Lecture 3

Transistor Models

Mark Horowitz
Computer Systems Laboratory
Stanford University
horowitz@stanford.edu

Overview

Reading

Chen	Predicting CMOS Speed
Pelgrom	Transistor Matching
Lovett	Transistor Matching

Introduction

Transistors are not all the same, we need to have some model of the variations too. This comes both in local variations (matching), and run to run variations. In this lecture we will briefly review the transistor models, and spend most of the lecture talking about device variations.

There is also a struggle in doing simulations: on the one hand you would like the model to be accurate, and yet you need to understand the model to be able to reason about it and predict the results. This leads to a number of strategies of building simple models first and then building up the complete model. To do this it is helpful to 'calibrate' a technology, and use these simple models to help you reason about the technology, and circuit.

MOS Device Behavior

Assume you know MOS device issues from EE313

- Briefly review what the most important issues are:
  - Look at iV and CV curves
  - Look in Hodges Jackson if you need more information

- For iV curves need to understand:
  - basic shape, threshold voltage, mobility effects, velocity saturation, subthreshold leakage, scaling (and variations in these parameters)

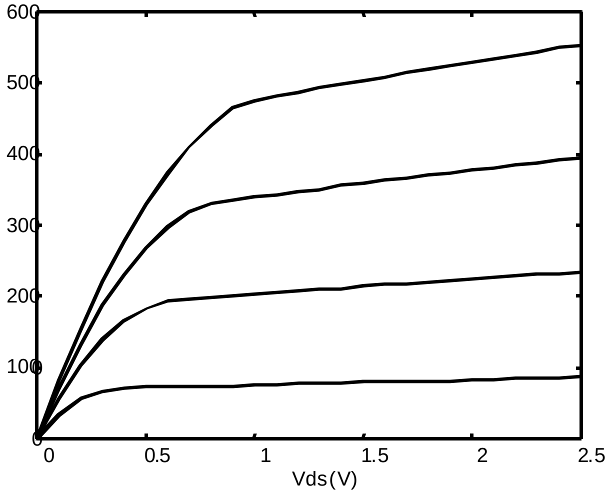
- For CV curves need to understand:
  - overlap capacitance, channel charge, junction capacitance

Basic Shape

Two plots are used, one vs. Vds, and one vs. Vgs

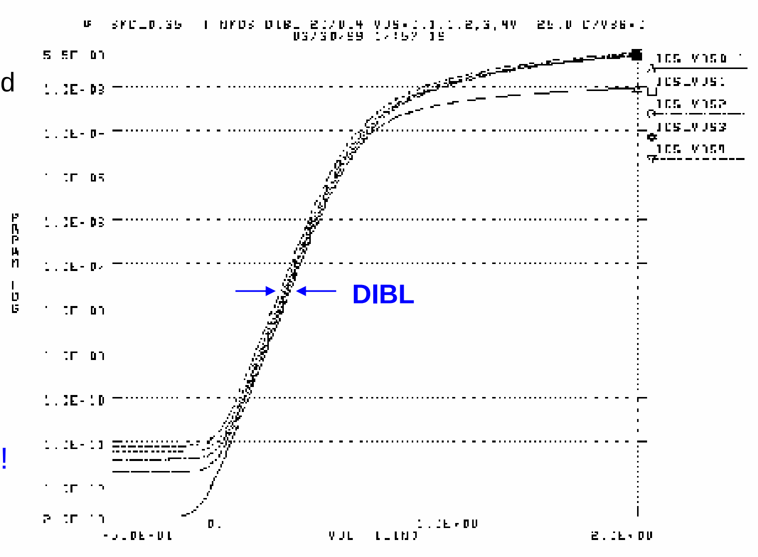
- Ids - Vds plot
  - Two regions
    - Linear (low Vds)
    - Saturated (high Vds)

- Linear region
  - Effective Resistance
- Saturated region
  - Current
  - g_m
  - g_{ds}



Basic Shape

- Ids vs. Vgs
  - Linear Ids
    - Used above threshold
    - Current vs Vgs
    - gm
  - Log (Ids)
    - Use below threshold
    - Subthreshold slope
    - Threshold voltage
- Also good to look at Ig
  - Gate tunneling current!



Mobility

- Has a strong temperature dependence:

$$\mu = \mu_o \left(\frac{T}{T_o}\right)^{-3/2}$$

  - Temp change from 27º to 130º decreases current to 0.65.
  - The circuit will run 1.6 time slower.
- Also decreases with high vertical field, and channel doping
  - New models say it is completely set by vertical field

$$\mu_e(V_{gs}, V_{th}, T_{ox}) = \frac{540}{1 + \left(\frac{V_{gs} + V_{th}}{0.54T_{ox}}\right)^{1.85}}$$

μ in cm^2/Vsec, Tox in nm

For the second equation, (Vgs+Vth) term may be counter intuitive To explain this, one can go though a quite simple derivation where Eff is proportional to Qb + 0.5 Qinv, Qinv is proportional to Vgs-Vt and Qb is proportional to Vt with the same proportional constant

Velocity Saturation

Mobility, μ, relates carrier velocity to electric field,

Relationship is not linear, max velocity is around 8×10^6 cm/s

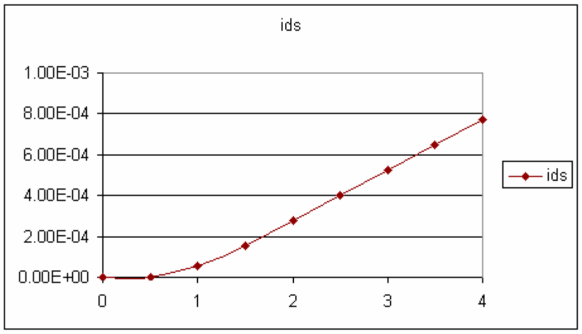
Approx i_{dsat}

$$i_{dsat} = \frac{W\upsilon_{sat}C_{ox}(V_{gs} - V_{th})^2}{V_{gs} - V_{th} + \frac{2\upsilon_{sat}L}{\mu_{eff}}}$$

Equation works in both limits.

When completely velocity saturated:

$i_{ds} = W\,C_{ox}\,(V_{gs} - V_{th})\,\nu_{max}$
which is independent of L



ids for a 0.35μ technology

ν_{max} L term around 2.3V

Subthreshold Conduction

The threshold voltage is not a magic place

- Voltage where the channel charge is roughly equal to the doping
- Have channel charge when Vgs less than Vth
    - Feedback is not as strong
    - Gate directly controls Φ_s, not channel charge
    - Channel charge exponentially related to Φ_s

$$Ids = Is \times e^{\frac{Vgs - Vth}{\alpha Vt}}$$

Vt = kT/q = 26mV @RT

Is depends on definition of Vth, around 0.3μA/μ

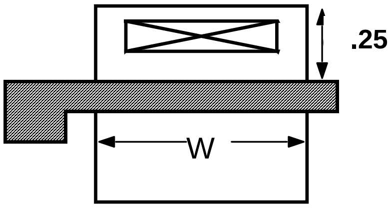
α is from cap voltage divider, around 1.3

MOS Capacitance

There is a lot more to worry about than channel capacitance

For 0.1μ technology, 2.5nm gate oxide

- Cox = 14 fF/μ^2 = 1.4fF/μ width
- Gate overlap cap ~ 0.35 fF/μ (per edge)
- Diffusion cap
  - 1.5 fF/μ^2 bottom plate
  - 0.2 fF/μ sidewall
- Total
  - C_{gate} = 1.4 W
  - $C_{overlap}$ = 0.7 W
  - C_{bot} = 0.4W
  - C_{side} = 0.4W + .small



MOS Scaling

With high fields, quadratic model is not very accurate for estimating scaling effects. A better model is:

$$I_{dsat} = K\,W\,L_{eff}^{-0.5} T_{ox}^{-0.8} (V_{gs} - V_{th})^{1.25}$$

If L,T_{ox} V all scale (note V scaling will be limited by V_{th} scaling)

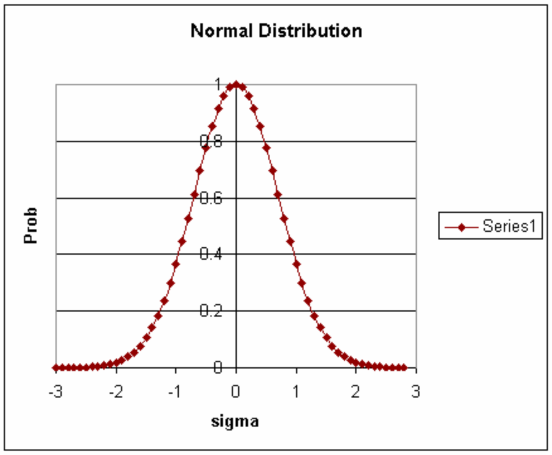
- Current should remains constant per micron.
- It will be 0.6 to 0.8mA/μ
- Current for a scaled transistor scales down by α
- Voltage scaled down by α
- Capacitance scales down by α
- $\Delta t = CV/i = \alpha\,\Delta t$
  - This assumes that Vth continues to scale, and this is not likely to happen

Parameter Variations

Talking about transistors like all transistors are the same

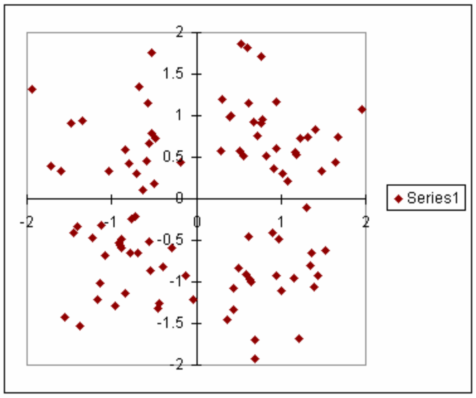
Not true -- no two are exactly the same

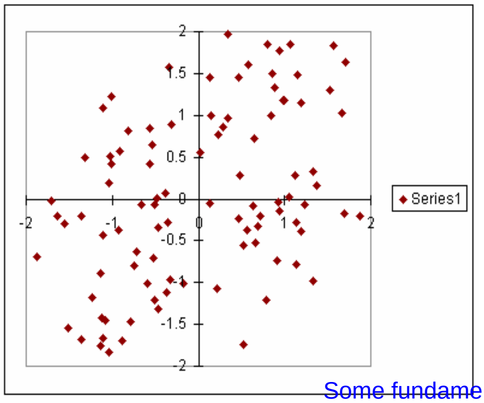
Parameters of a fabrication run are generally normally distributed



Variations

There are really many parameters:





Some fundamental physical parameters affect both transistor types (Tox, print/etch) and hence are the reasons that n/p can be correlated.

Sometimes they are correlated, and sometimes they aren't

- Corners are the extreme points on the distributions
- For correlated parameters corners are pessimistic

Process Corners

Extreme points in parameter distributions (spec limits)

- Probably stress circuits at these points
- Good place to test your design

Corners:

- Process parameters:
  - Poly linewidth, nMOS Vt, pMOS Vt, Tox,
  - metal width, oxide thickness
- Operating conditions
  - Temp (0-100 die temp)
  - Operating voltage (die voltage)

EE371 Corners

Group parameters into transistor, and operating effects

- nMOS can be slow, typ, fast
- pMOS can be slow, typ, fast
- Vdd can be high, low
- Temp can be hot, cold

Use library file to get models

- This file sets up the temp, device models parasitics and voltage
- file is `/usr/class/ee371/lib/opConditions.lib`
- Contains the following conditions

Library

Label is nMOS, pMOS, Temp, Voltage

TTTT = typ nMOS, typ pMOS, room temp, nominal supply
SSSS = slow nMOS, slow pMOS, hot temp, low supply
FSSS = fast nMOS, slow pMOS, hot temp, low supply

Note that wires are not included. We will either provide a separate wire model, or you will have to create your own.

The temp and voltages are the levels that the die sees, and this is worse then the package spec.

w.c. temp 110oC, w.c. voltage +10%, -15%

What To Look For

TTSS	Must meet timing specification for part
SSSS	Sometime need to hit spec here too, signal collapse
FFFF	Power issues, edge rates on IO outputs, pulse collapse
SFSS	Failure of ratio circuit, race conditions
FSSS	Failure of ratio circuits, race conditions

etc.

Style, cont'd

To use library file:

```
.lib '/usr/class/ee371/lib/opConditions.lib' SSSS
```

- This calls out the set of models for the slow simulation case
- Could have used any of the others.

```
.lib '/usr/class/ee371/lib/opConditions.lib' TTSS
```

Can use .alter command to run simulation in many cases

- Device models are always nmos, pmos

```
Mname      drain gate source substrate nmos
Mname      drain gate source substrate pmos
```

Philosophical Struggle

How to analyze circuits?

- Use your intuition and your pencil and paper analysis. These are things that you understand. SPICE is prone to Garbage In / Very Pretty Garbage Out. Need to understand circuit to check SPICE, and not vice versa.

- There is enormous complexity and ugly nonlinearity in VLSI circuity, making it potentially very difficulty to do hand analysis. Also competitive market pressure requires sophisticated circuitry (which need SPICE) on short schedules or you will be steam rolled by your competitors

So you end up doing both

Calibrating a Technology

When you get a set of models for a technology, it is a good idea to run some simple simulations to get a feeling for the technology. Here are some things that I like to run:

- Run iV curves for a few device sizes.

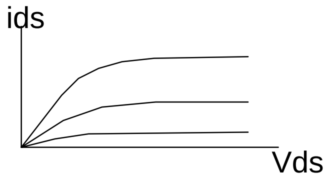
  Do the curves seem reasonable?

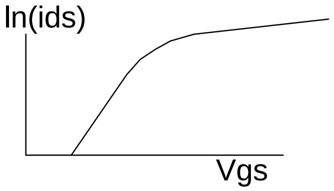
  What do they say about:

    velocity saturation, output conduction, Vbb sensitivity

    subthreshold conduction, Vth effects (DIBL, DVth(W,L))

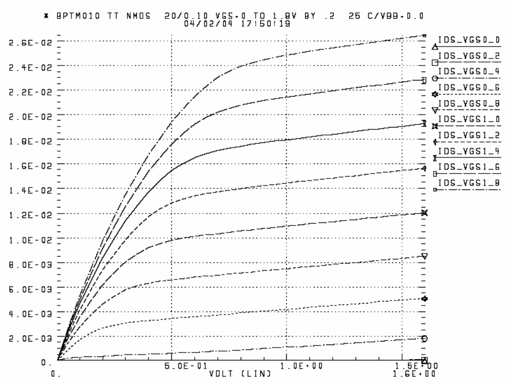
  To get a feeling for these effects, you will need to do a few simulations:

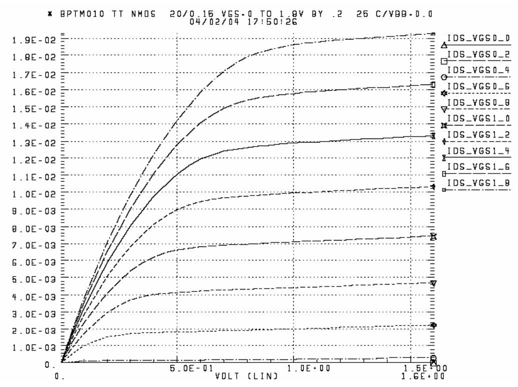




Ids vs. Vds

Look at different channel lengths (nMOS):

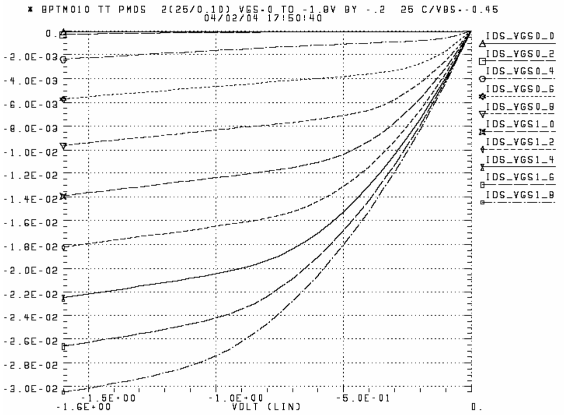


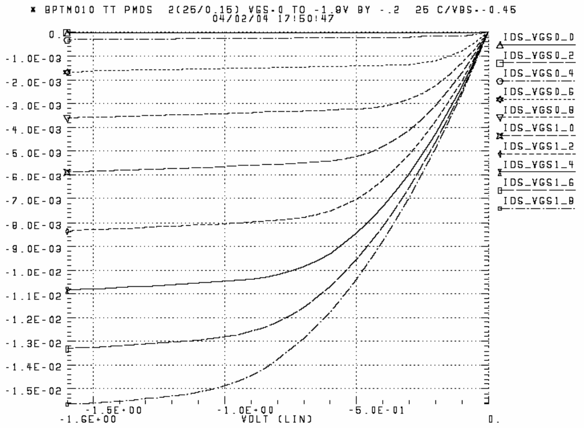


- Notice:
  - Difference in output slope
  - Linear gm in longer channel device

Ids vs. Vds

Look at different channel lengths (pMOS):

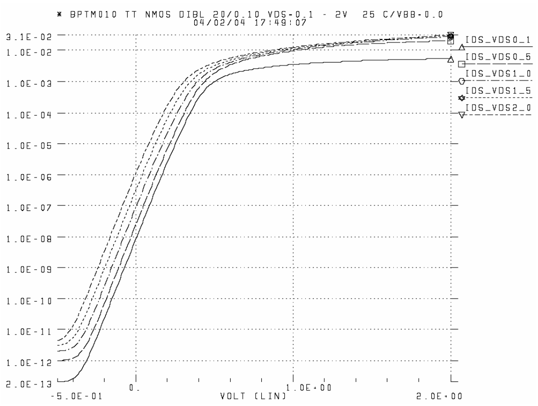


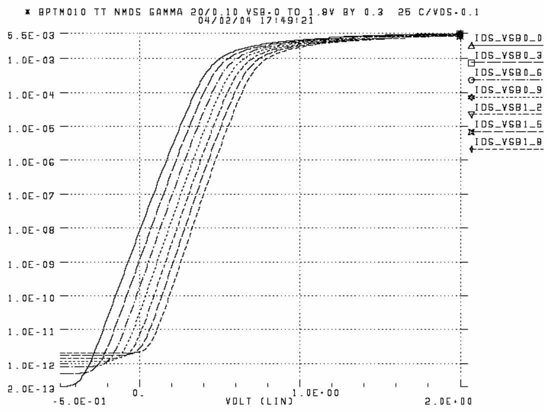


- Notice:
  - Difference in saturation voltage from nMOS
  - Linear gm in longer channel device, change in output slope

Ids vs. Vgs (nMOS)

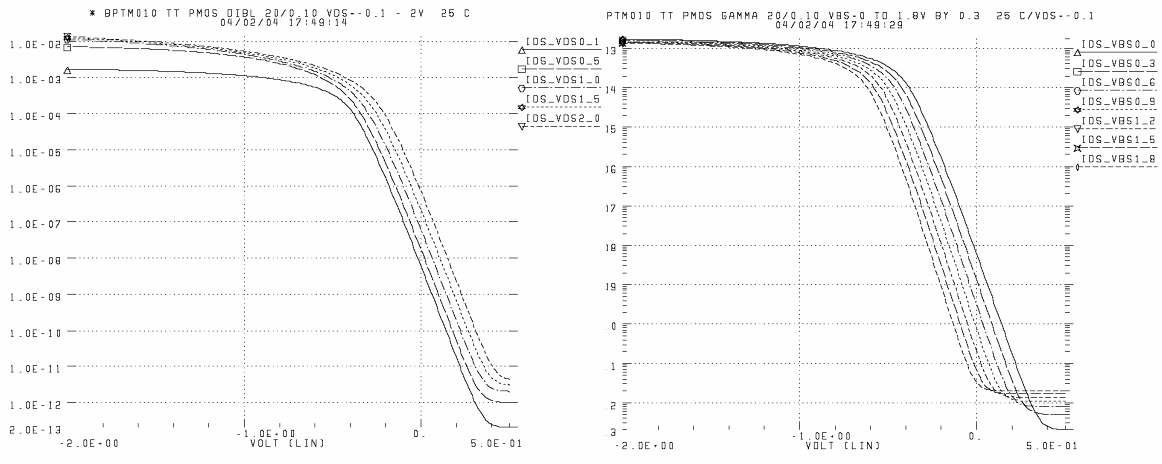
Look at Vds Vbs:





- One shows DIBL, and the other shows gamma:
  - DIBL is drain induced barrier lowering, it is when the voltage at the drain reduces the threshold voltage

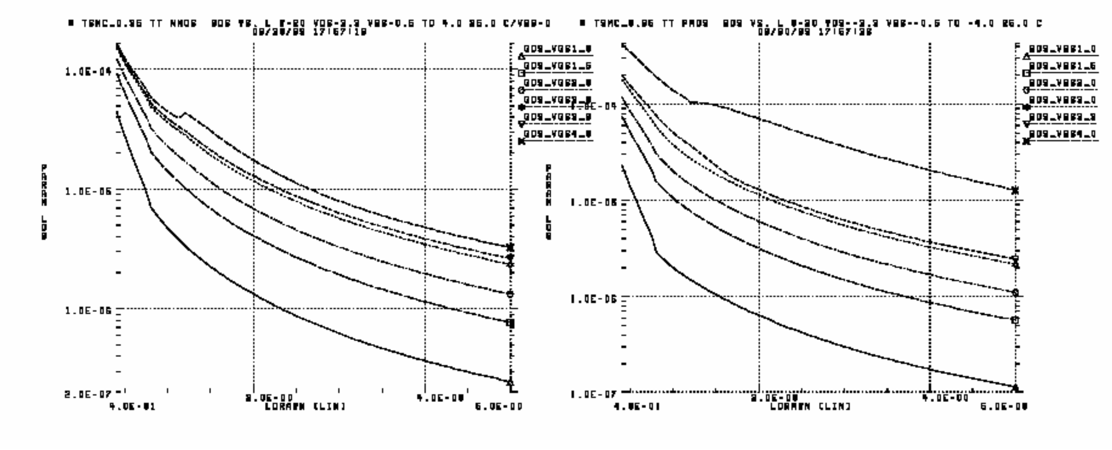
Ids vs. Vgs (pMOS)


Vds

Vbs:

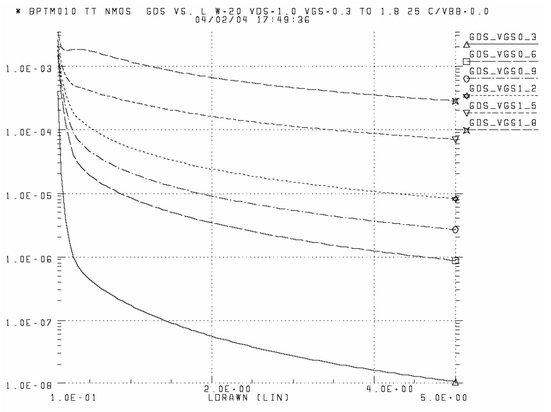
Gds vs. L (for .35μ)

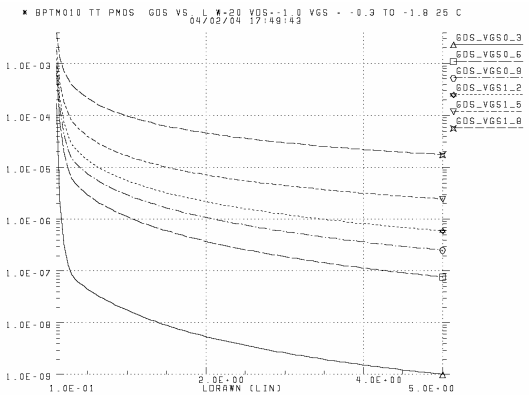
(shows effects of model binning)


nMOS

pMOS

Gds vs. L

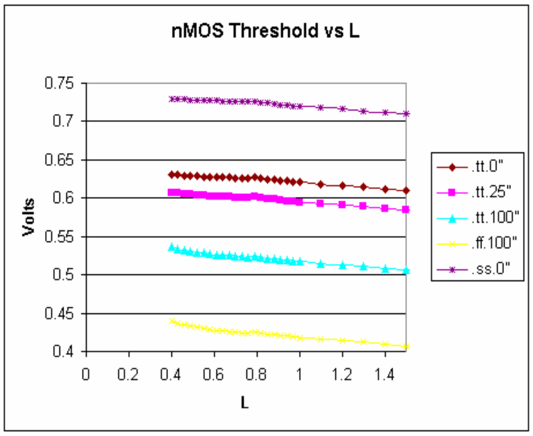


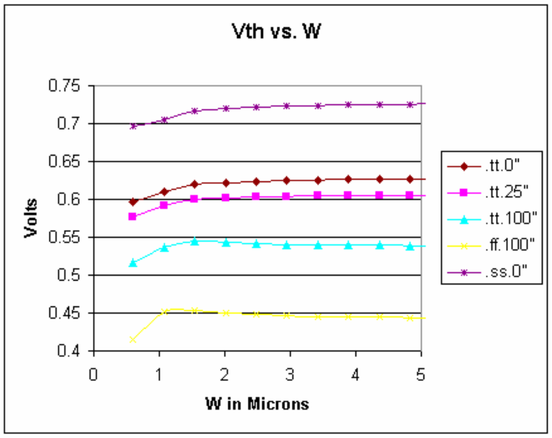


- Scale on sim run was wrong – Max L should be probably 1μ

Threshold Voltage nMOS (0.35μ)

Notice change with temp, channel length, width:

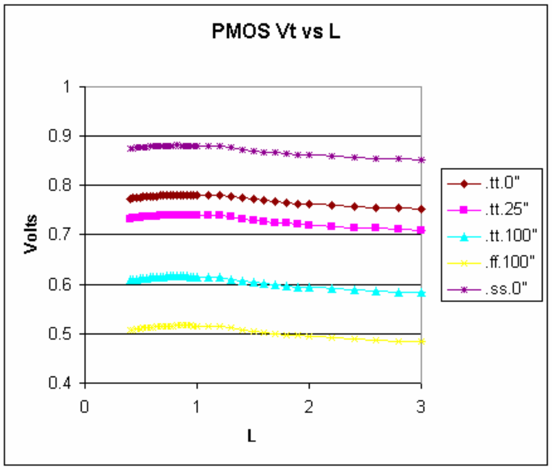


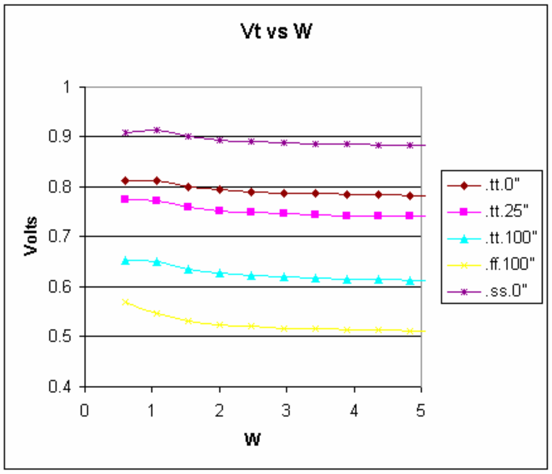


Vth vs. W dependency is affected by: isolation types, dopant segregation effects:

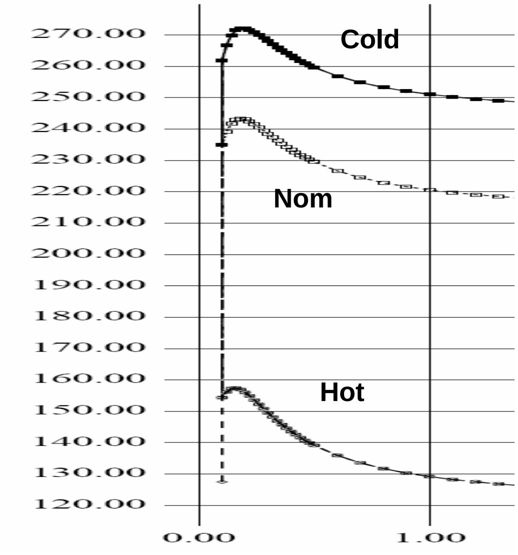
LOCOS (old) isolation scheme tends to see Vt rise as W shrinks, because of excess silicon region the gate has to deplete; Trench isolation tends to see Vt fall as W shrinks, due to higher effective Cgate/Width. Boron segregation into oxide tends to cause Vtn to decrease as W reduce and P/As pile up in silicon tends to cause Vtp to increase as W increase. Overall effect is an interaction of all the mechanism.

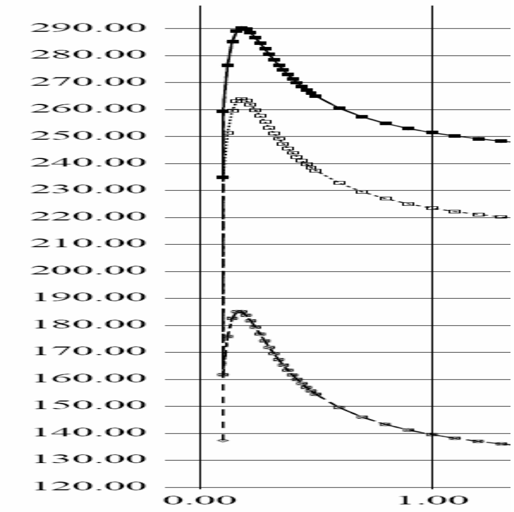
Threshold Voltage pMOS 0.35μ





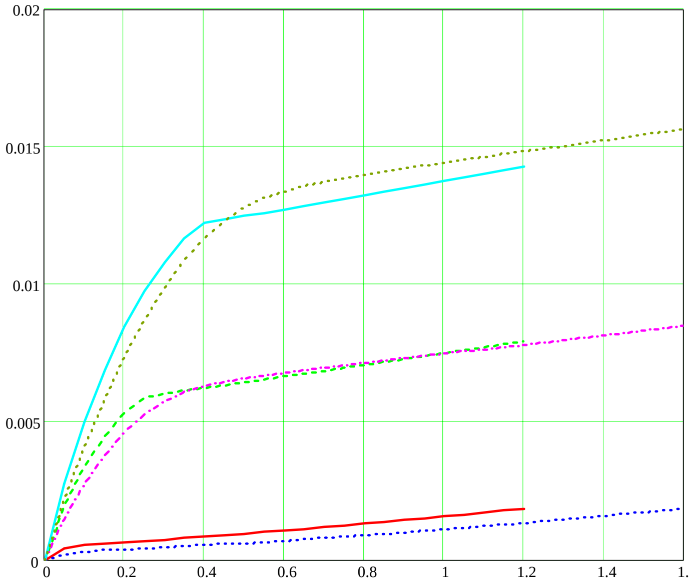
Threshold Voltage





Checking the EE 313 Vsat Model

- Solid is model
  - Dashed is data
- Very good fit!
  - High DIBL
  - Causes low gds



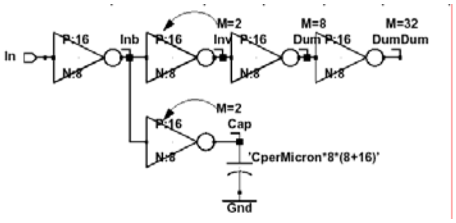
Cg Calibration (Delay)

We like our RC model, so we need to figure out what R and C are

- Gate Capacitance -- fF/μ

  Used for two reasons, delay and power, and they are not the same

  How does it depend on input slope, output slope, temp, V



Why have so many gates? The cap of the M=8 gate is the one we are trying to match, by matching the delay of the M=2 gate. 1x is to make input to 2 2x nominal.

Find C so the delay of 4x gate is the same in both paths

Can change pre/post gate to change input/output slope

Cg Calibration (Power)

Gate cap for power:

- Many options, and a little tricky

- If you measure current pulled by M=2 gate, I(Vdd)
  - Includes short circuit current
  - Includes parasitic cap current

    diodes, and gate overlap on M=2 gate!
- Measure the current going into M=8 gate
  - Add 0V voltage source between driver and gate
    - Remember the average current will be zero
    - Measure the current to charge capacitor
    - C= Q/Vdd; Q = integral of I
    - This should give you the correct answer

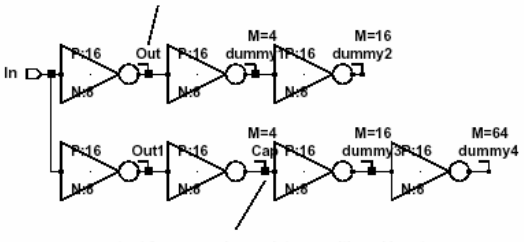
R_{tran} Calibration

Resistance of a transistor -- Ωμ/W

Know gate effective cap, so R = GateDelay/C

How does R vary with Temp, input slope, V

output for measuring resistance with step input



output for measuring resistance with real input

AD, AS, PD, and PS are zero for all transistors

- How do R's add (two transistors in series)?

  replace inverter with enable tristate inverter, but watch parasitic cap

Watch for parasitic capacitance changing your measurement. Since can't remove some parasitics, measure resistances with large load caps (large FO), and set all parasitics you can to zero.

Better approach is to measure the delay for a couple of fanouts. From the slope of the delay vs fanout, you get the effective resistance.

Parasitic Capacitance Calibration

Effective capacitance of transistor parasitics -- fF/μ or fF/ μ^2

- This is hard since many depend on gate W
  - Gate overlap, diffusion edge under gate
  - Source and drain series resistance also depends on W
    - Making W very small can lead to large R
- To extract Gate overlap, diffusion edge, area
  - Replace M=8 inverter with diode model with grounded gate
  - This will give you the sum of all the components
  - Changing gate width, PS, AS can allow you to estimate components
  - W, AS=0, PS=0 give gate overlap + junction under gate
  - W=0.2μ is dangerous, since it is likely to have large series R

- For diffusion cap C for rising and falling transitions are different!

Wire Parasitics

Wire parameters

- Cap fF/μ, ratio to gate cap
- Coupling
- Resistance, RC product

Now What?

Use your simple models to reason about circuit

- Look at different trade-offs
- Try to determine what is important
- If you need more information, do some sims to build new model
- Come up with ’good’ first pass design

And then simulate it!

- First look at a few of the corners that might be interesting
- See if the results make sense

  If not, check the circuit schematics

  Check you models (where does it break down -- learn)
- Check it over many corners

Simulation Issues

Conflict:

- Simulation is cheap, silicon revs are VERY expensive
  - Don’t scrimp when you construct a SPICE deck
  - Simulate the real stuff under real conditions

    Include the real input waveform and load devices

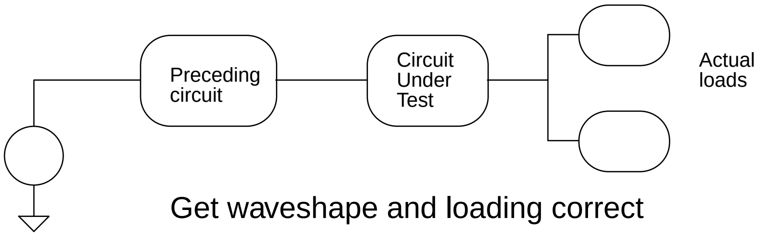
- The more complex the deck, the more confusing the results
- will be
  - Easier to make mistakes in entry
  - Simulating the wrong thing
  - Interaction of lots of small mistakes
  - Hard to debug
  - Slow simulations / long revision times

Compromise

Start simple, and then add complexity

- Start with an understandable and predictable simulation deck
- Add more complexity, and check at each step that the results make sense
- End up with complete simulation file

Need to remember to add all the effects you need to model



Get waveshape and loading correct

Matching

If you want two transistor to match you need to be very careful

Almost anything will make them different

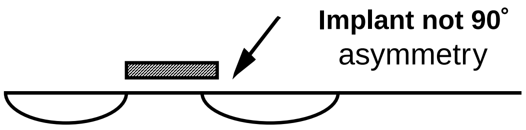
In SPICE all transistors match perfectly,

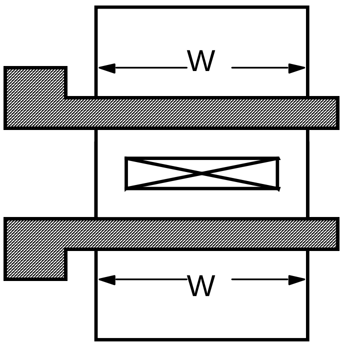
You need to add mismatch explicitly to model problems

Things you need to avoid:

X, Y in silicon are not exact the same.

Match -> orient them the same way

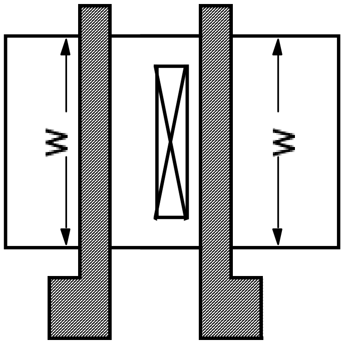

Implant not 90˚
asymmetry



These transistors will not match

Matching

Need to worry about poly alignment



In this case diffusion resistance, and cap will not match

Solution:

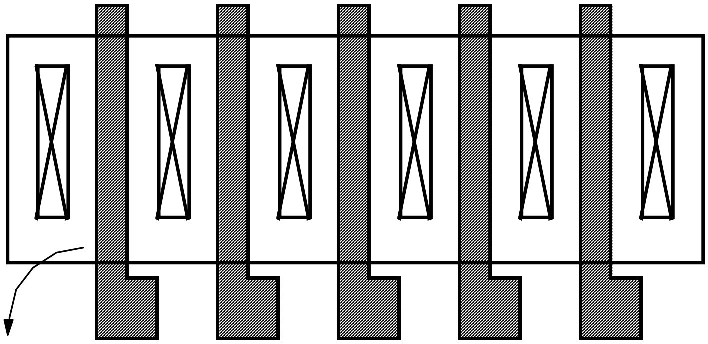
Make sure current flow in the same direction in matched transistors

Easy if all the transistors are folded

Matching

Poly width needs to be carefully controlled

- Etch rate depends slightly on the local density of poly
- To match transistors local poly density needs to be the same.

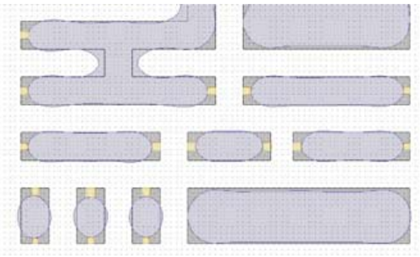


Will not match

End transistors will be different, next to end might be effected too

Welcome To Modern Technology

- Feature size is now below the wavelength of light
  - Very hard to print sharp anything
  - Layouts are preprocessed first
    - Resolution Enhancing Technology



e.g. Line end shortening

- To control transistor properties
  - All transistors must be vertical
  - Poly edges must be far from diffusion
- Moving toward regular arrays of transistors
  - Looks similar to old gate array designs

Statistical Matching

- The errors we have been talking about were systematic
  - You can (in theory) make them zero
  - And you generally can figure out what happened
- But fundamentally even if you do everything right
  - There will still be some random mismatches between transistors
  - These are caused by random doping variations in the device
  - And small random variations in the etching process
- These effects can be models by adding an uncertainty to
  - Vth
  - K, or β, the current prefactor in the current equation

Statistical Matching

- Read Pelgrom's paper!
  - It is the classic paper in this area

- The equations that he set out are still being used today
  - Data indicates that the matching depends on the sqrt(LeffWeff)
  - σ(Vth) = 0.6V*Tox/sqrt(LeffWeff)
    - Need to measure Tox and L,W in same units
  - σ(Ids) = 2%μ/sqrt(LeffWeff)
    - This is in addition to the variation caused by the Vth variation