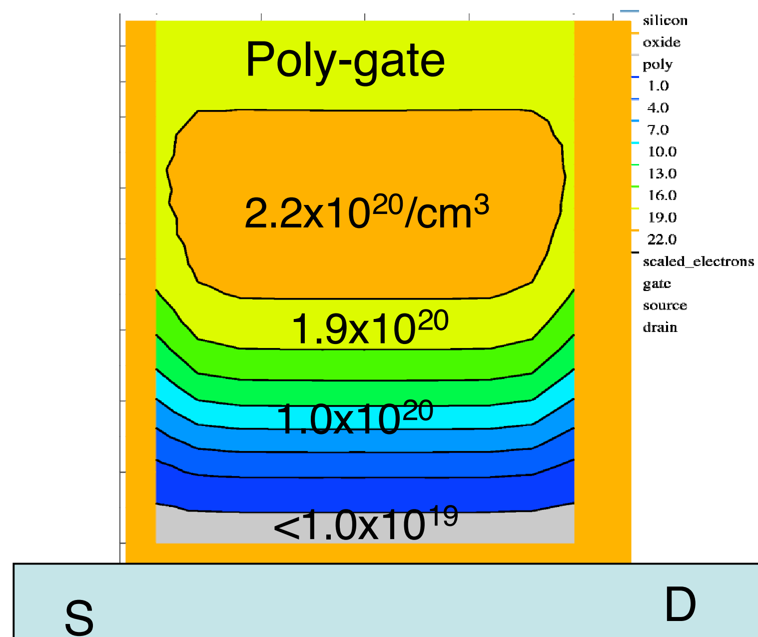
Surface region of bulk MOS device that shows: well potential created by the conduction band energy (- * -); discrete energy levels imposed by QM (E0, E1…); electron distribution associated with E0 level; electron distribution for classical theory (- - -)

QM "Poly Depletion"

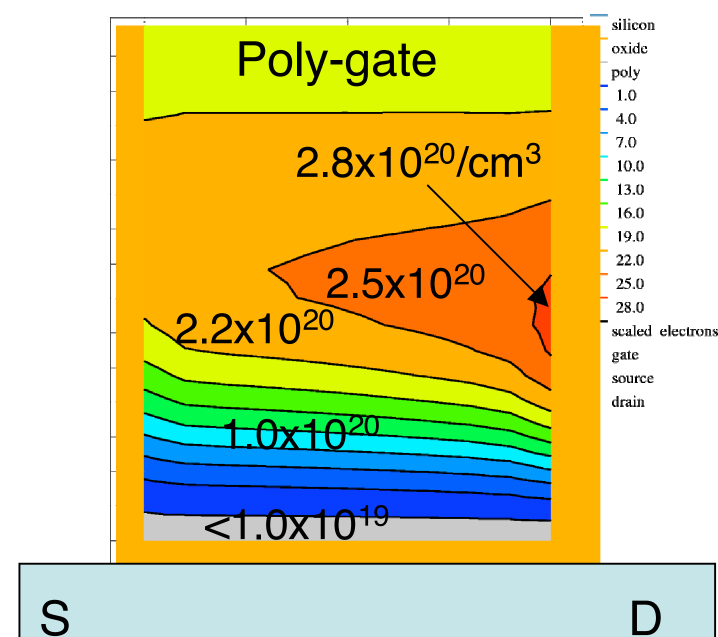


Electron distributions in the n-type poly-silicon gate of an NMOS transistor, based on classical theory versus that for a quantum-based (QM) solution for charge. The peak region shows a "QM depletion" resulting in V_{th} and C-V shifts

Implications of QM Poly Depletion (Fully Depleted Double-Gate-SOI)



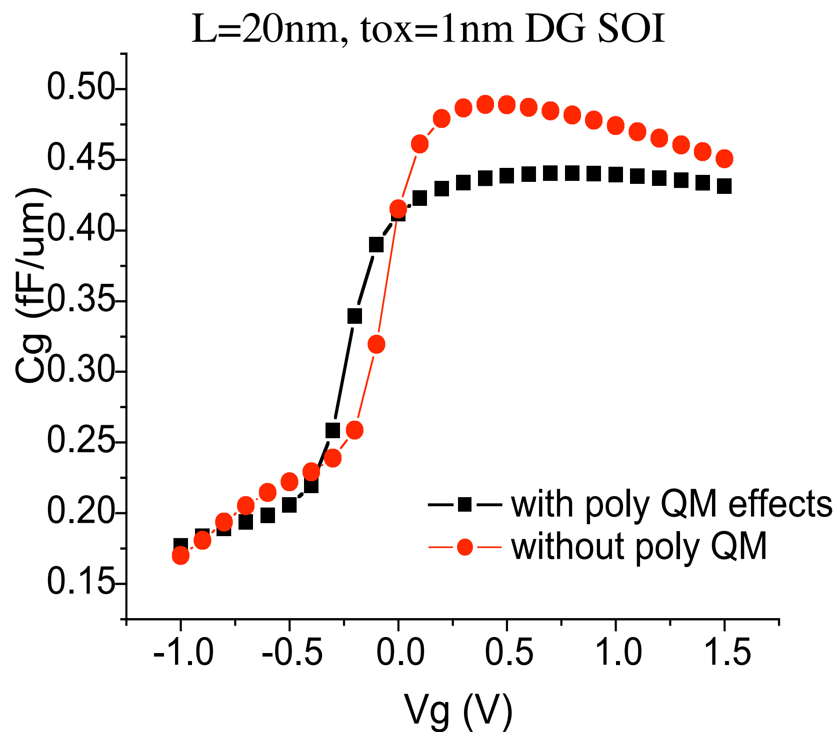
Electron contour for Vd=0V

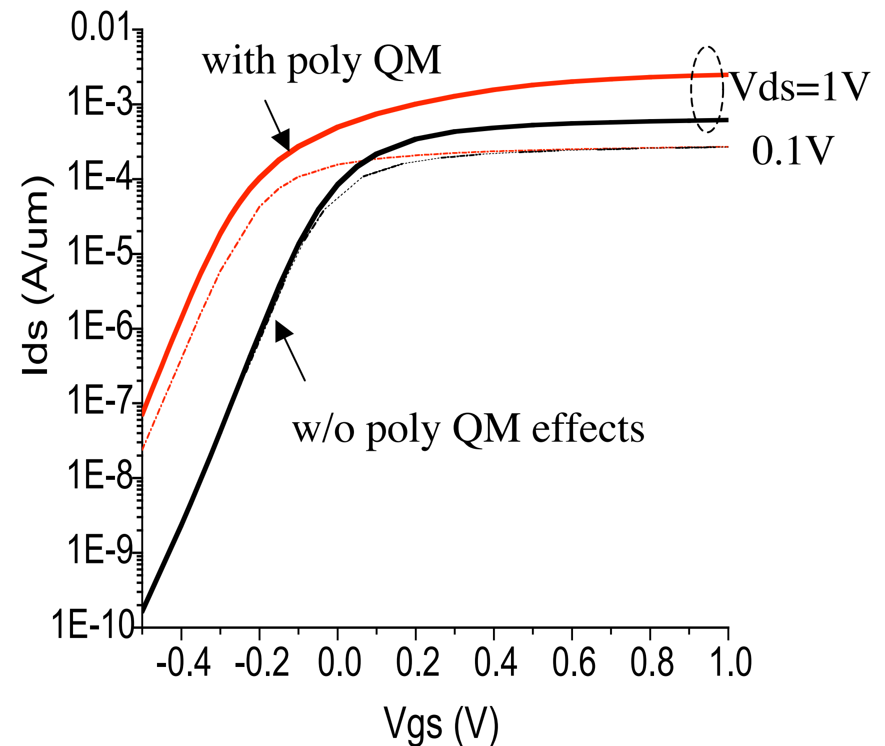


Electron contour for Vd=1V

2D simulations of QM poly depletion effects for two drain bias conditions. Results show significant gate depletion (left-side) and lateral effects that influence drain-induced barrier lowering (right-side)

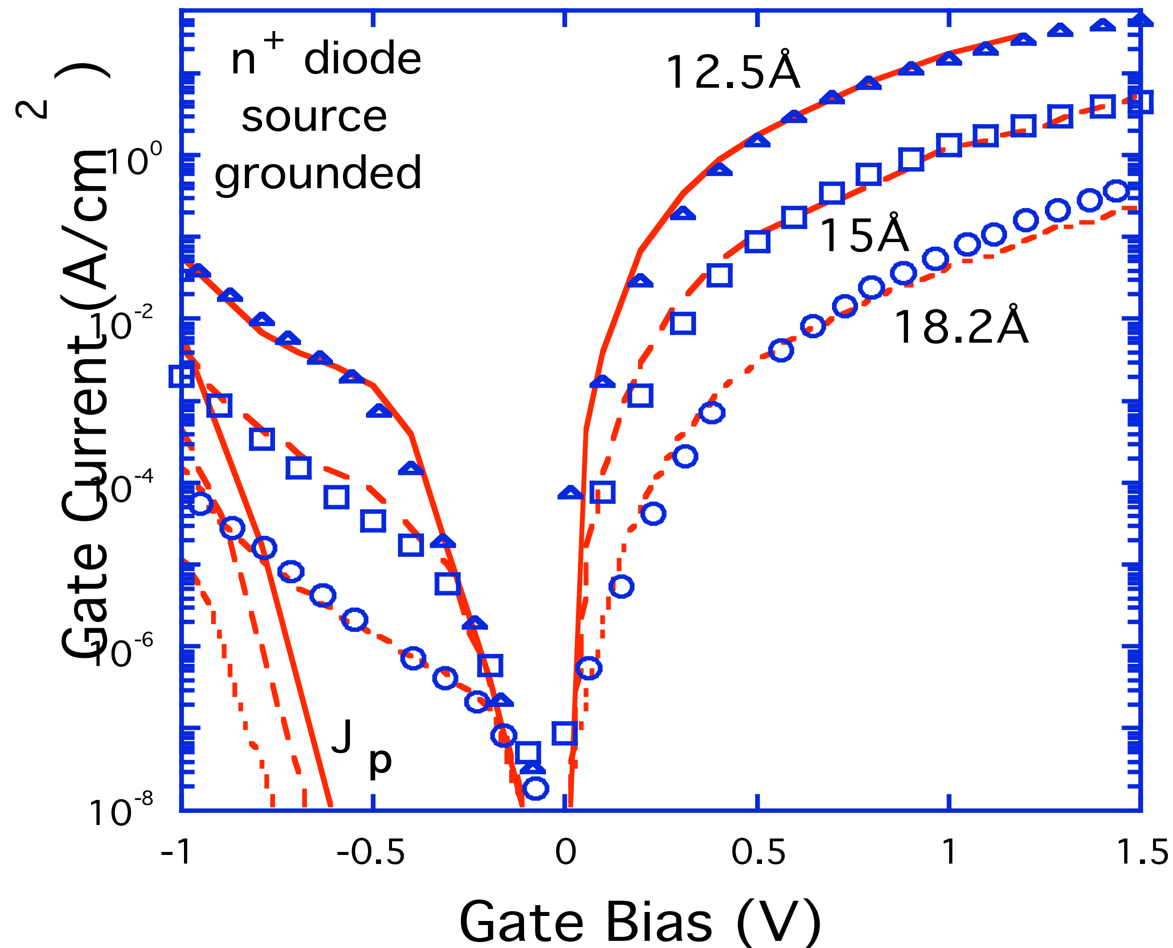
Threshold and Sub-Threshold Effects





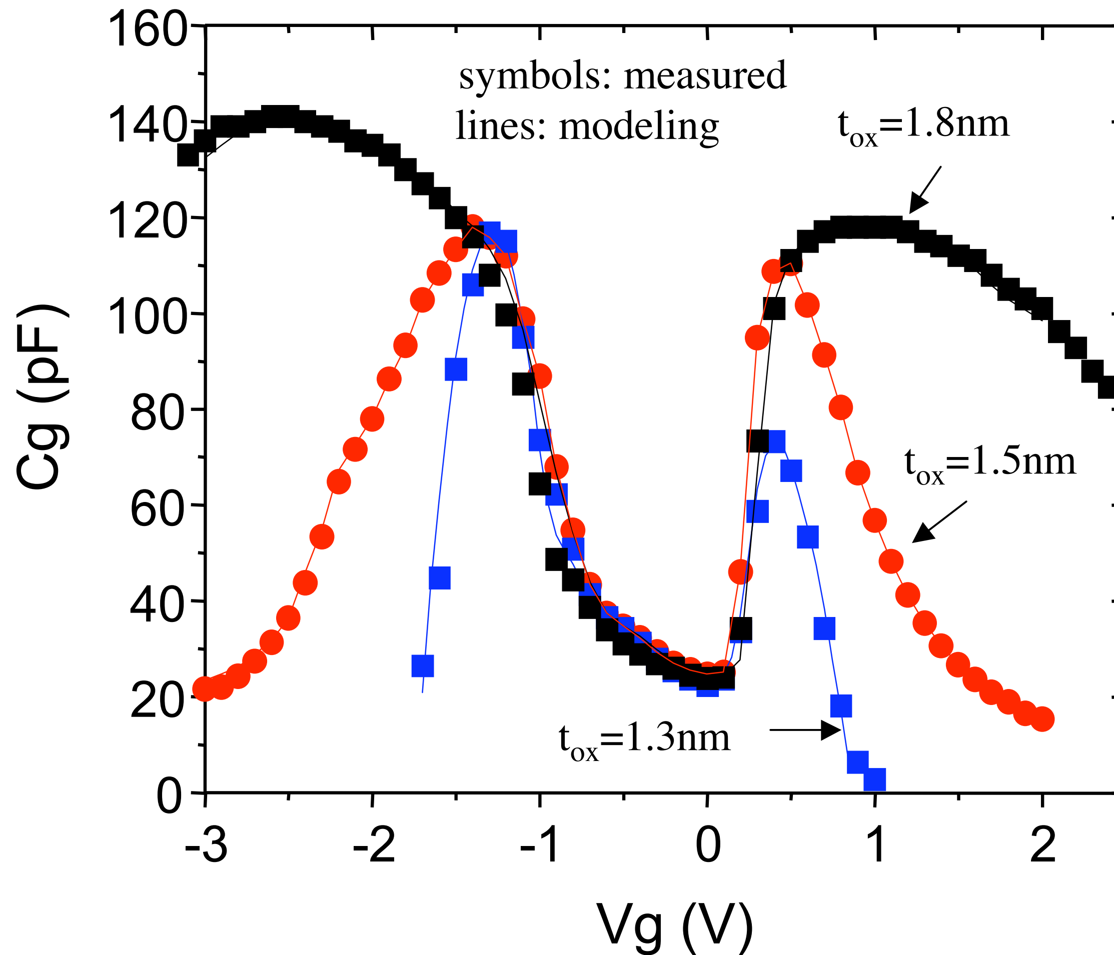
C-V and I-V characteristics for a 20nm channel length DG SOI transistor, comparing QM effects in the poly gate region with idealized (metal-like) gate.

Gate Current (Tunneling) vs. t_{ox}



Simulated and measured data for gate current in an MOS capacitor as a function of gate oxide thickness using NEMO; data provided by Hewlett-Packard.

Serious Limits to Scaling t_{ox}

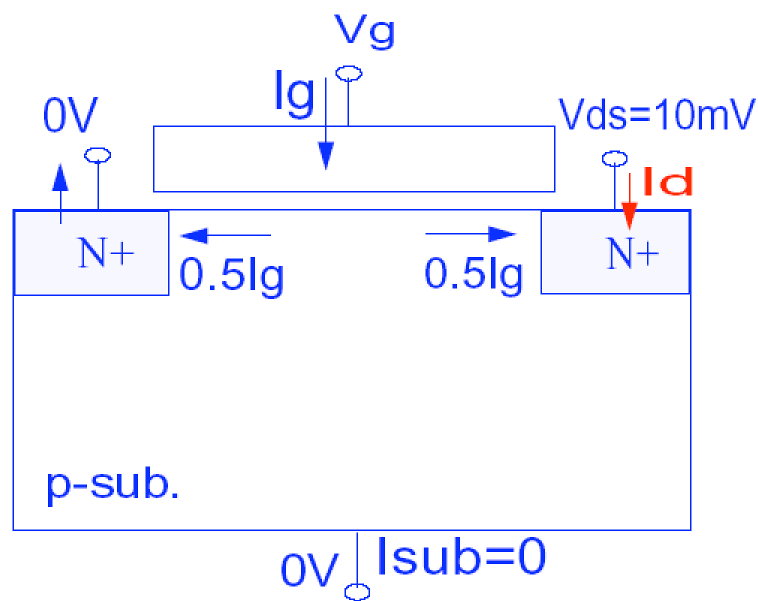


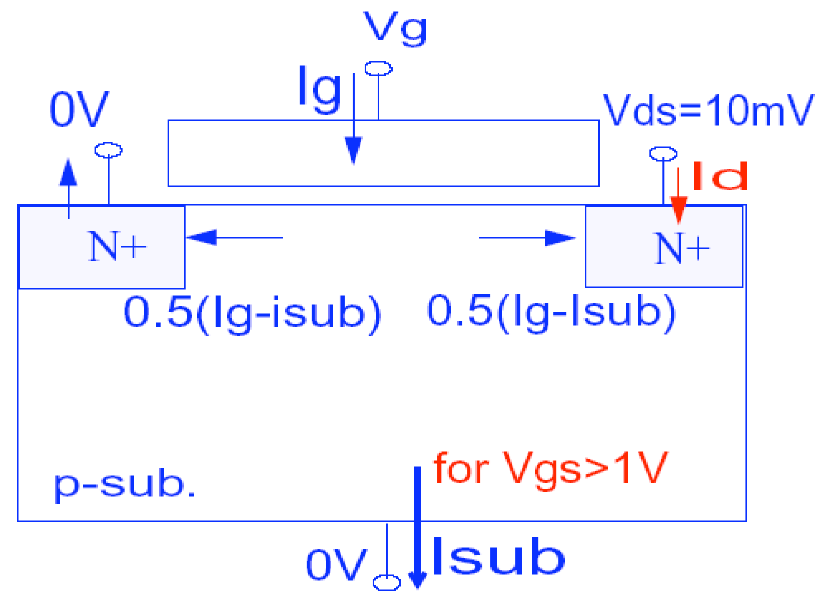
Measured and equivalent circuit simulations of imaginary component of input admittance, large area MOS capacitor structure with oxide thickness as parameter.

Drain Current with I_{sub}

- Distorted drain current (I_d) from the real drain current (I_{d0}) due to gate current (I_g): $I_{d0}=I_d+0.5I_g$ (Zeitzoff, EDL'03)

- But, note that $I_{d0}=I_d+0.5(I_g-I_{sub})$ in the presence of I_{sub}

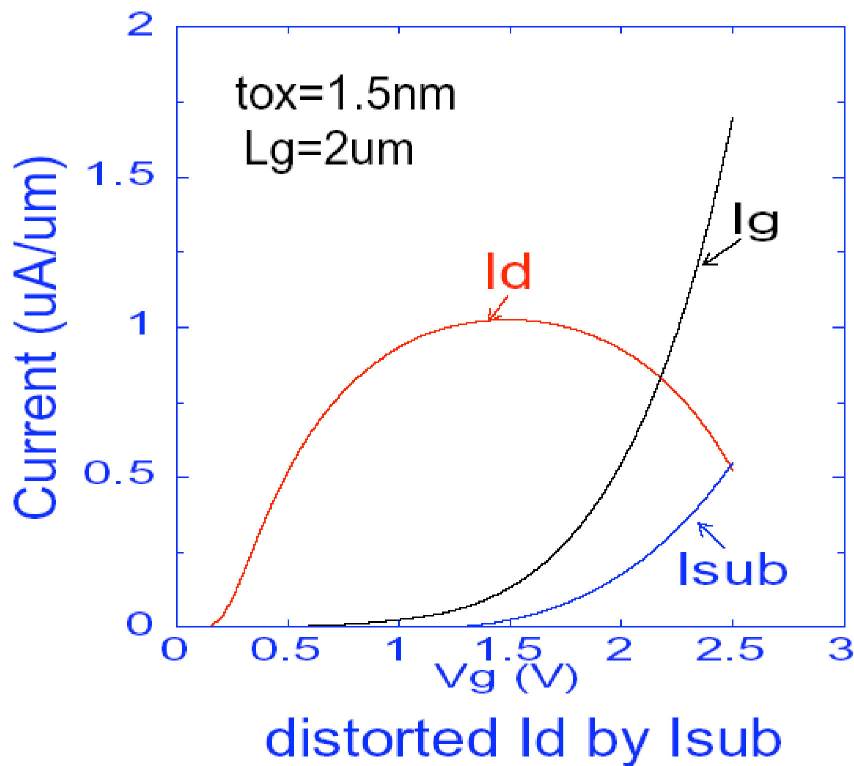




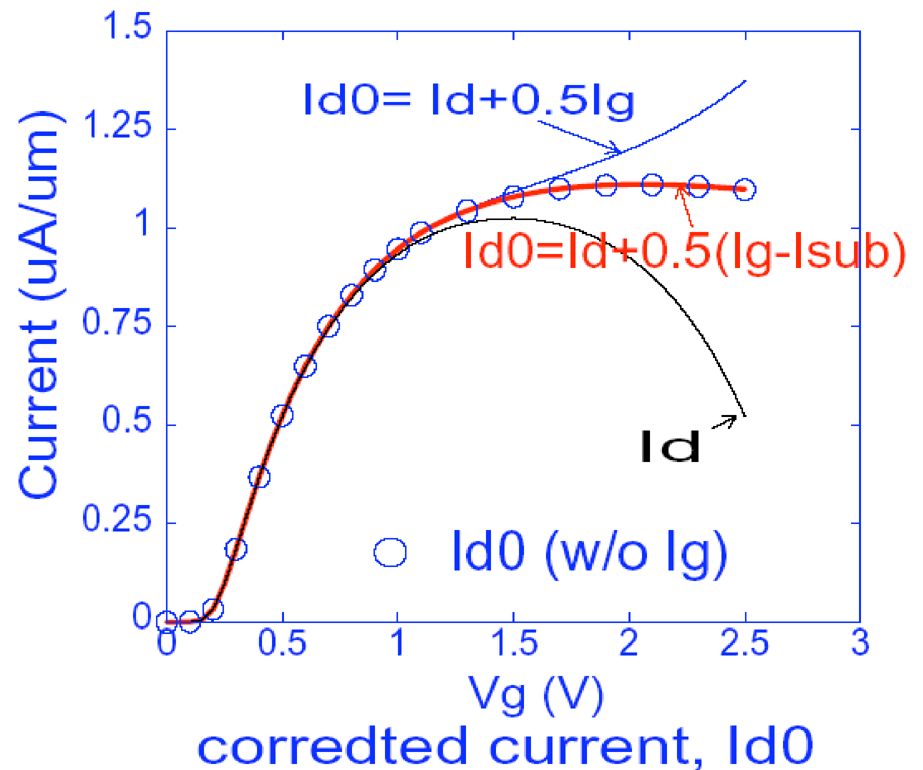
Correction of Drain Current

- Over-estimation of corrected drain current (I_{d0}) for
  $I_{d0}=I_d+0.5I_g$ expression

- New expression: $I_{d0}=I_d+0.5(I_g-I_{sub})$ in the presence of I_{sub}



distorted Id by Isub

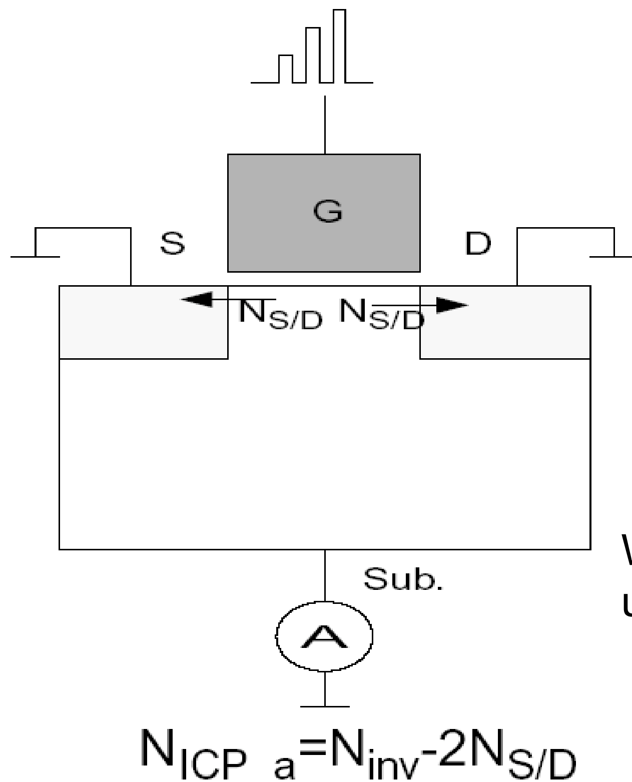


corredted current, Id0

Correction of Charge-Pumping Current

- Measurement of true inversion charge (N_{inv}) based on device symmetry (Keber, VLSI tech.'03)

- Elimination of N_{tun} from N_{inv} for high-K or thin oxide



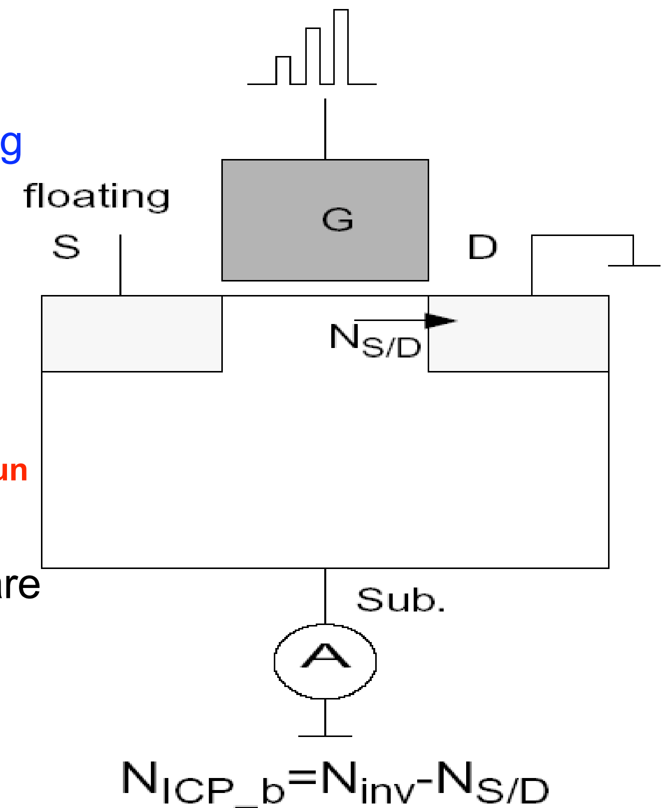
Inversion charge pumping

$N_{inv}=2N_{ICP_b}-N_{ICP_a}$

Modified N_{inv}

$N_{inv}=2N_{ICP_b}-N_{ICP_a}-N_{tun}$

Where DC values of N_{tun} are used (Masson et. al 1999)

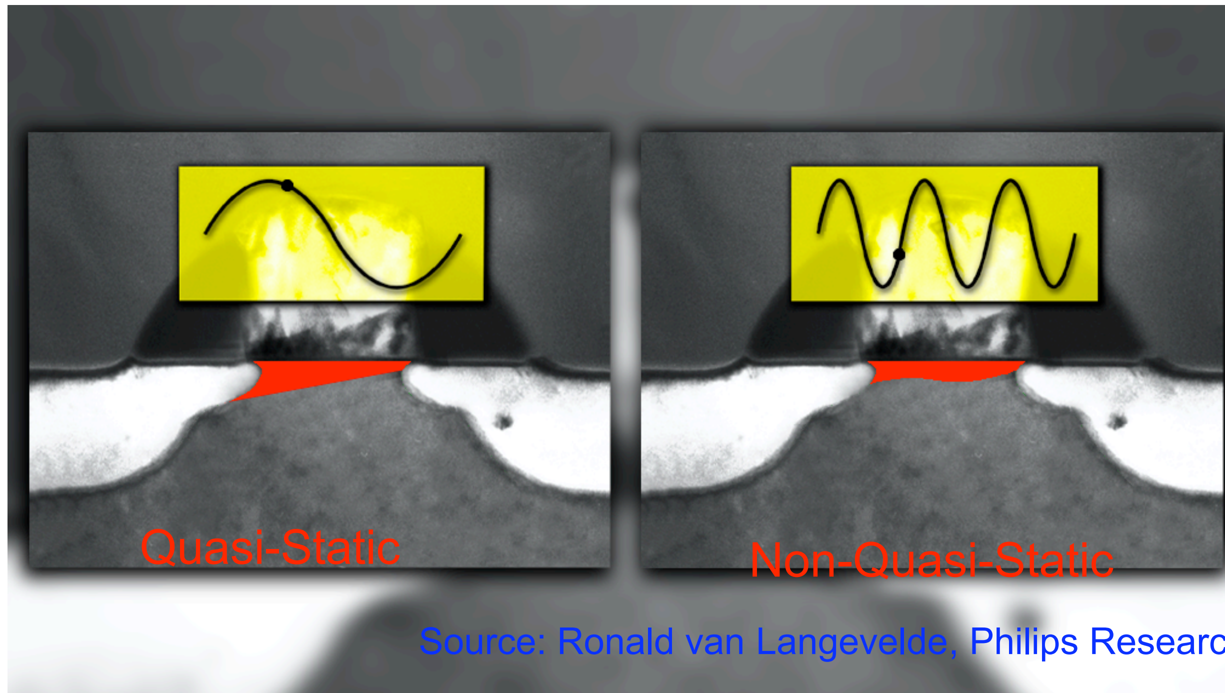


Small-Signal Modeling of RF MOSFET

Jaejune Jang
PhD 2004

Center for Integrated Systems
Stanford University

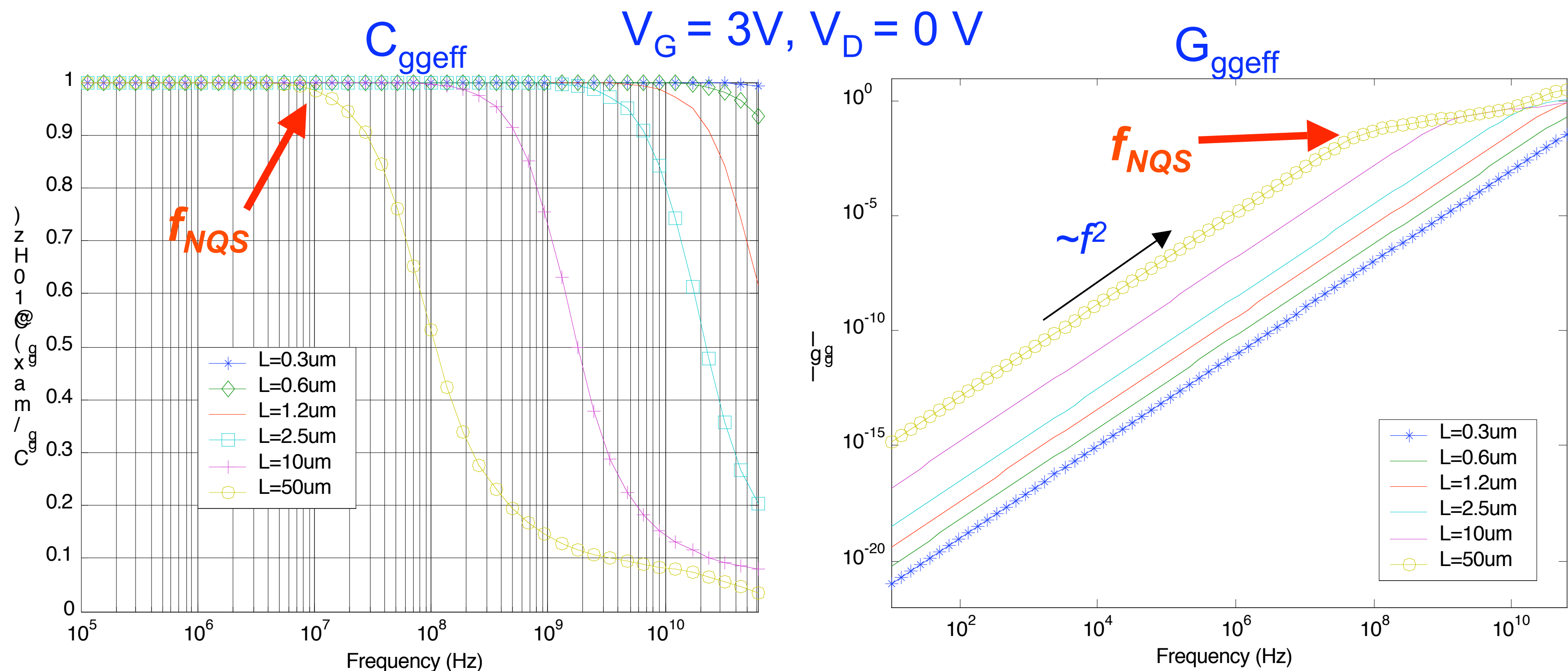
Non-Quasi-Static Effect



Source: Ronald van Langevelde, Philips Research Laboratories

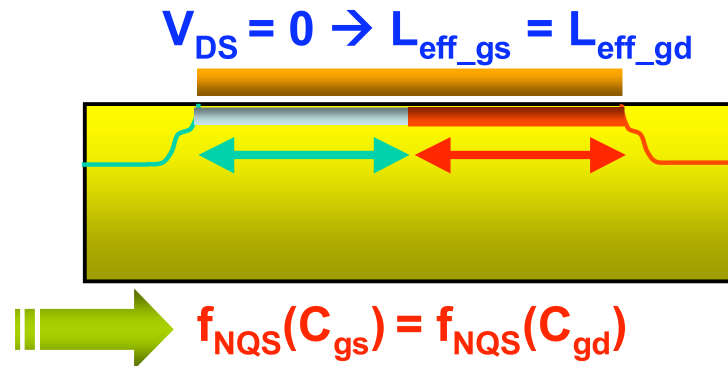
- Quasi-Static is when each terminal responds instantaneously to applied signal
- If switching time of v_g is compatible with transit time of inversion charge, QS approximation fails

NQS Effect on C_{ggeff} and G_{ggeff}



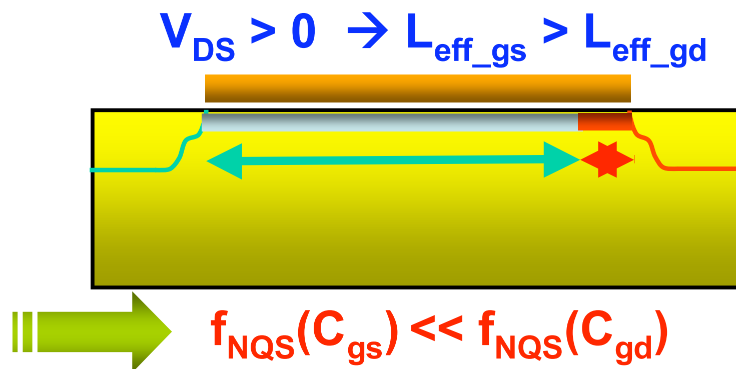
- Long channel device falls off at lower frequencies
- $G_{ggeff} \sim f^2$: directly proportional to gate induced noise
- f_{NQS} is defined as a frequency when NQS start

Bias Dependency of f_{NQS}



$$L_{eff_gs} \approx L \cdot \frac{C_{gs}}{C_{gs} + C_{gd}}$$

$$L_{eff_gd} \approx L \cdot \frac{C_{gd}}{C_{gs} + C_{gd}}$$



$$f_{NQS} = n\left[\frac{\mu(V_{GS} - V_T)}{2\pi L_{eff}^2}\right]$$

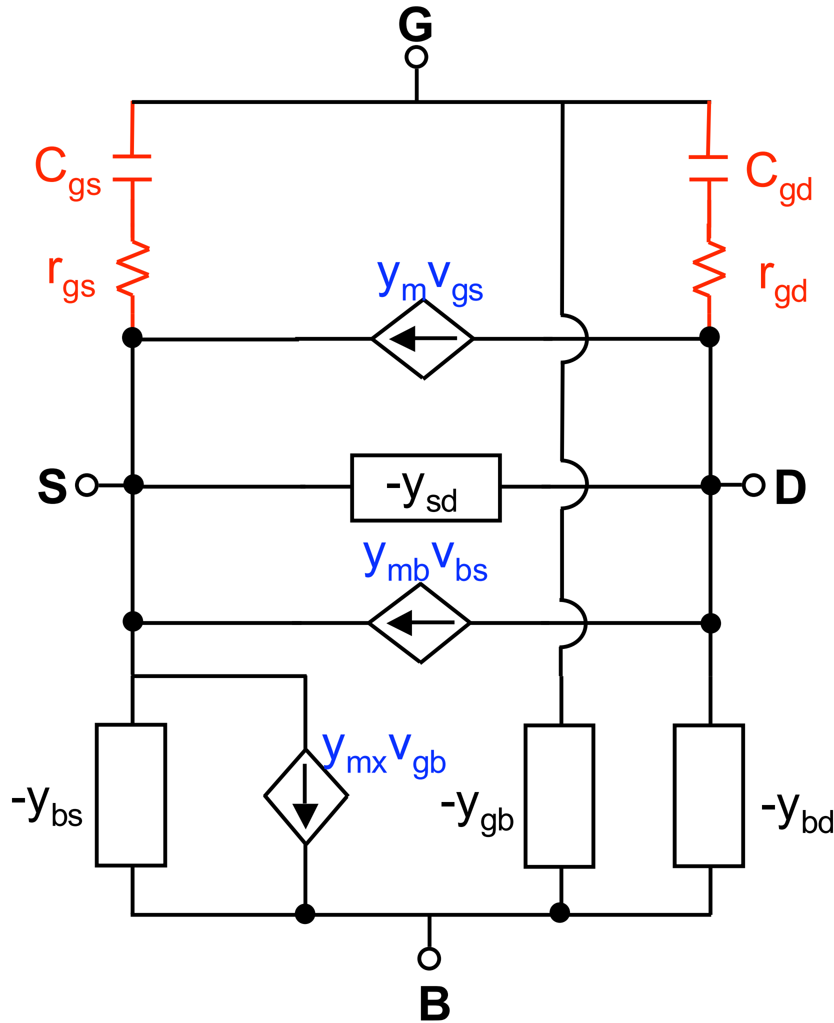
→ Long Channel

$$f_{NQS} = n\left[\frac{v_{sat}}{2\pi L_{eff}}\right]$$

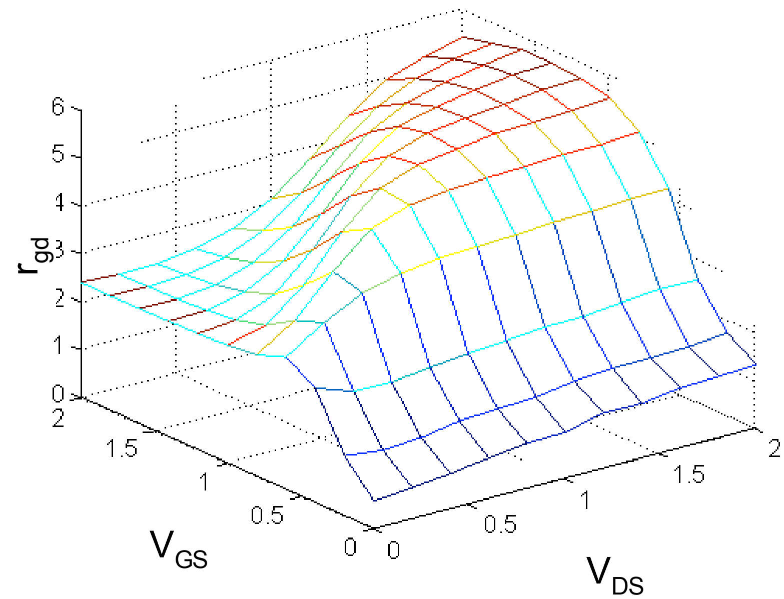
→ Short Channel

- L_{eff} is determined by charge sharing b/w source and drain
- Bias dependency of f_{NQS} is primarily function of L_{eff} , mobility (μ) and, saturation velocity (v_{sat})

y_{gs} and y_{gd}

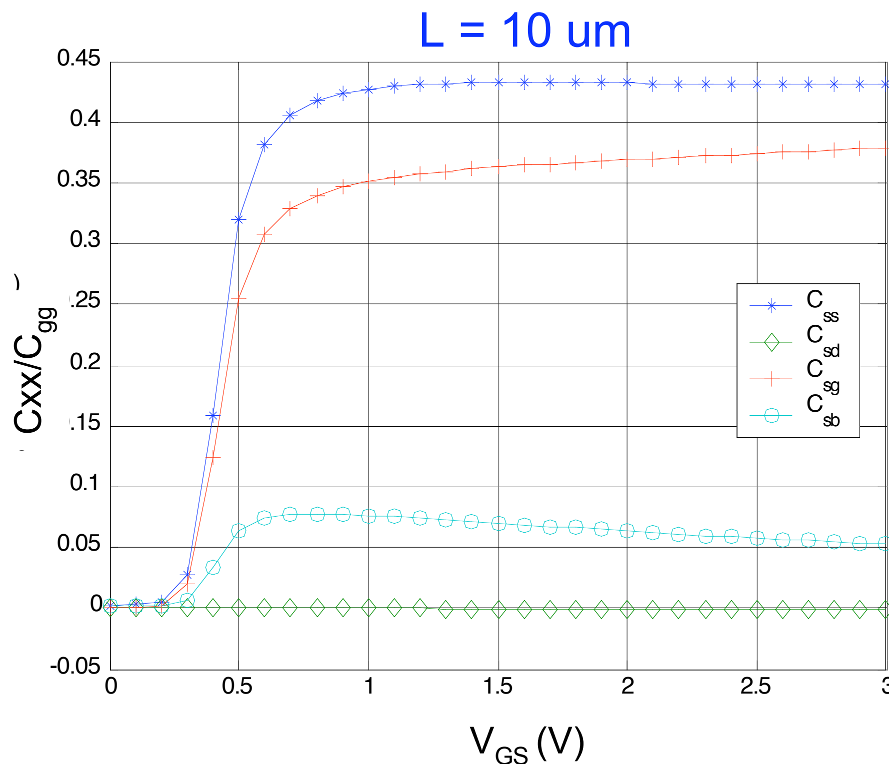


- y_{gs} and y_{gd} can be represented as series RC
- $C_{gs}r_{gs}/C_{gd}r_{gd}$ determines f_{NQS}

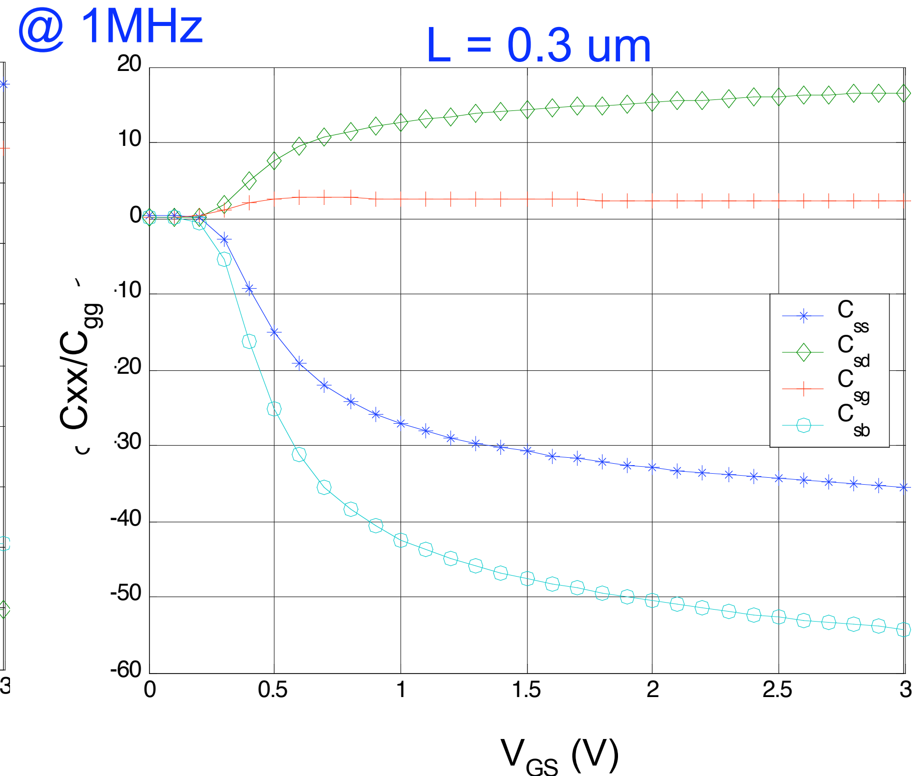


L = 1.2 um (measured)

Capacitance vs. Channel Length

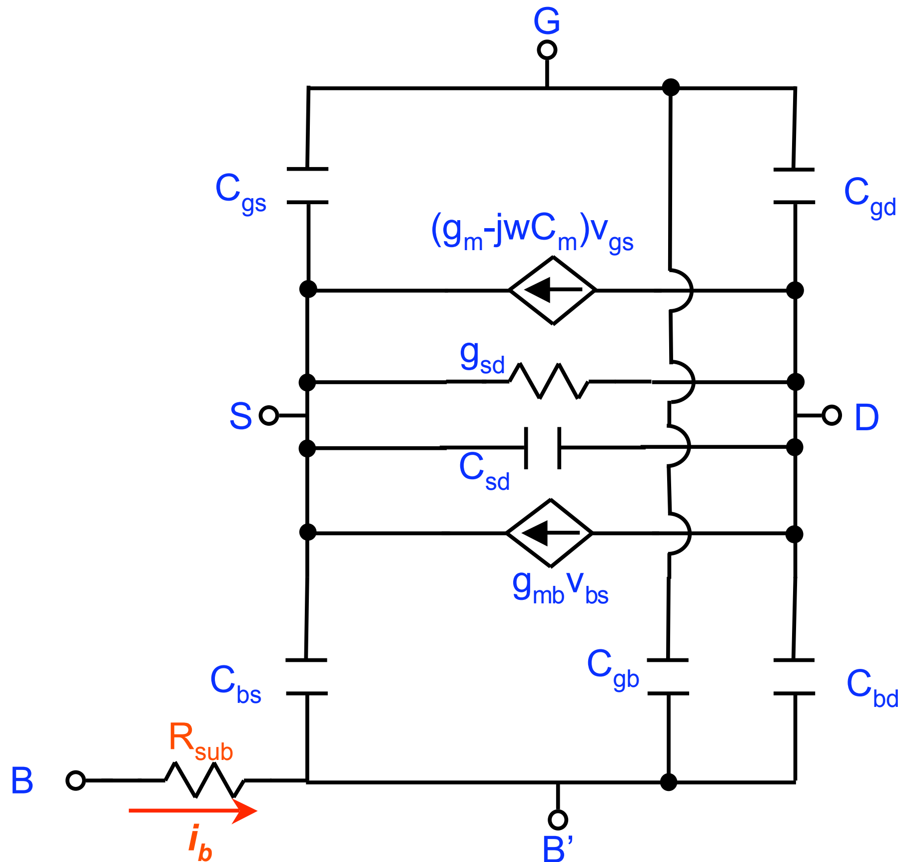


@ 1MHz



- All 16 capacitances agree well with charge-based capacitances for long channel device
- Short channel capacitances show quite different behavior
- This is due to existence of substrate resistance (R_{sub})

Impact of R_{sub} on Terminal Admittance



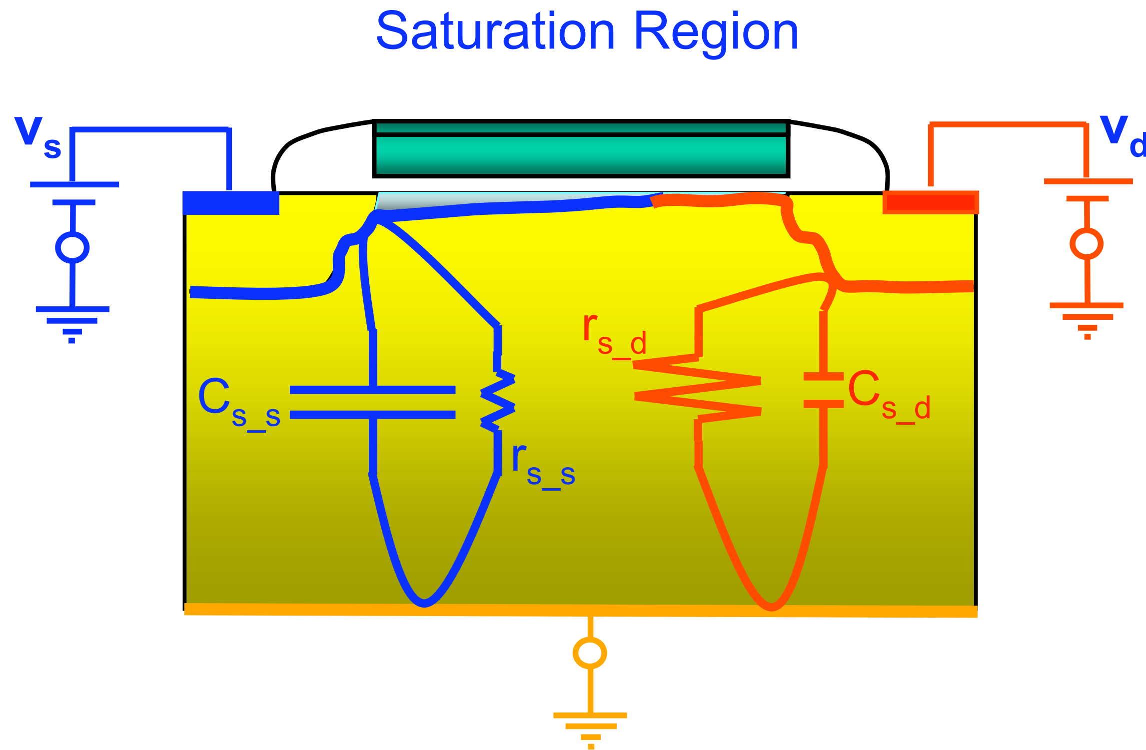
$$C_{sdeff} = C_{sd} + \frac{R_{sub} g_{mb}}{1 + \omega^2 R_{sub}^2 C_{bb}^2} C_{bd}$$

$$g_{sdeff} = g_{sd} + \frac{\omega^2 R_{sub}^2 g_{mb}}{1 + \omega^2 R_{sub}^2 C_{bb}^2} C_{bb} C_{bd}$$

As L decreases, R_{sub} and g_{mb} increase

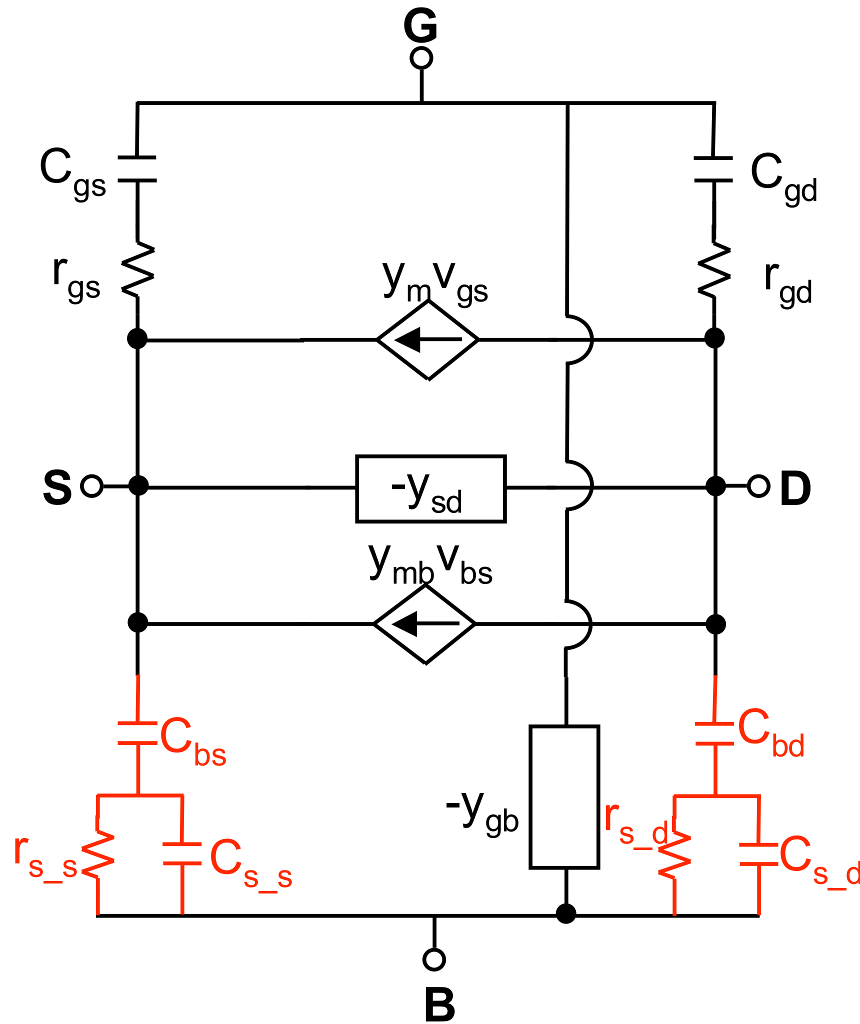
- R_{sub} is amplified through g_{mb}
- Impacts following y-parameters (y_{dd}, y_{ds}, y_{db}, y_{dg}, y_{ss}, y_{sg}, y_{sd}, and y_{sb})

Bias Dependency of Substrate Network

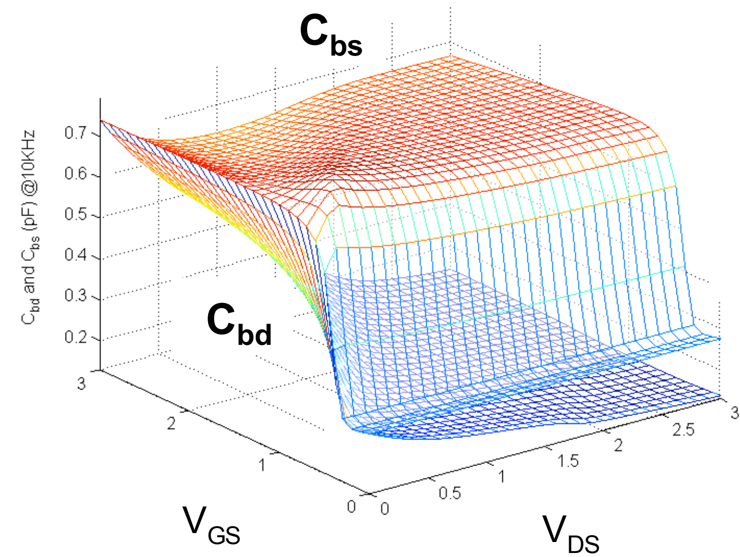


- It's been reported that substrate network is bias independent → not true
- $C_{s_s} > C_{s_d}$ & $r_{s_s} < r_{s_d}$ in saturation region
  - Surface area of effective electrode is bias dependent due to charge sharing

y_{bs} & y_{bd}



- C_{sub} and R_{sub} is a strong function of bias
- C_{bs} and C_{bd} are also strongly bias dependent



Noise Analysis of deep-submicron MOSFETs

Tae-young Oh

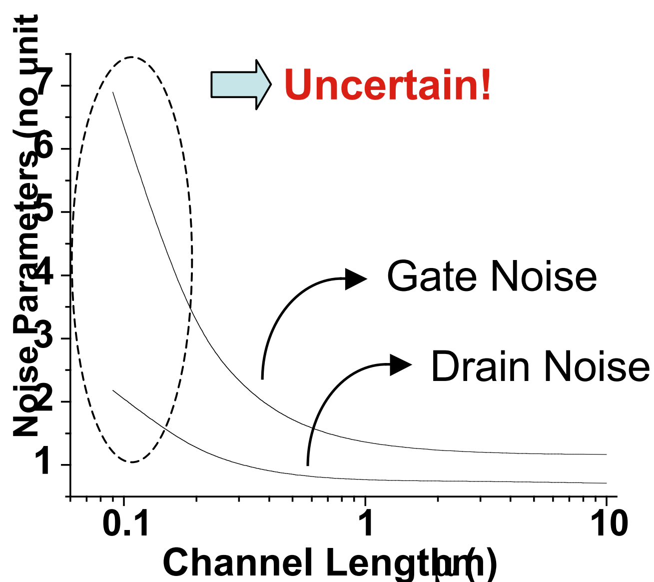
PhD 2004

Center for Integrated Systems

Stanford University

MOS Channel Length and Excess Noise

- Higher speed devices offer great promise for RF
- However, the changes in intrinsic noise of scaled MOS devices is not clearly understood

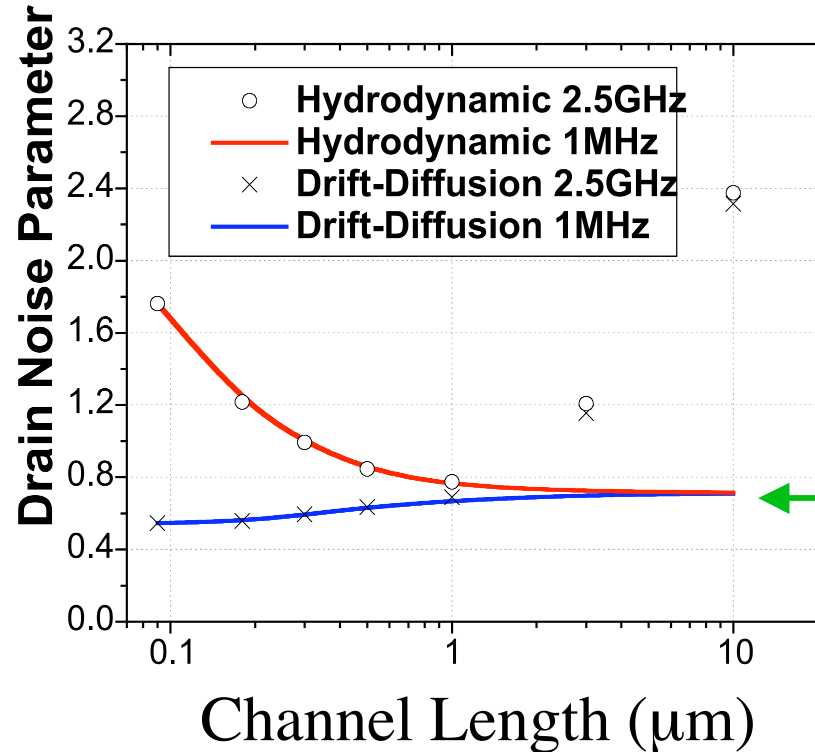


- The amount and source for this excess noise were still uncertain
- High electric field in short channel MOS device should have relation with this excess noise

  - Simulation has to handle this carefully

(**D**rift-**D**iffusion vs. **H**ydro-**D**ynamic models)

Channel Length vs. Drain Noise Parameters

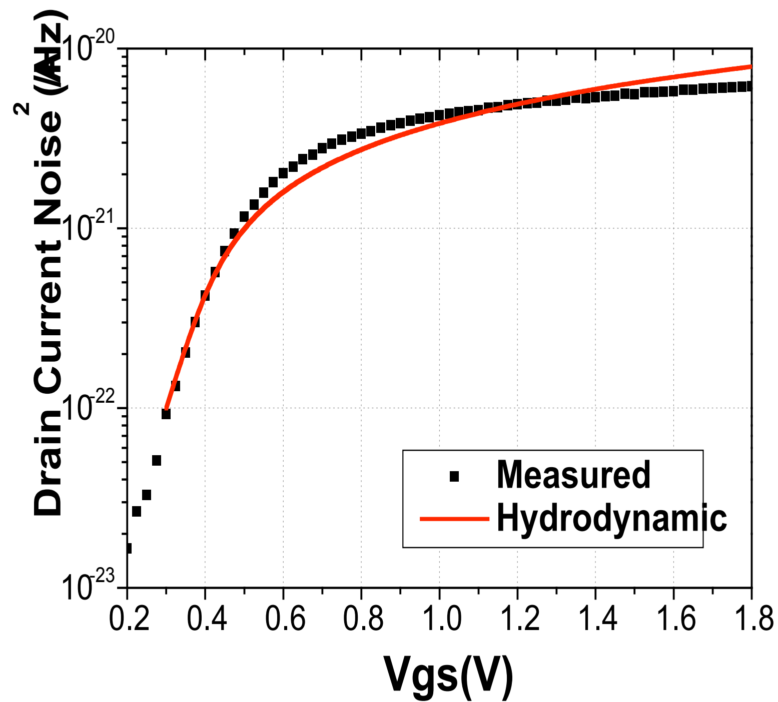


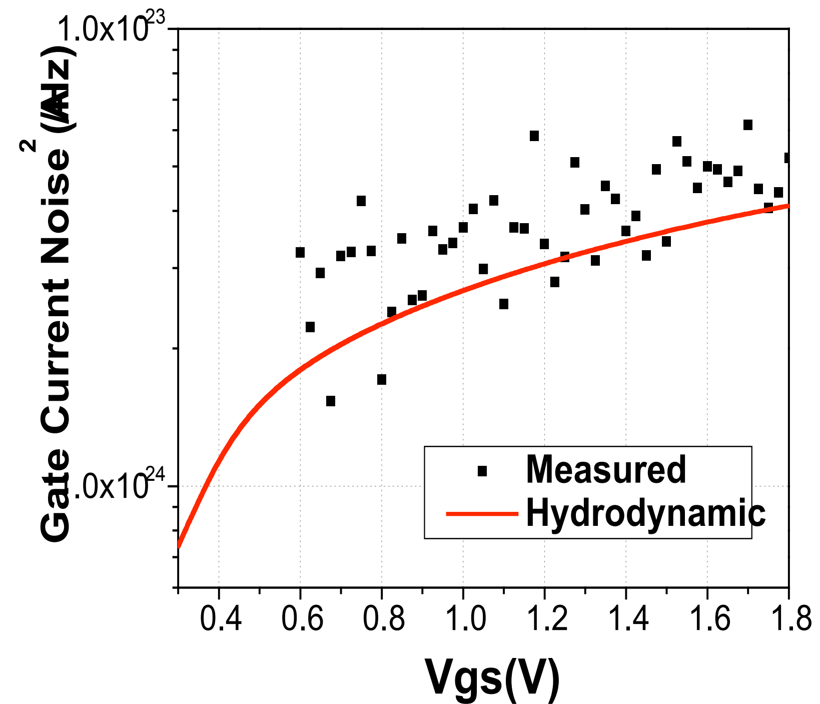
- Noise increases with reduced L
- Modeling of Carrier Transport very important

Classical Limit
(Van der Ziel)

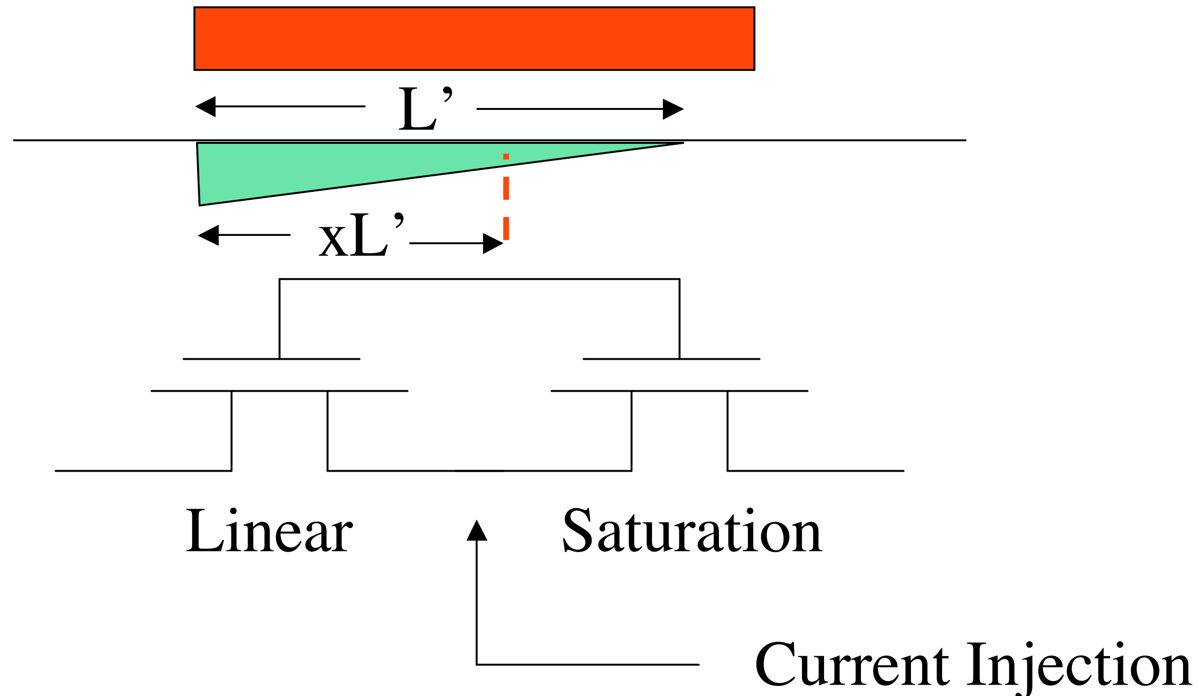
Measurement vs. Modeling--0.18μm

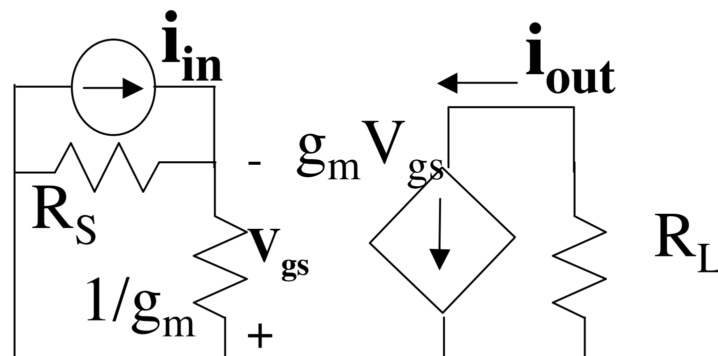
0.18 μm Channel Length MOS--Measurement and Simulation at 5 GHz (Data from Philips (Scholten, IEDM 02))





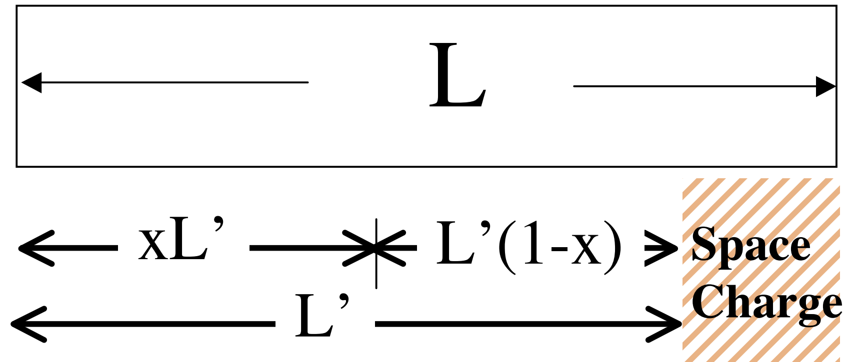
Two-Lump Impedance Field Model





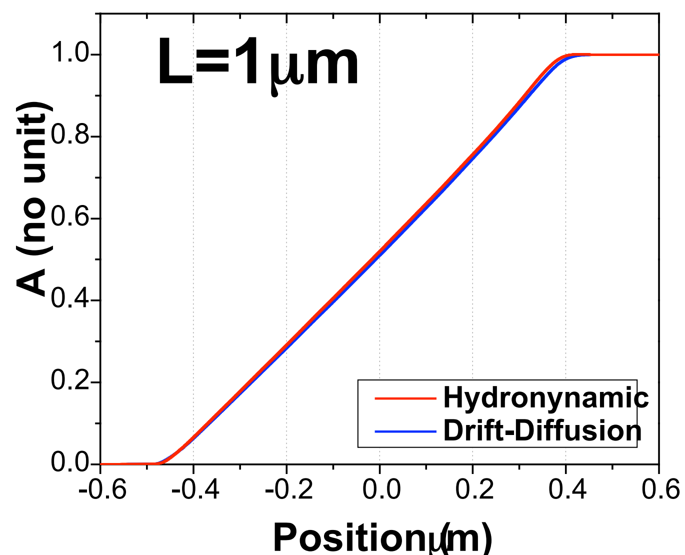
$$S_{i_d} = \left|\nabla A_k\right|^2 S_{i_n}$$

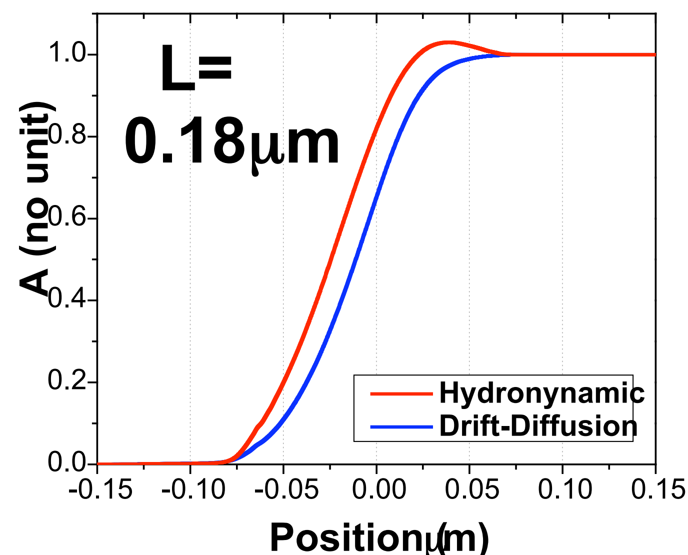
First-Order View of Impedance Field



$$A(x) = \frac{R_S}{R_S + \frac{1}{g_m}} = \frac{\frac{x}{\sqrt{1-x}}}{\frac{x}{\sqrt{1-x}} + \sqrt{1-x}} \approx \frac{x}{x + (1-x)} \approx x$$

Comparison of A(x)--1μm vs 0.18μm

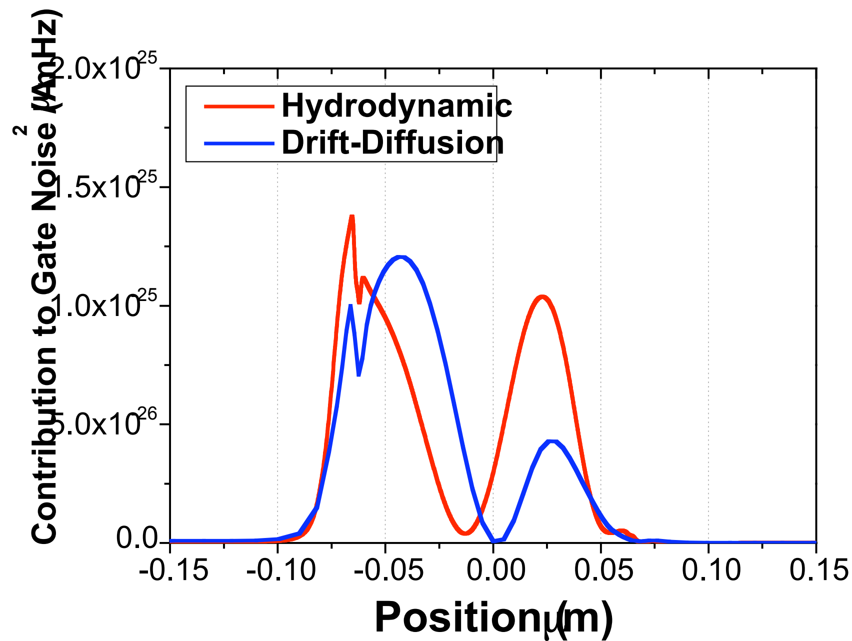


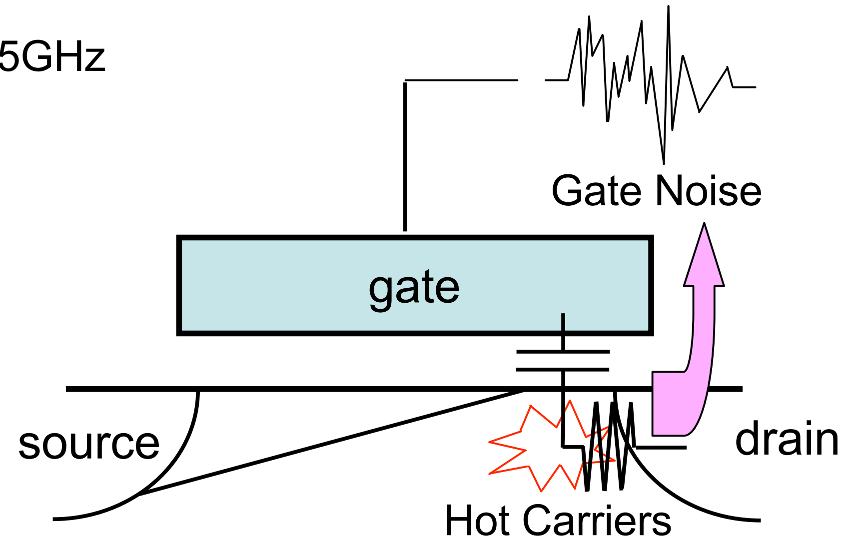


- Effects on noise in short-channel devices more pronounced
- Field-effects impact source-end noise, the most critical area base on impedance field analysis
- Clear differences between Drain- and Gate-current noise contributions

Gate Current Noise--Short Channel MOS

- **0.18 μm** channel length. Vds = 1.8 V f=5GHz





- Noise from high energy carriers has direct impact on gate noise in short channel MOS devices.

Self-Heating and Scaling of Silicon Nano-Transistors

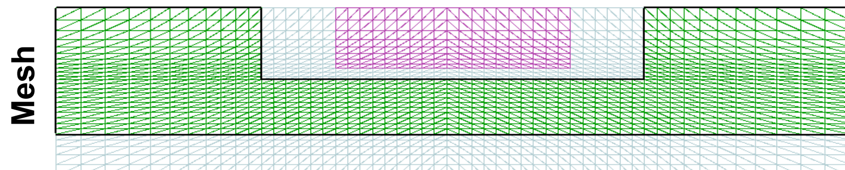
Eric Pop

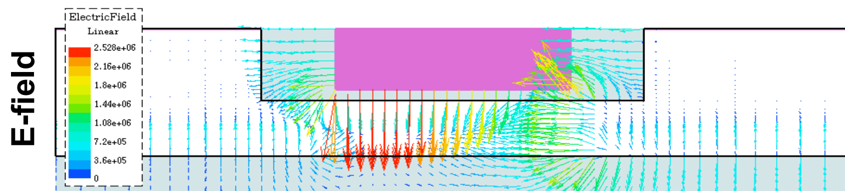
PhD Orals 2004

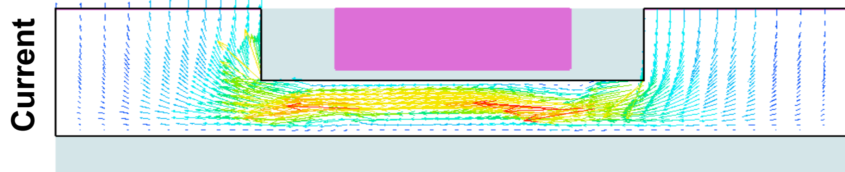
Center for Integrated Systems

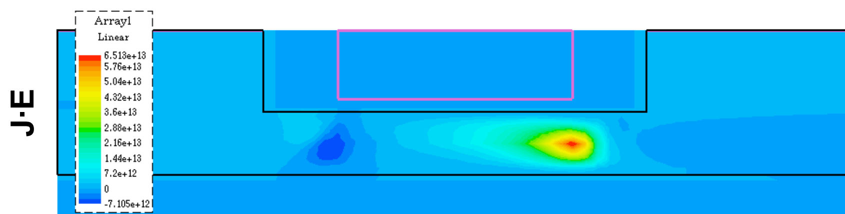
Stanford University

2-D: Thin Body SOI (L_g = 18 nm)







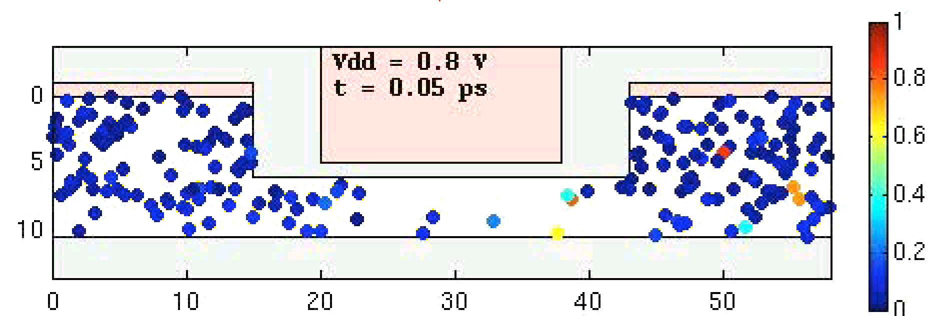


Engineer to ITRS Specs:

L_G=18 nm, t_{SI}=4.5 nm, t_{OX}=1 nm
N_{SD}=1e20 cm^{-3}, N_{CH}=1e15 cm^{-3}
I_{ON}=1000 μA/μm, I_{OFF}=1 μA/μm
Φ_{GATE}=4.53 eV (Mo), V_{DD}=0.8 V

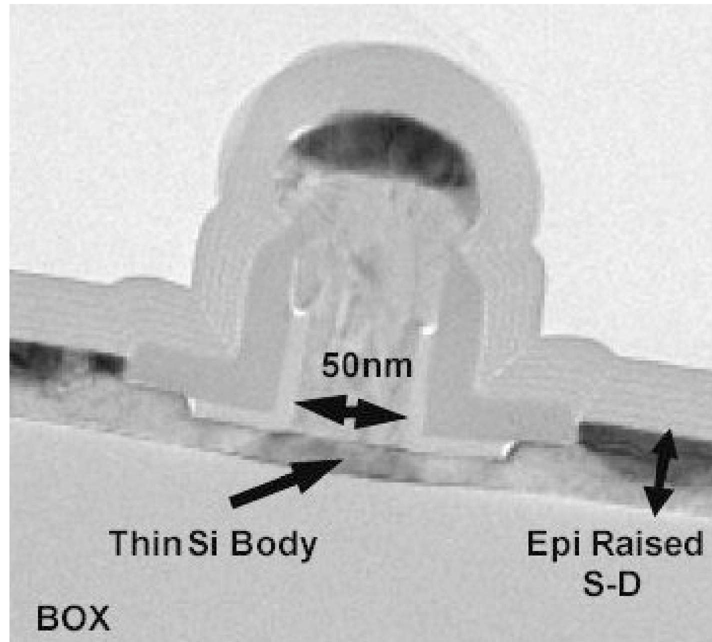
if W/L = 4 then N_{elec} ~ 2500 total!

MONET* **(no Poisson)**

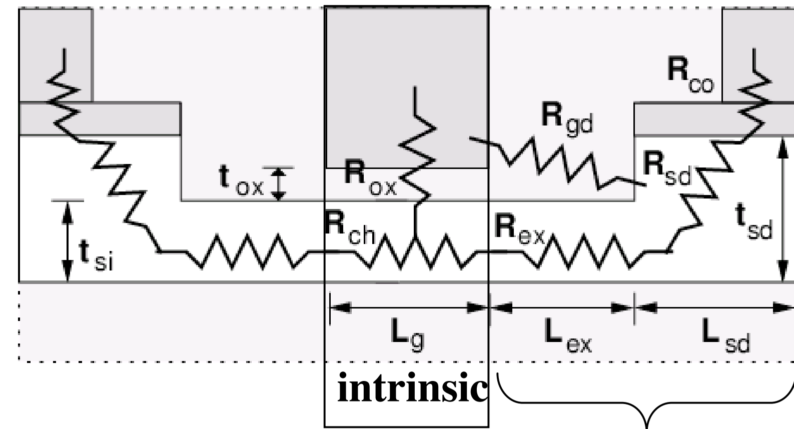


***nanoheat.stanford.edu**

Ultra-Thin Body SOI Scaling



Intel DST/SOI Transistor (*IEDM 2001*)



extrinsic:
link-up (L_{ex})
elevated S/D (R_{sd})

$R_{ox} = (R_i + t_{ox}/k_{ox})/A$

$R_i = 2 \times 10^{-8}$ m²K/W

$R_{co} = 6.75 \times 10^{-9}$ m²K/W

other → $R = L/(kA)$

baseline scaling:

$L_{ex} \sim L_g/2$
$L_{sd} \sim L_g$
$t_{si} \sim L_g/4$
$t_{sd} \sim 2t_{si}$
$W \sim 3L_g$
$A_{co} \sim 2L_g \times L_g$

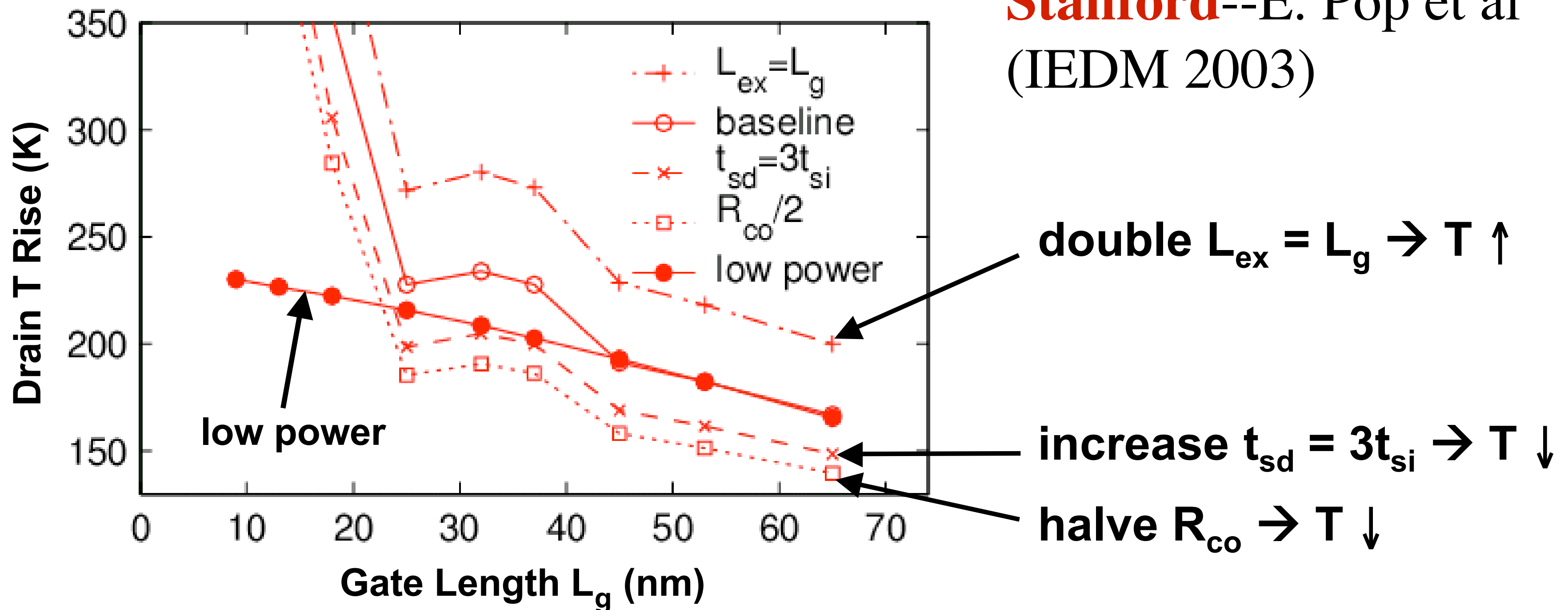
other parameters:

(I_{on}, V_{dd}, t_{ox}) from ITRS

Eric Pop, Stanford PhD 2004

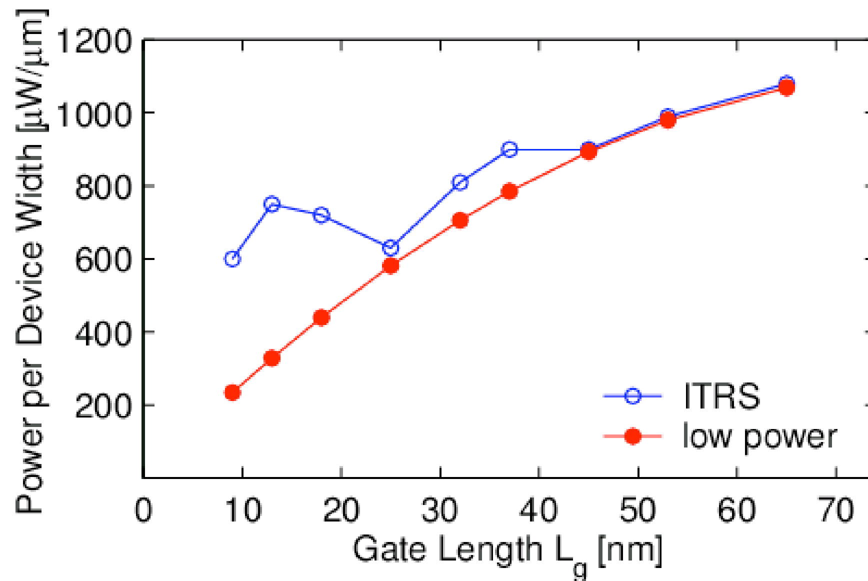
Thermal Modeling for SOI

Stanford--E. Pop et al (IEDM 2003)



- "Baseline" (ITRS power scaling): $t_{sd}=2t_{si}$, $L_{ex}=L_g/2$
- "Low power" case uses proposed quadratic power scaling guidelines--reduce power, consistent with volume

Proposed Power (I·V) Scaling



Proposed quadratic power scaling

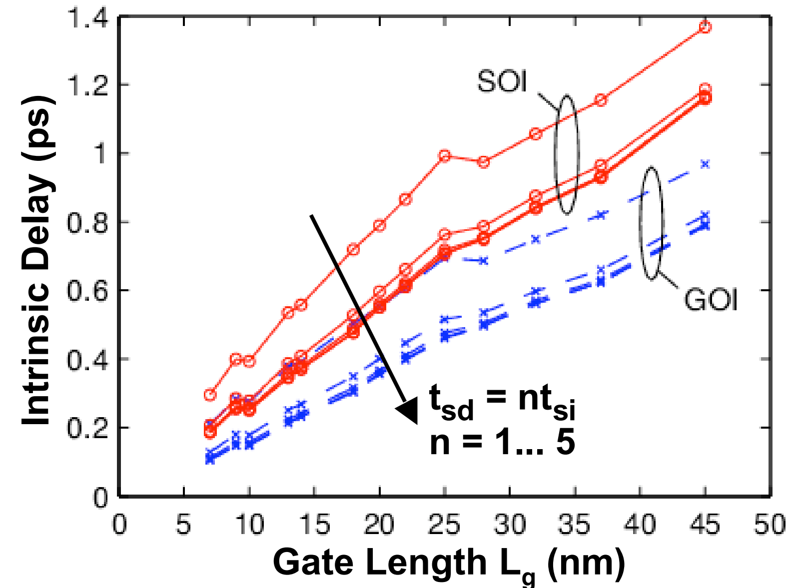
$$Q = I \cdot V = -0.17L_g^2 + 27.5L_g$$

(L_g in nm)

- ITRS power scaling non-uniform
- Quadratic scales power closer with device dimension (and volume) scaling
- Device temperatures near-isothermal
- 250 W/m power budget → V_{dd} = 0.25 V @ I_{on} = 1000 A/m

SOI-GOI Gate-Delay Comparisons

Stanford--E. Pop et al
(IEDM 2004)

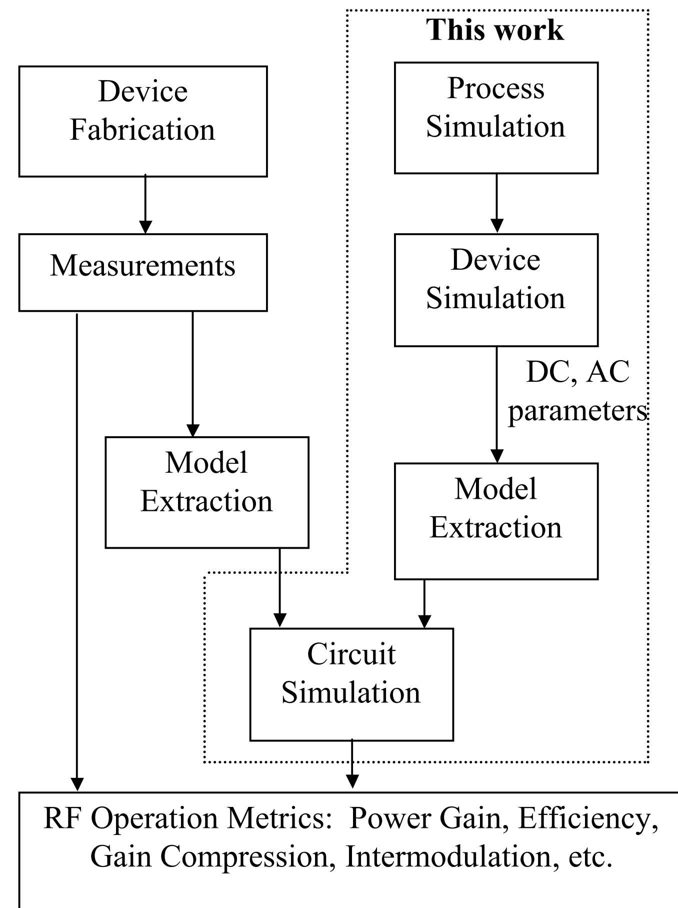


Bottom-Line:
GOI can achieve lower delays; improved drive current results in lower heat generation

- Ge-O-I → assume $t_{Ge} = 3/4t_{Si}$ where $t_{Si} = L_g/4$
- Si more k_{thin} reduction due to larger phonon mean free path
- Ge has 2x mobility advantage, 40% lower V_{dd}
- Delay not lowered for S/D raised beyond ~ 3 x t_{film}

Proposed Methodology*
(RF Power MOS Devices)

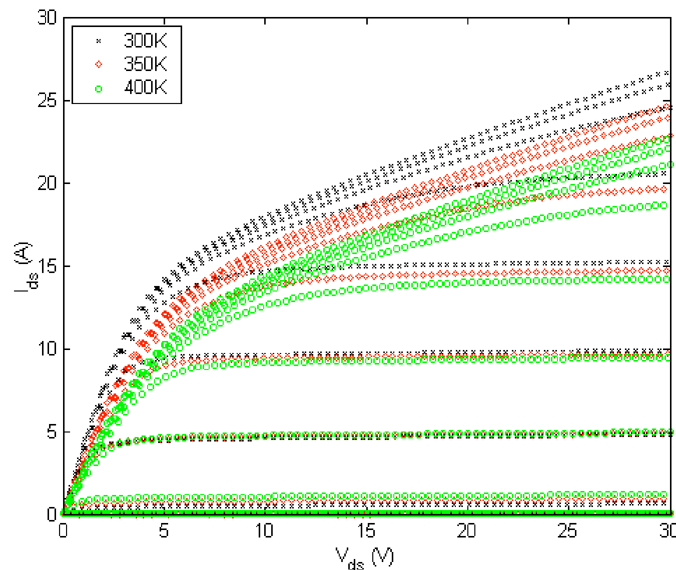
- Minimize fabrication-characterization time by using device simulation
- Generate table-based model from device simulation for use in circuit simulation
- Resulting model:
  - "Black box" model
  - Easy enough to generate for quick evaluation of design changes

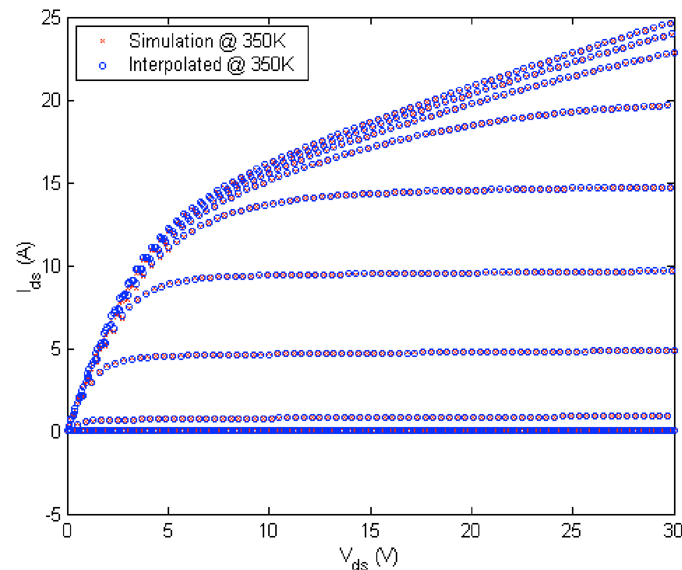


*C. Ito et al, Title, SISPAD 2004

Thermal Characteristics
(Simulation Results)

- RF power amplifiers require temperature-dependent models
- Implemented using several tables extracted at different temperatures
- Characteristics for arbitrary temperatures may be linearly interpolated between tables

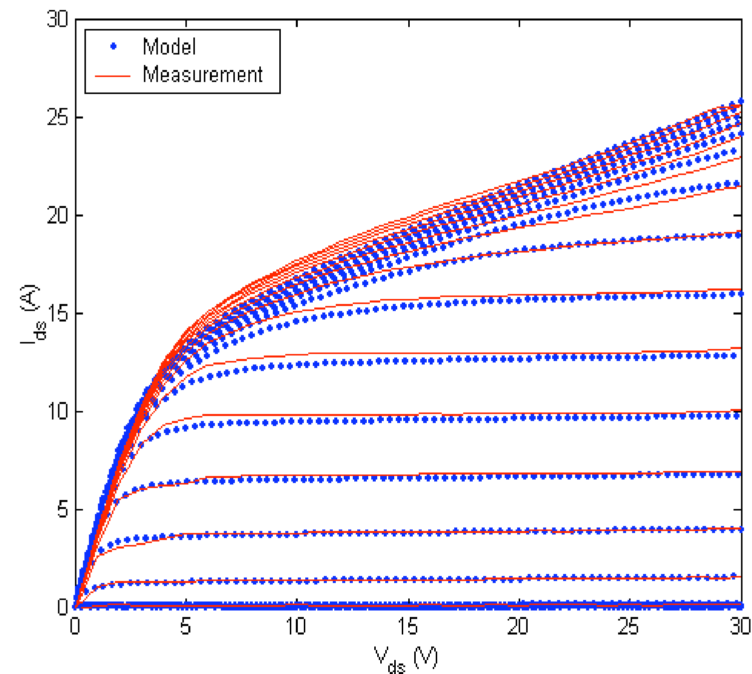




DC Characteristics

(Model vs. Measurements)*

- Isothermal (pulsed) current-voltage (IV) measurements and simulations
- IV characteristics match well between model and measurement
- Small discrepancy possibly due to differences between simulation doping profile and actual device profile



**Infineon Device*

Methodology for Modeling

- Compact models are becoming more abundant and complex
- Implementation is becoming bottleneck and potential weak link
- Cross-platform model portability is important to technology deployment

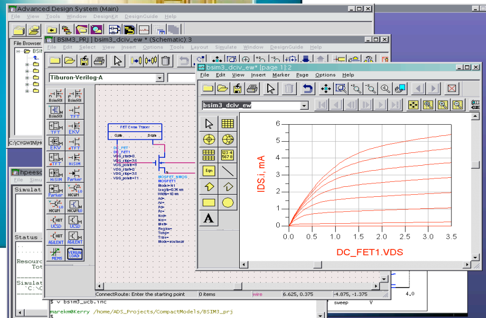
- **HLD Languages (Verilog/VHDL…) offer powerful solution**

Models Implemented in Verilog-A

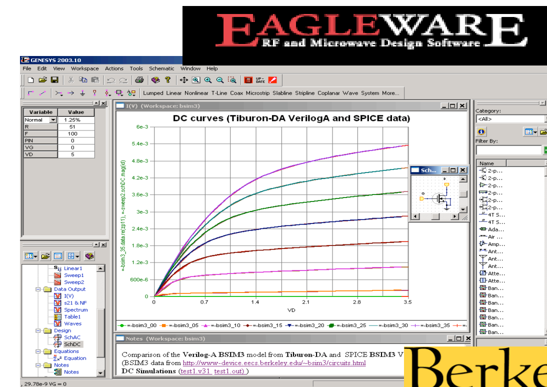
- BSIM3, BSIM4.3, BSIMSOI
- HiSIM, Shur-RPI TFT, EKV
- HICUM
- SPICE Gummel-Poon, Diode, JFET
- Philips MEXTRAM, MOS 9, MOS 11
- Triquent, Curtice, Parker-Skellern, Raytheon-Statz, Angelov
- UCSD (Aspeck) GaAs HBT

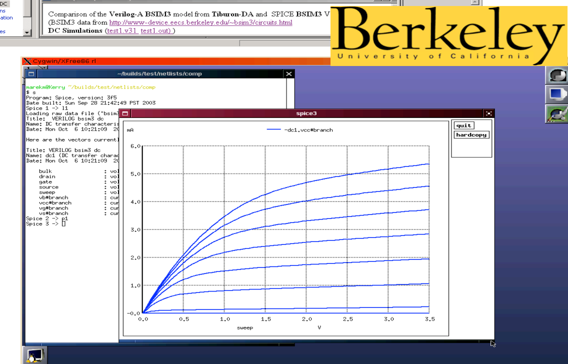
Multiple Simulators, Single Model

- **Compiled Verilog-A devices can be shared among diverse simulators**
- **Same compiled object file linked to each simulator**
- **Develop in one simulator, same results in all simulators**



BSIM3.cml





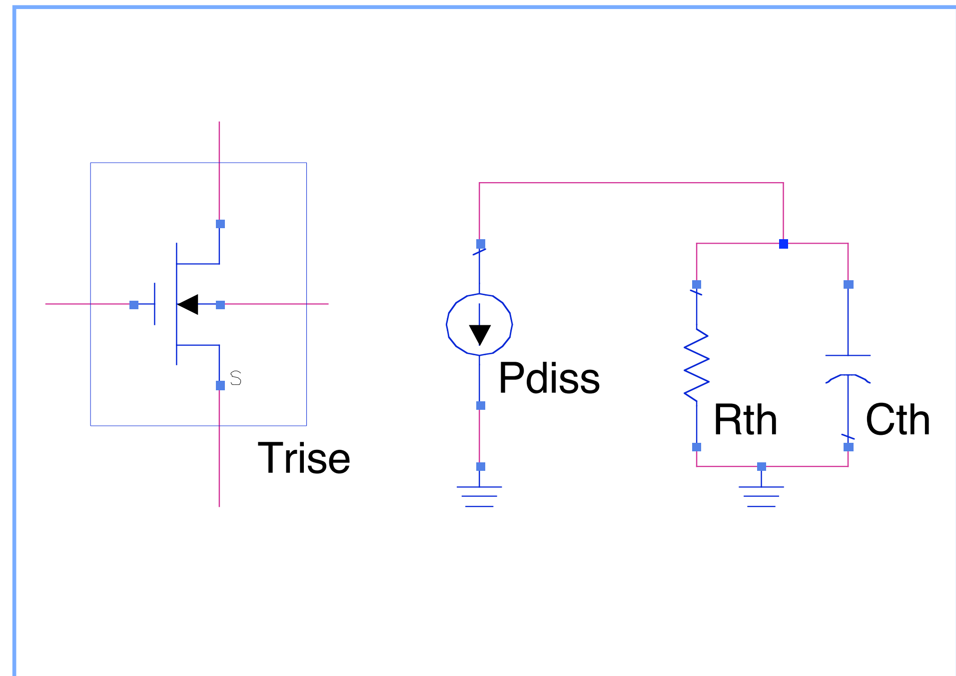
Tiburon Design Automation

Example: Adding Self-heating to BSIM3

- Adding a thermal circuit to a compact model can be a complicated project, as all of the derivatives with respect to the temperature must be provided to the simulator.
- Since Verilog-A does that for the developer, the additional lines of code are minimal, typically just a few dozen, including the parameter definition.

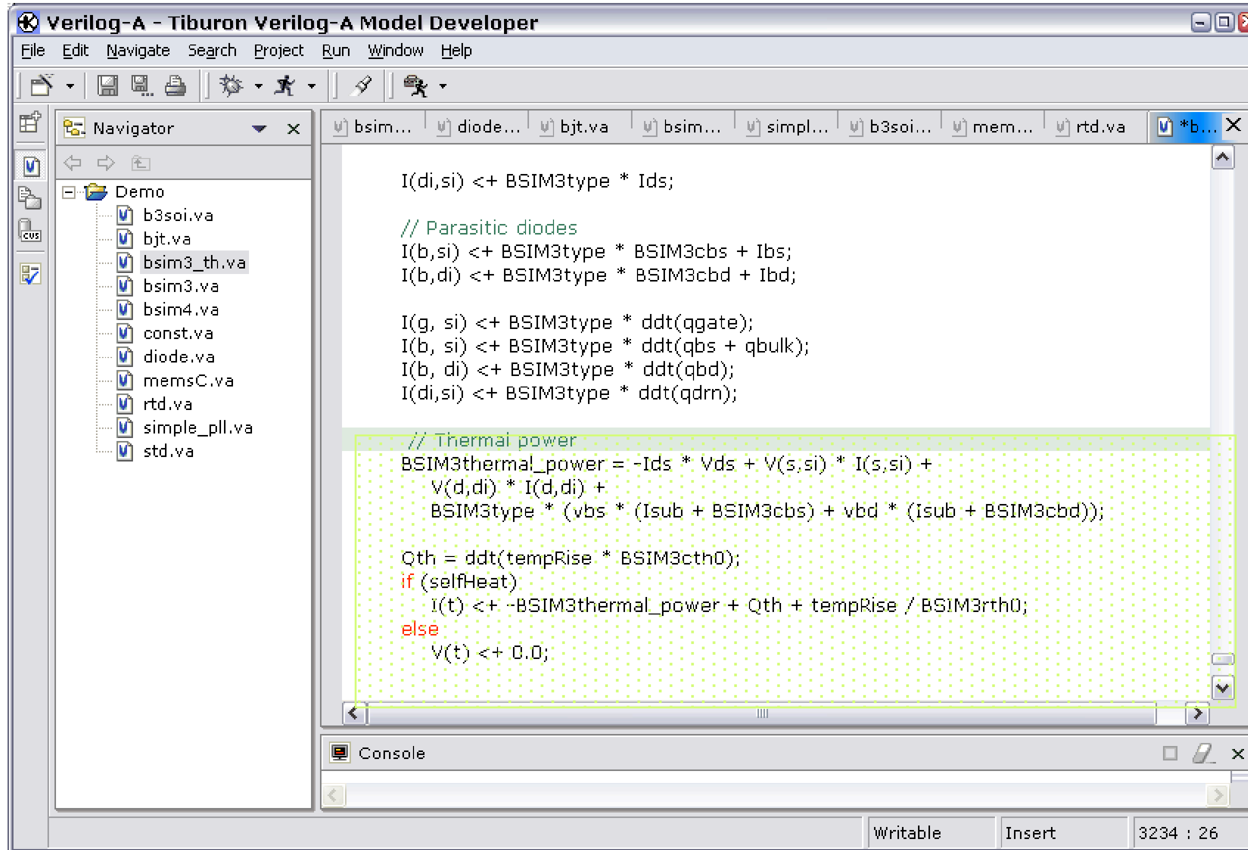
Electrical Model for Self-heating

- An R-C circuit is added to the intrinsic model
- The current source represents the power dissipated in the device
- The voltage across the thermal resistance and thermal capacitance represents the associated temperature rise
- This temperature rise is fed back to the device model



Verilog-A Implementation

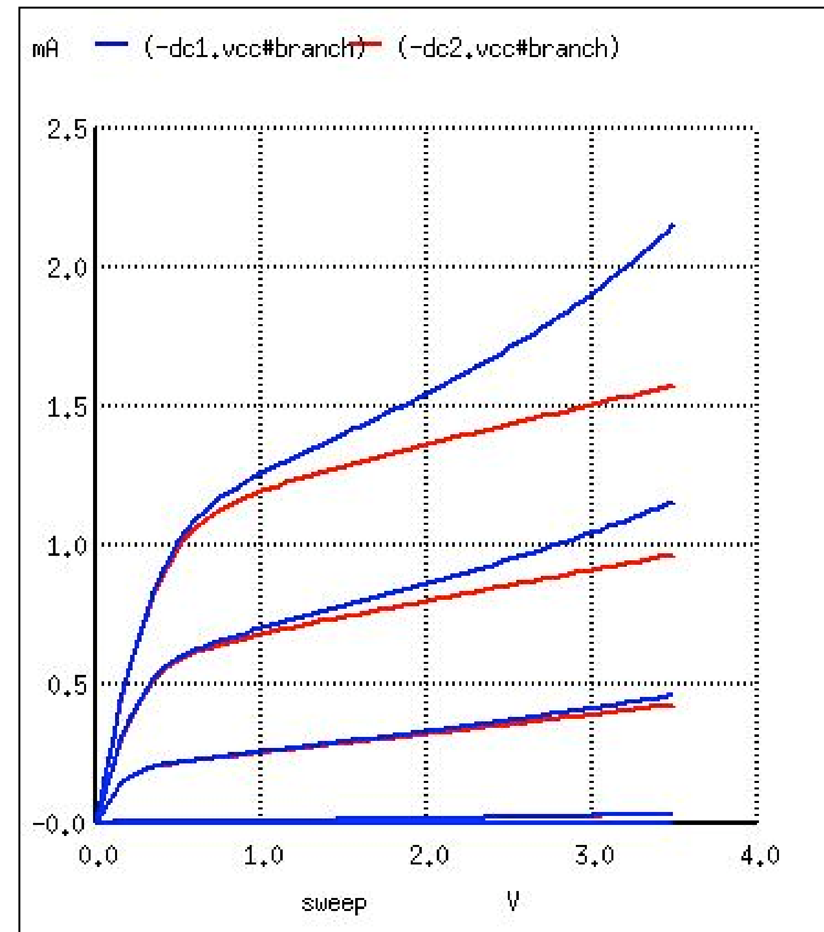
- The code change is minimal



Tiburon Design Automation

DC I-V Results

- The (standard) BSIM3 compared to the BSIM3 with self-heating



Tiburon Design Automation

Summary and Conclusions

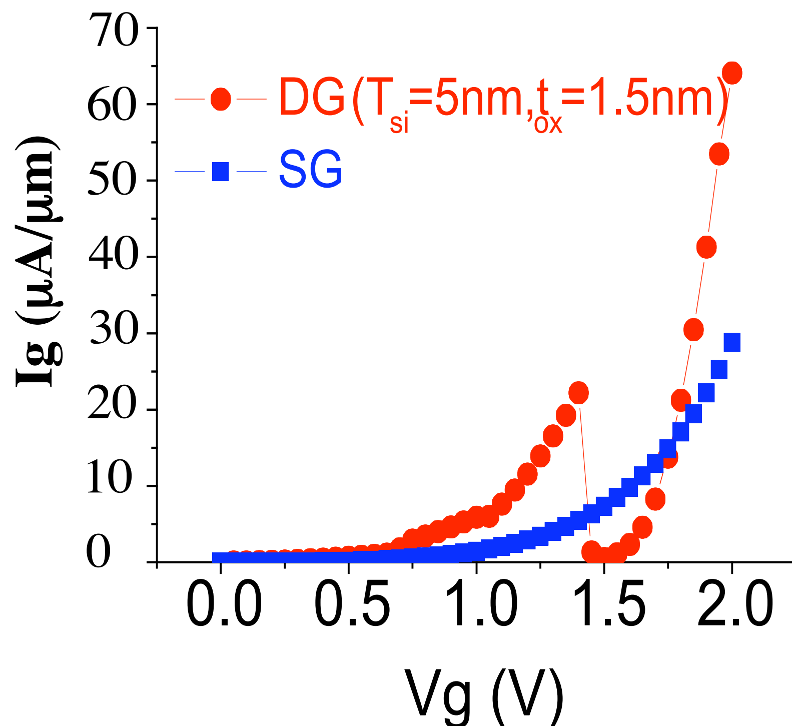
- Development of scaled-MOS compact models face many challenges: intrinsic and parasitic device effects and “other” challenges:
  - Intrinsic Limits--mobility, gate current, “leakage”
  - RF Issues--NQS, substrate effects, noise
  - Thermal Issues--new models, scaling laws
- Methodology for compact model deployment shifting to HLL approach and compilers; “hand-coded” models (for “efficiency”) inherently limit portability
- “We live in interesting times (still!)”

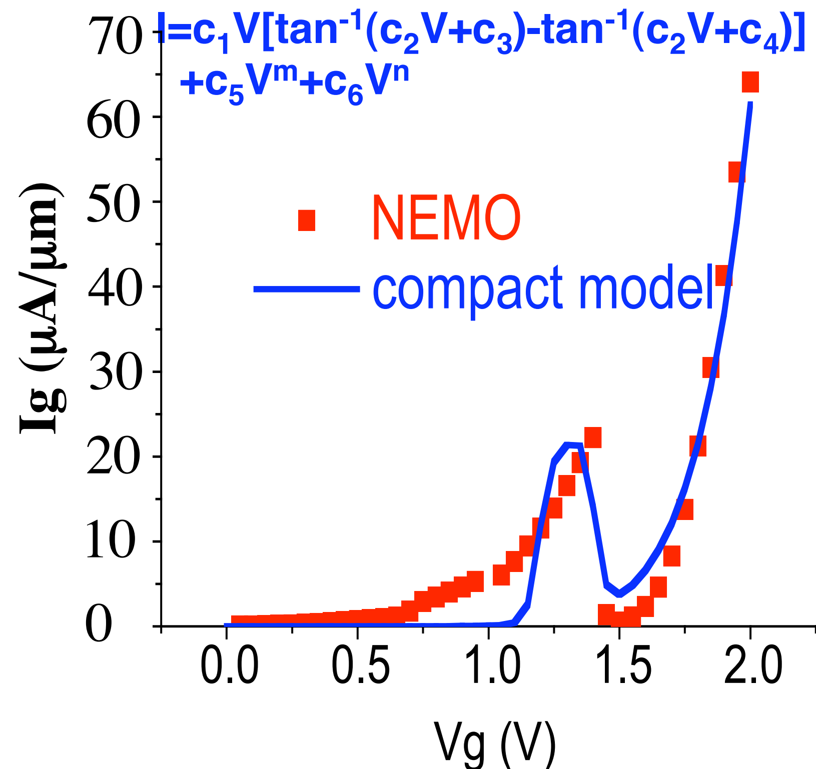
Back-up Slides

- Power RF Table-based model (Root)
- Silicon RTD--physics and Verilog-A model

Resonant Tunneling

- Higher gate tunneling in DG than SG (bulk) MOS
- Simulated resonant gate tunneling current for thin layer DG SOI by using NEMO: possibly Si-based resonant tunneling diodes (RTD)

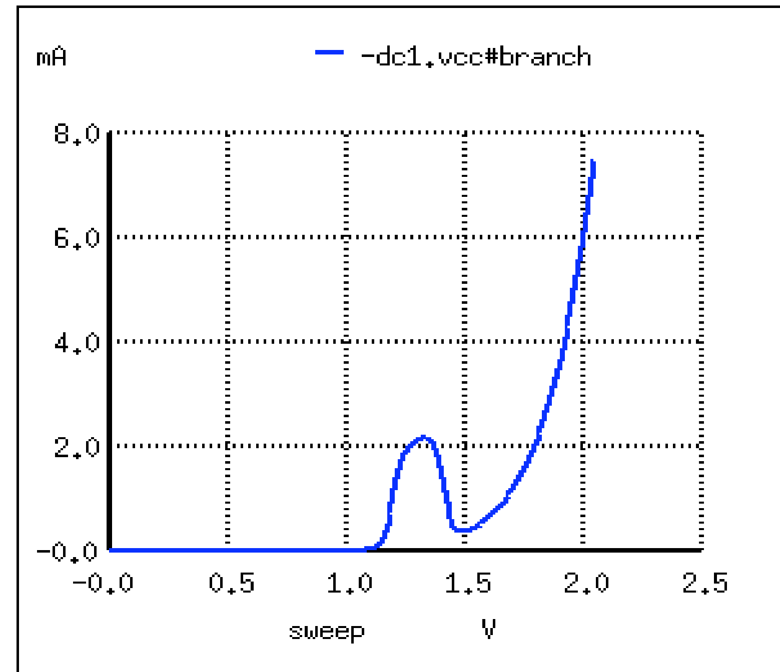




Advanced Device Development

- Verilog-A provides a simple way to implement new devices
- Resonant tunneling diode

```
`include "disciplines.vams"

module rtd(anode,cathode);

   inout anode, cathode;
   electrical anode, cathode;

   parameter real c1=8.0e-4;
   parameter real c2=23.5;
   parameter real c3=-28.0;
   parameter real c4=-33.0;
   parameter real c5=5.0e-7;
   parameter real c6=6.0e-6;
   parameter real m=1.0;
   parameter real n=10.0;

   real v;

   analog begin

     v=V(anode,cathode);
     I(anode, cathode) <+ c1*v*(tanh(c2*v+c3) - tanh(c2*v+c4)) +
        c5*pow(v,m) + c6*pow(v,n);

   end

endmodule
```



Tiburon Design Automation

Table-Based Models

- Table lookup + interpolation
- Separate models can easily be generated for different designs
  - Possible due to efficient model generation
- Root Model
  - Consists of 5 variables over the V_g-V_d space
  - I_{gs}, I_{ds}, Q_g, Q_d, I_{dh}

$$\nabla Q_G = \frac{\mathrm{Im}\{y_{11}\}}{\omega}\overrightarrow{v_{gs}} + \frac{\mathrm{Im}\{y_{12}\}}{\omega}\overrightarrow{v_{ds}}$$

$$\nabla Q_D = \frac{\mathrm{Im}\{y_{21}\}}{\omega}\overrightarrow{v_{gs}} + \frac{\mathrm{Im}\{y_{22}\}}{\omega}\overrightarrow{v_{ds}}$$

$$\nabla I_D^h = \mathrm{Re}\{y_{21}\}\overrightarrow{v_{gs}} + \mathrm{Re}\{y_{22}\}\overrightarrow{v_{ds}}$$