Lecture 6

Technology Trends and Modeling Pitfalls: Transistors in the "real world"

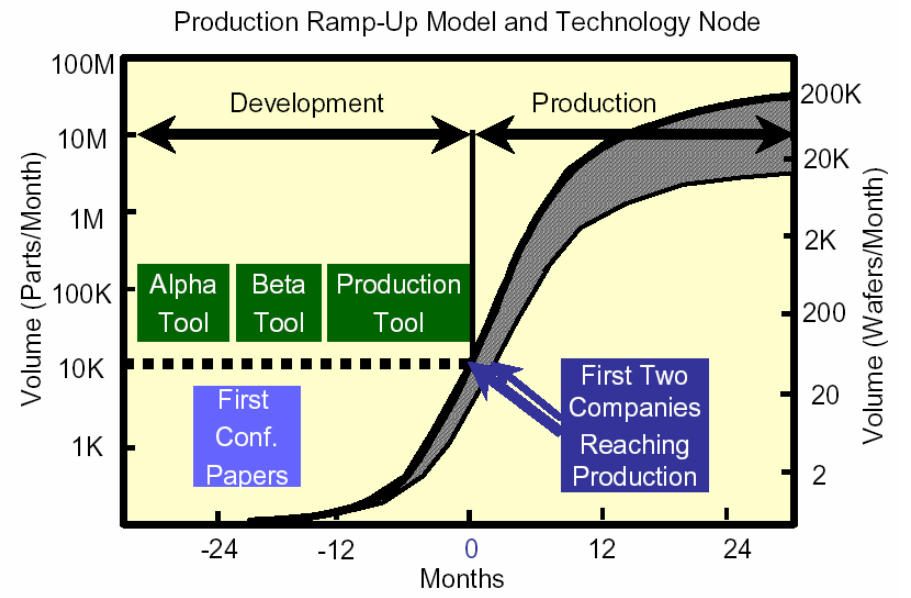
Guest lecturer: Jared Zerbe
Rambus Inc
jared@rambus.com

Some Figures courtesy of C. Enz, M. Bucher, D.Foty, 2003 ITRS, IBM

Overview

- CMOS technology trends
- Deep submicron & SOI processes
- MOS Modeling & industry gotchas
- Next generation modeling

ITRS Node production ramp



CMOS Technology generations *(OLD : '99NTRS)*

95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12
350 nm	1	2	3	4	5												
-2	-1	250 nm	1	2	3	4	5										
-4	-3	-2	-1	180 nm	1	2	3	4	5								
-6	-5	-4	-3	-2	-1	150 nm	1	2	3	4	5						
-8	-7	-6	-5	-4	-3	-2	-1	130 nm	1	2	3	4	5				
-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	100 nm	1	2	3	4	5	
			-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	70 nm	1	2	3
						-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	50 nm

exact DATES are FLAWED

- Research (7), Development (5), Manufacturing (5)
- Technologies span ~17 years: unlikely to be totally surprised
- The rate at which things change is what's debatable

Technology Scaling & Moore's Law

- Scaling is extremely well predicted & controlled
- Driven by Moore's Law *# of DRAM bits 4X every 3 years*

Technology (2x) X Diesize (1.4x) X Innovation (1.4x) = 4X

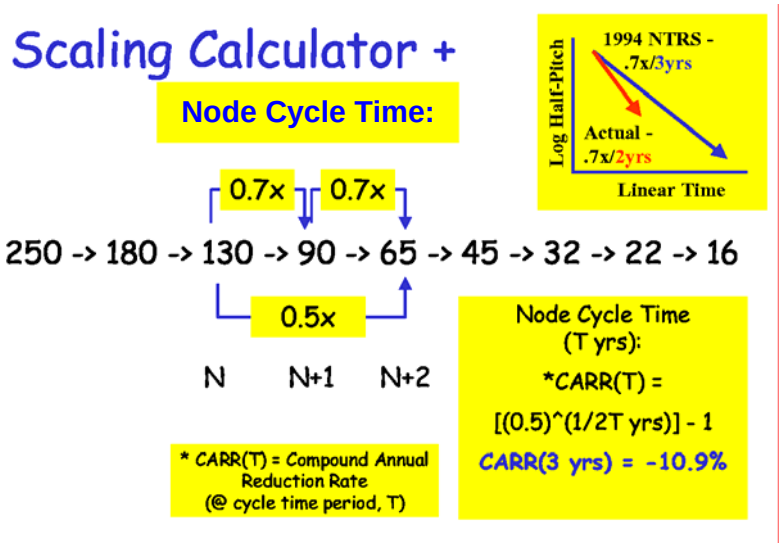
– BUT

Technology now making up for diesize (die per wafer) limits

=> Technology has been 2x every two years since 1995

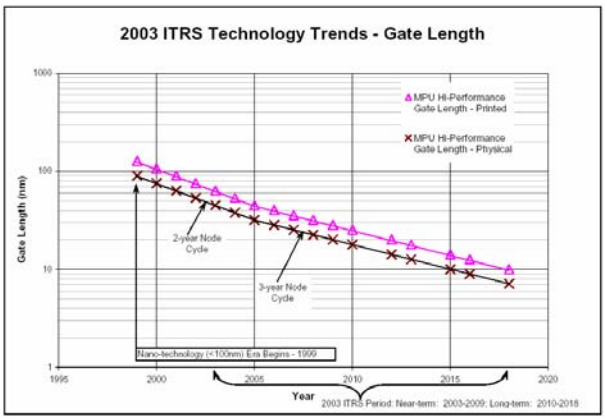
=> Allows *diesize* to remain virtually constant

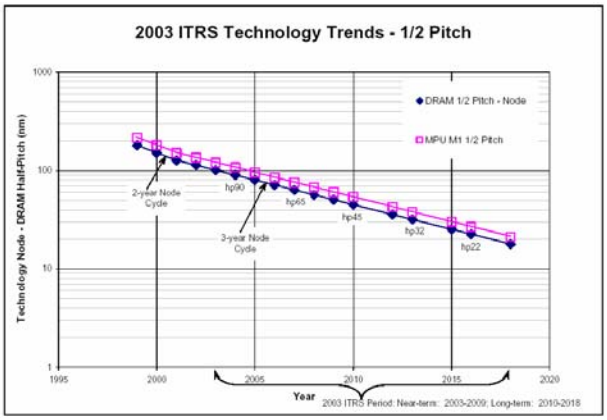
Technology Scaling & CARR



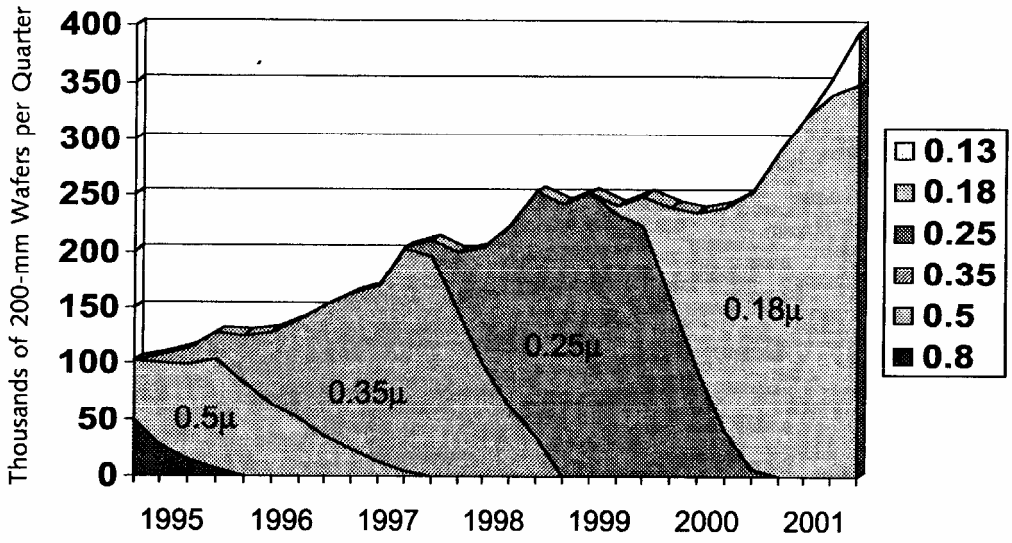
- This is important : lots of time & $$ spent to keep it on track!

Gate & Interconnect Scaling
From 2003 IRTS



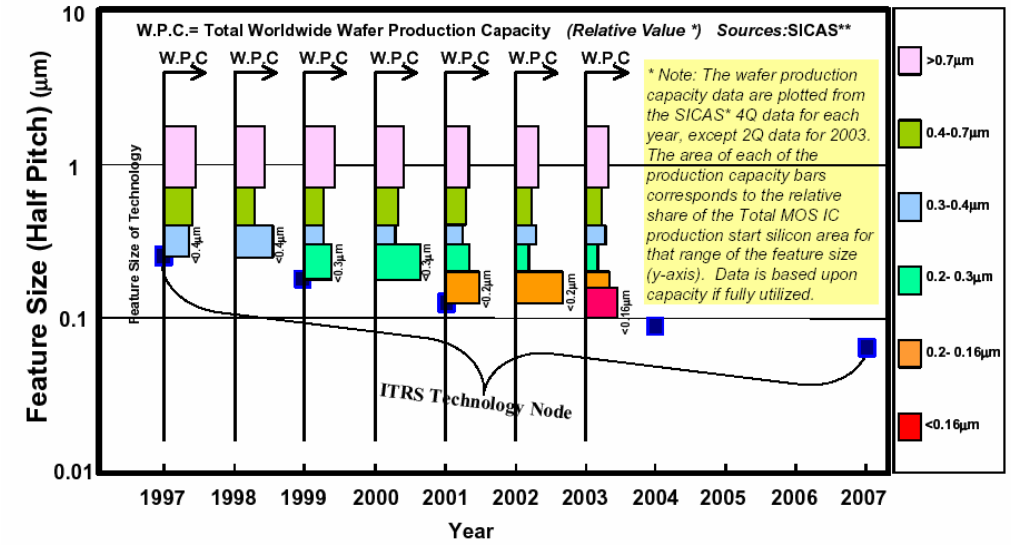


Technology & Intel wafer capacity(μP)



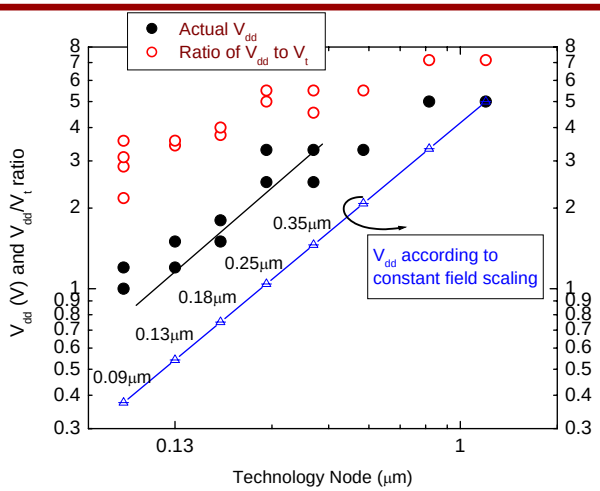
- **In uP's Useful years per technology is shrinking, but total volume same or growing!**

Technology & Worldwide Production Capacity



- When you're not uP's technologies stay around a lot longer
- Majority of WW capacity continues to be at 0.3u+

Technology trends: Vdd & Vt scaling



- After 5V -> 3.3V the "Berlin wall" cracked & Vdd has been dropping
- Current is not increasing : speed comes from lowered capacitance

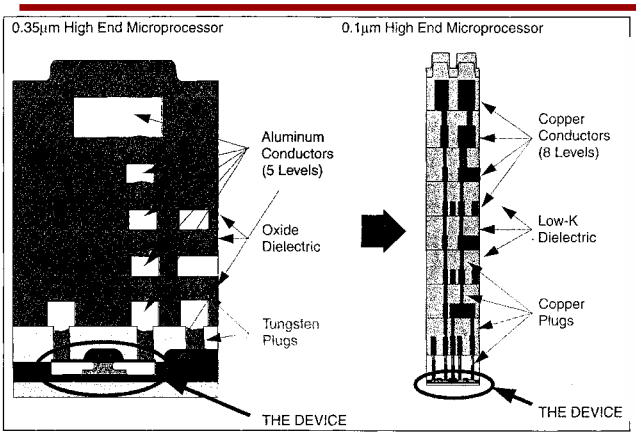
Technology trends: Vdd scaling & you

- Scaling not occurring on Vt at the same rate!
  - Device counts going up & leakage too high to lower Vt : power!
  - Headroom a major issue for analog circuits -> stacked structures are tough
    - Migration of analog designs a serious headache
    - What was once “Vdd = 5 Vt’s” becomes “Vdd = 3Vt’s”
    - Budget your headroom carefully
  - Different Vt devices are coming… or already here
    - Low & High-Vt devices (separate implant = $)
    - Native device is ‘free’
    - BEWARE: Using any special device make your design less portable
      - But may become inevitable

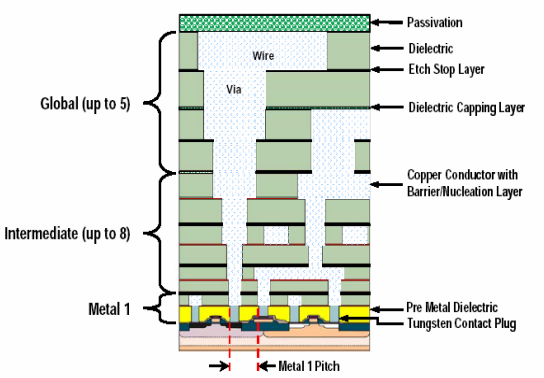
Technology trends: multiple supplies

- Frequently multiple supplies on same die
  - Further reduce power or jitter (on-chip regulators)
  - Compatibility w/different devices (I/Os)
  - Further reduce leakage (DRAM)
- Be careful when crossing domains
  - Watch pass-gates & forward-biasing a diode
  - Timing issues, power consumption
  - Multiple oxides, multiple different device types

Wires are very important…
and becoming even more important



1. Process architecture challenge.



- Metal layers are not equal: top layer is special
  - Which layer is top? Hierarchical scaling?
- Fringing much more important. $C_{fringe} > C_{area}$ below 0.25µ
  - Your tools must be up to the job

Overview

- CMOS technology trends
- Deep submicron & SOI processes
- MOS Modeling & industry gotchas
- Next generation modeling

Deep Submicron : 2001 ITRS Predicted Performance Logic Technology Characteristics

Table 2. High-Performance Logic Technology Requirements, Data from 2001 ITRS

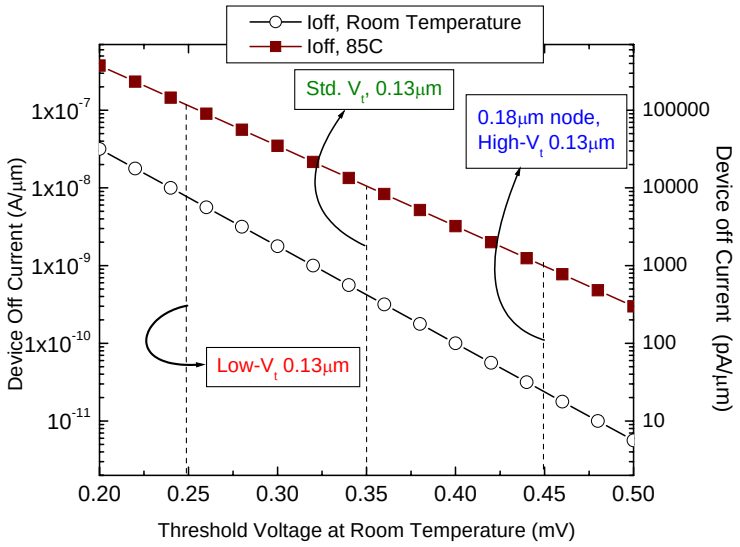
		Near Term							Long Term		
Calendar Year		2001	2002	2003	2004	2005	2006	2007	2010	2013	2016
DRAM Half Pitch	nm	130	115	100	90	80	70	65	45	32	22
Physical Gate Length, L_g	nm	65	53	45	37	32	28	25	18	13	9
Equivalent Oxide Thickness, T_{ox}	nm	1.3-1.6	1.2-1.5	1.1-1.6	0.9-1.4	0.8-1.3	0.7-1.2	0.6-1.1	0.5-0.8	0.4-0.6	0.4-0.5
Nominal Power Supply Voltage (Vdd)	V	1.2	1.1	1.0	1.0	0.9	0.9	0.7	0.6	0.5	0.4
Nominal High-Performance NMOS Sub-Threshold Current (@25C)	µA/µm	0.01	0.03	0.07	0.1	0.3	0.7	1	3	7	10
Nominal NMOSFET Saturation Drive Current, I_{on}	µA/µm	900	900	900	900	900	900	900	1200	1500	1500
Required Percent Current-Drive "Mobility/Transconductance Improvement"		0%	0%	0%	0%	0%	0%	0%	30%	70%	100%
Parasitic Series S/D Resistance, $R_{sd,series}$	Ω-µm	190	180	180	180	180	170	140	110	90	80
Parasitic Capacitance Percent of Ideal Gate Capacitance		19%	22%	24%	27%	29%	32%	27%	31%	36%	42%
NMOSFET Intrinsic Transistor Delay, τ_i	ps	1.65	1.35	1.13	0.99	0.83	0.76	0.68	0.39	0.22	0.15
NMOSFET Intrinsic Transistor Switching Frequency, $f_i = 1/\tau_i$	GHz	606	742	888	1007	1205	1320	1463	2570	4445	6514
Relative Device Performance		1.0	1.2	1.5	1.6	2.0	2.1	2.5	4.3	7.2	10.7
Energy per (W/Lgate=3) Device Switching Transition (Cgate*(3*Lgate)*V^2)	fJ/Device	0.347	0.212	0.137	0.099	0.065	0.052	0.032	0.015	0.007	0.002
Static Power Dissipation Per (W/Lgate=3) Device	Watts/Device	5.6E-09	6.7E-09	1.0E-08	1.1E-08	2.6E-08	5.3E-08	5.3E-08	9.7E-08	1.4E-07	1.1E-07

Deep Submicron : 2003 ITRS Predicted Performance Logic Technology Characteristics

Year of Production	2003	2004	2005	2006	2007	2008	2009
Technology Node		hp90			hp65		
DRAM ½ Pitch (nm)	100	90	80	70	65	57	50
MPU/ASIC Metal 1 (M1) ½ Pitch (nm)	120	107	95	85	76	67	60
MPU/ASIC ½ Pitch (nm)	107	90	80	70	65	57	50
MPU Printed Gate Length (nm)	65	53	45	40	35	32	28
MPU Physical Gate Length (nm)	45	37	32	28	25	22	20
Physical gate length high-performance (HP) (nm) [1]	45	37	32	28	25	22	20
EOT: equivalent oxide thickness (physical) for high-performance (nm) [2]	1.3	1.2	1.1	1.0	0.9	0.8	0.8
Electrical thickness adjustment for gate depletion and inversion layer effects (nm) [3]	0.8	0.8	0.7	0.7	0.4	0.4	0.4
Equivalent electrical oxide thickness in inversion (nm) [4]	2.1	2.0	1.8	1.7	1.3	1.2	1.2
Nominal gate leakage current density limit (at 25°C) (A/cm^2) [5]	2.2E+02	4.5E+02	5.2E+02	6.0E+02	9.3E+02	1.1E+03	1.2E+03
Nominal power supply voltage (V_{dd}) (V) [6]	1.2	1.2	1.1	1.1	1.1	1.0	1.0
Saturation threshold voltage (V) [7]	0.21	0.20	0.20	0.21	0.18	0.17	0.16
Nominal high-performance NMOS sub-threshold leakage current, I_{sdLeak} (at 25°C) (µA/µm) [8]	0.03	0.05	0.05	0.05	0.07	0.07	0.07
Nominal high-performance NMOS saturation drive current, I_{dsat} (at V_{dd} at 25°C) (mA/mm) [9]	♦ 980	1110	1090	1170	1510	1530	1590
Required "mobility/transconductance improvement" factor [10]	1.0	1.3	1.3	1.4	2.0	2.0	2.0
Sub-threshold slope adjustment factor (full depletion/multiple-gate effects) (0–1) [11]	1.0	1.0	1.0	1.0	1.0	0.8	0.7
Effective saturation carrier velocity enhancement factor (due to quasi-ballistic transport) [12]	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Parasitic source/drain series resistance (R_{sd}) (Ohm-µm) [13]	♦ 180	180	180	171	162	153	144
Ideal NMOS device gate capacitance (F/µm) [14]	7.40E-16	6.39E-16	6.14E-16	5.69E-16	6.64E-16	6.33E-16	5.76E-16
Parasitic fringe/overlap capacitance (F/µm) [15]	2.40E-16	2.40E-16	2.40E-16	2.30E-16	2.20E-16	2.00E-16	1.90E-16
High-performance NMOS intrinsic delay, $\tau = C_{gate} * V_{dd} / I_{dsat}$ (ps) [16]	♦ 1.20	0.95	0.86	0.75	0.64	0.54	0.48
Relative NMOS intrinsic switching speed, $1/\tau$, normalized to 2003 [17]	♦ 1.00	1.26	1.39	1.60	1.86	2.20	2.49
Nominal logic gate delay (NAND Gate) (ps) [18]	♦ 30.24	23.94	21.72	18.92	16.23	13.72	12.13
NMOSFET power-delay product (J/µm) [19]	1.41E-15	1.27E-15	1.03E-15	9.66E-16	1.07E-15	8.33E-16	7.66E-16
NMOSFET static power dissipation due to drain and gate leakage (W/µm) [20]	3.96E-07	6.60E-07	6.05E-07	6.05E-07	8.47E-07	7.70E-07	7.70E-07

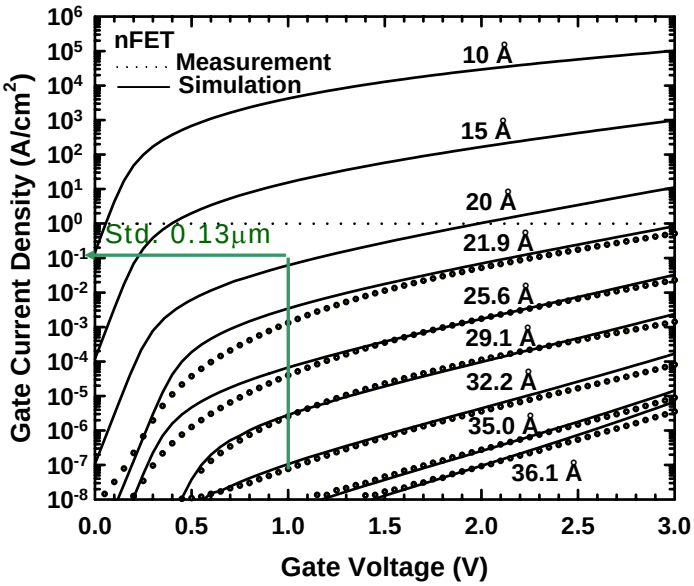
Device Off Current as a Function of Threshold Voltage

- Device leakage increases dramatically from low temperature to high temperature due to V_t drop and degradation of the subthreshold swing.
- The "Standard" 0.13μm node has at least 10X higher device leakage than the 0.18μm node due to approximately 100mV lower V_t.
- This is driven by the reduced supply voltage required for the 0.13μm.
- In calculating a part's standby power one has to consider the highest rated temperature (normally 85C – 125C depending on application)



Gate Leakage Due to Oxide Scaling

- At the 0.13μm process (standard process) this leakage is 1000pA for a 1μmX1μm device, and about 100pA for a device with WxL of 1μmX0.13μm.
- Note that the current actually becomes larger for longer channel devices, which is the opposite behavior of off current.
- At room temperature, this current can be higher than the off current of the high-Vt device and close to off current of the standard-Vt device.
- Fortunately, this current is a very weak function of temperature, in contrast to the off current.
- The current is an exponential function of oxide thickness. For example, the low voltage version of 0.13μm technology has more than 10X gate current.
- It is a weak function of supply voltage.



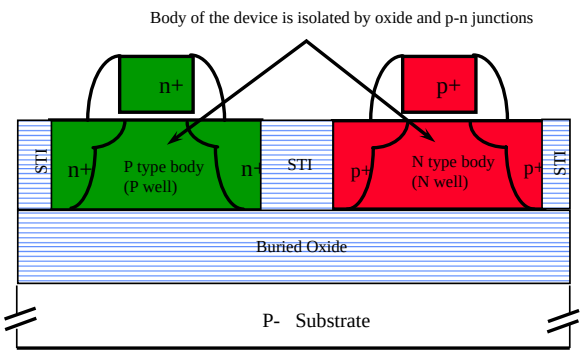
S.-H. Lo et al., *IBM Journal of Research and Development*, vol. 43, no. 3, p. 327, 1999

Deep Submicron Significant Issues

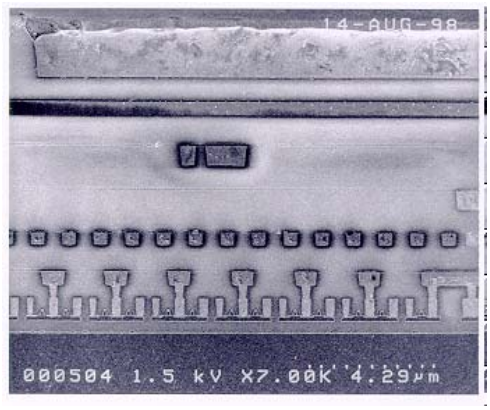
- Cost
  - Cost of 90nm and below devices are skyrocketing
    - Mask set alone is > 0.25 $M
  - Cu interconnect, low-K dielectrics, fancy lithography all $ ↑
- Power
  - Leakage current is now comparable to switching current!
  - Vdd scaling -> Vt scaling -> off currents going up
  - Thinner oxides -> gate current going up
- Wires
  - Metal density rules for planarization
    - Usually get by with “fill routines”
    - Density numbers are getting very challenging
    - Extraction with metal fill in place usually not done

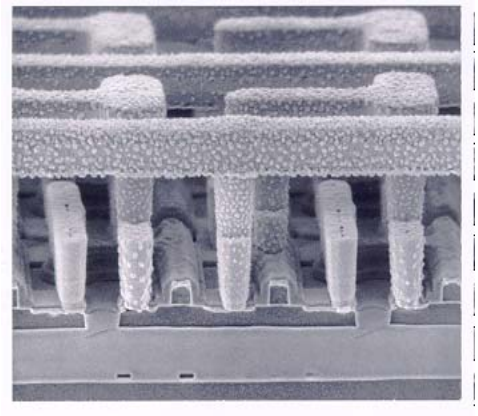
What’s SOI?

- Body terminal of a MOSFET is treated as its fourth terminal (affecting its V_t, current drive, leakage, and capacitance).
- In a bulk MOSFET, the device body is either tied to the V_{dd} (PFET) or to the ground (NFET).
- In an SOI MOSFET, the device body is isolated by oxide insulators and p-n diodes. Therefore, the body of an SOI MOSFET is left floating and is called NFD (non fully depleted) or PD (partially depleted)



SOI Process Cross-Section





Courtesy IBM

Current Issues in Designing on SOI

- NFD Devices Sometimes used for Digital SOI Due to:
  - Better Control of Threshold Voltage
  - Larger Design Window for Control of Short-Channel Effects
  - Ease of Manufacturing
- Area of Concern for NFD MOSFETs: DC and Transient Effects Associated With the Floating Body
- Analog designs usually grab control of the body

Technology trends - conclusions

- Processes take a long time to develop & make manufacturable
  - You can make one of anything...
- Devices are getting less friendly
  - Lower Vdd/Vt ratio makes analog more challenging
  - Multiple supplies on-chip
  - Multiple Vt's, multiple oxides
  - Static 'off' power becoming significant in performance processes
- Wires are more important than ever
  - Lots of layers, lots of fringe capacitances
  - Fill rules
  - Local vs. global clocking?
  - Tools
- Still lots of room for creative circuit design!

Overview

- CMOS technology trends
- Deep submicron & SOI processes
- MOS Modeling & industry gotchas
- Next generation modeling

MOSFET modeling : approaches

Two basic approaches over time:

- Physical

  Parameters have physical meaning
  Can be extracted from physical measurement (Tox, Ld, etc.)
  Usually simple, few parameters (“one page’r”)

- Empirical

  Use curve-fitting to match measured devices
  Parameters hard to understand, and there are LOTS
  Mostly mathematical approach

Reality is always a compromise

WARNING: Physical models can fit poorly
Empirical can break outside measured space

Modeling : The Big Problem

The biggest problem when it comes to MOS Modeling:

Circuit designers want a model that is 100% accurate, physically intuitive, very fast, preferably 6-months before the process is stable, and don’t want to pay for it.

Process designers want circuit designers to make their designs robust and tolerant to ‘minor variations’.

Fabs don’t get paid for having a better model…

…. but if your model is broken your circuit may be too!

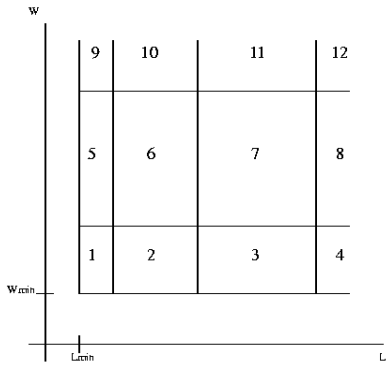
MOSFET Modeling: brief history

- First generation
  - Hspice Level 1, 2, 3
  - “Physical” analytical models with geometry in model equations
  - Holding onto hand-calculation…
- Second generation
  - Hspice level 13, 28, 39: Bsim, “MetaMOS”, Bsim2
  - Shift in emphasis to circuit simulation with lots of mathematical conditioning
  - Quality of outcome is highly dependent on parameter extraction methodology
  - Good luck with hand-calculation
    - BUT served industry well for almost 10years!

MOSFET Modeling : the present

- Third generation: Hspice level 49, 55: Bsim3v3, EKV
- Most new models are Bsim3v3
  - Bsim3 intent was return to simplicity... now >100 parameters!
  - Often start simple… and add complexity w/measured data
  - Binning still used to cover W&L space
  - Extensive mathematical conditioning
  - YOU will probably be using a Bsim3v3 model in your future
- Next Generation…

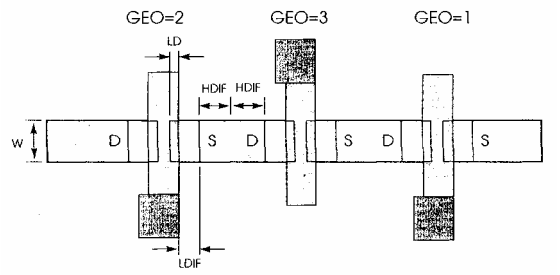
MOSFET Modeling : know your binning!



- Model binning often required for highest level of accuracy
  - Know your bin-space (remember process corners push you)
  - Beware of non-physical behavior at boundaries & beyond limits
  - Know what bin(s) your circuit is using

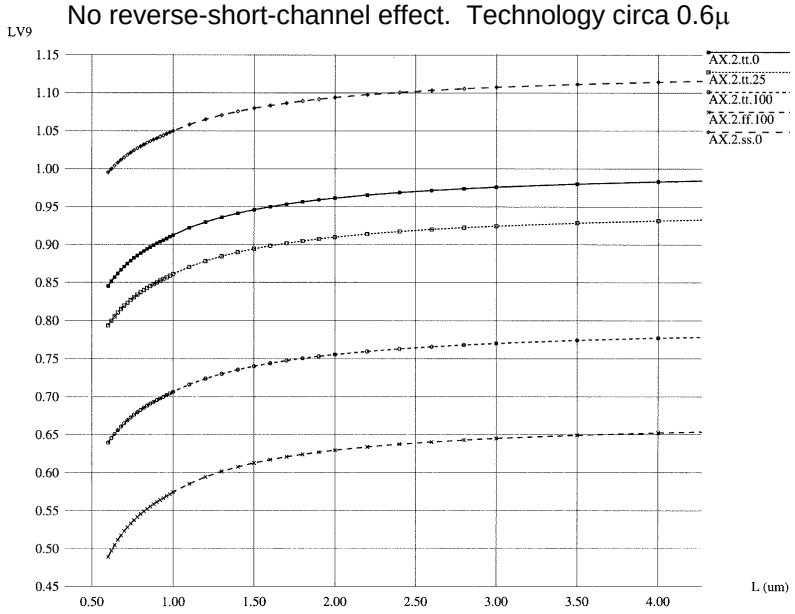
MOSFET modeling: check the basics

- Source/drain diode capacitances are critical - don't get into a "gate cap only" mentality
  - What is the ACM method used?
    - Are all parameters (i.e. Cjgate) included in the model?
    - Do you know about GEO (HSPICE)?
  - Does your model jive with your extraction tool?
  - Does HDIF jive with your layout style?

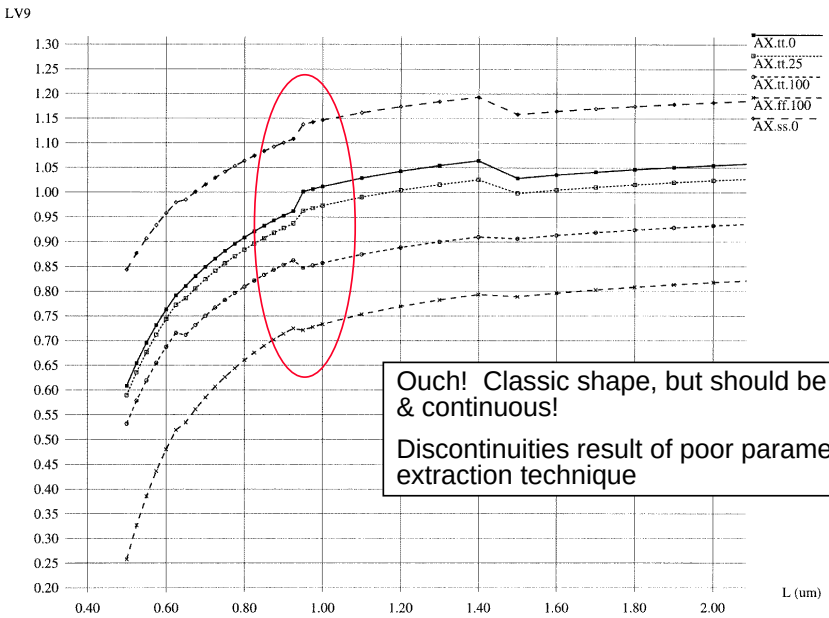


Modeling gotchas : Classic Vt vs. L

No reverse-short-channel effect. Technology circa 0.6µ



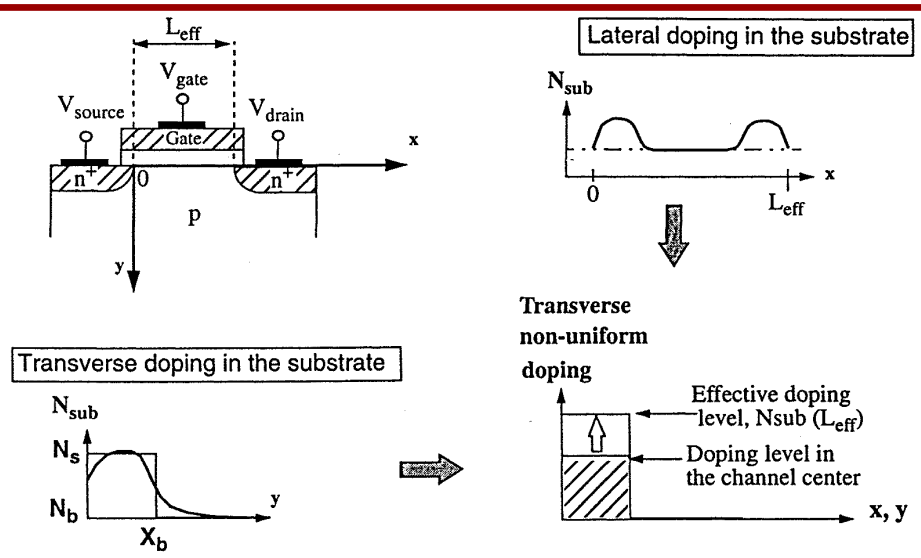
Vt vs. L : discontinuities at model boundaries



Ouch! Classic shape, but should be smooth & continuous!

Discontinuities result of poor parameter extraction technique

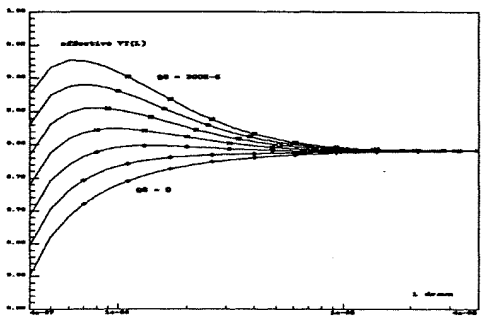
Reverse short channel effect (RSCE)



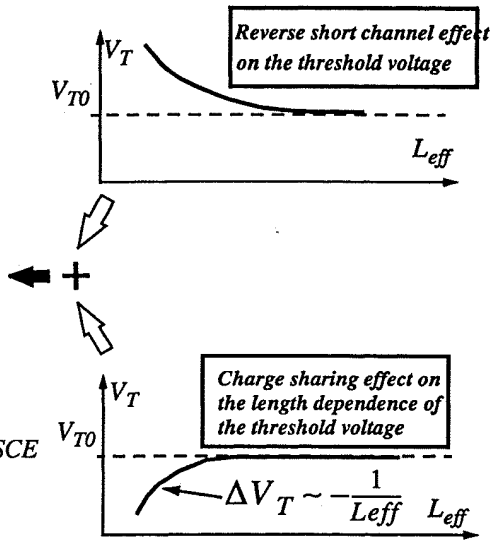
Impurities effect lattice during high-temp process steps

RSCE con't

Really a combination of two effects

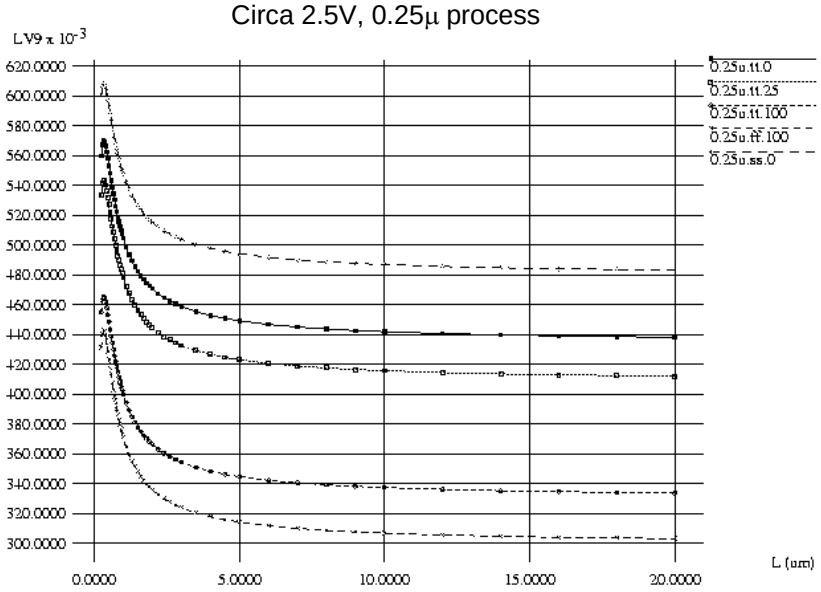


Variation of the threshold voltage due to the RSCE and charge-sharing effects with Leff



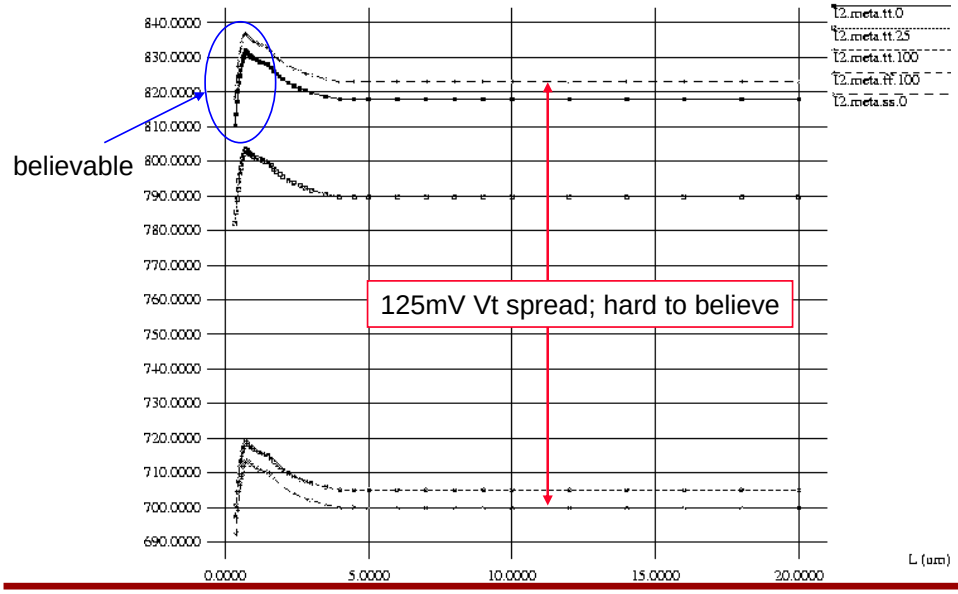
Vt vs. L with RSCE

Circa 2.5V, 0.25μ process



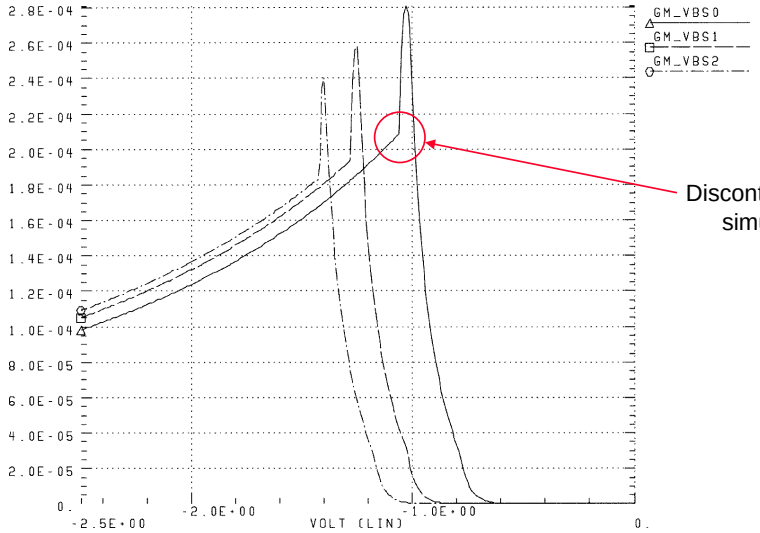
Vt spread : process & temp

Check Vt spread between ff/100C & ss/0C



Modeling gotchas: g_M vs. Vgs

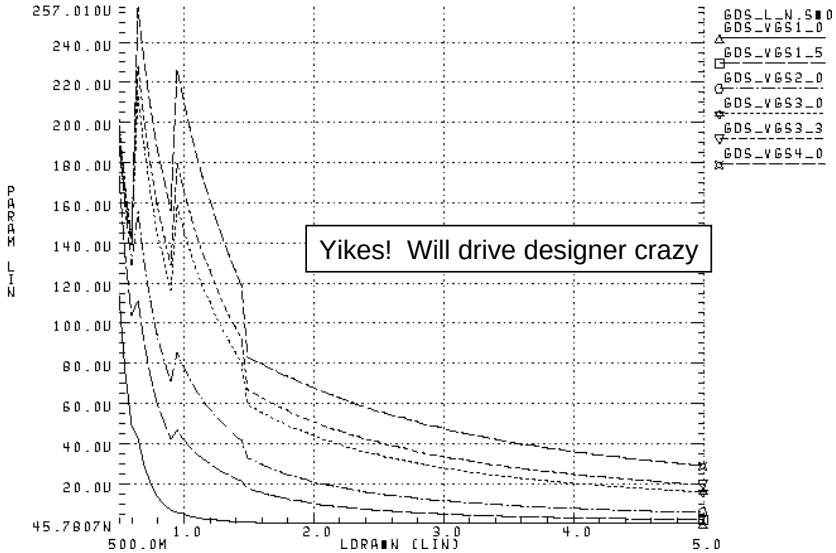
All first-derivatives should be smooth & continuous



Discontinuity will drive simulator crazy

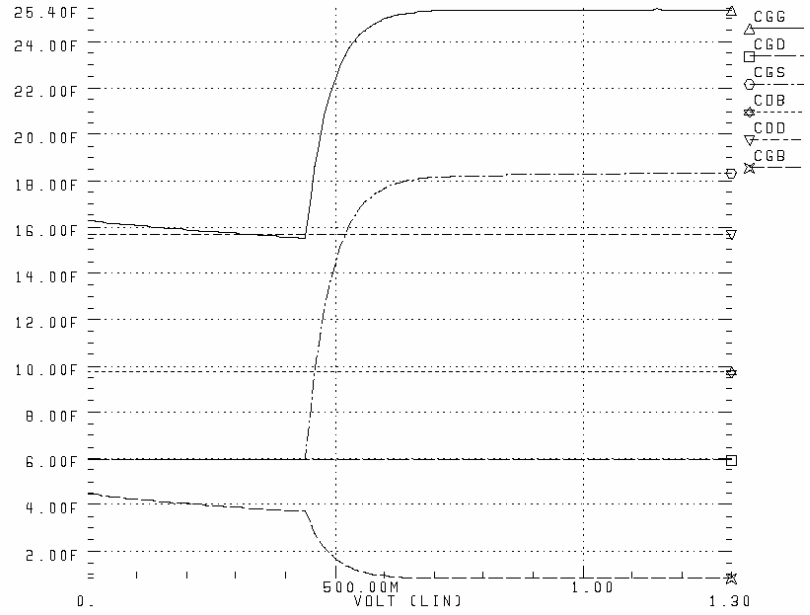
Modeling gotchas: g_{DS} vs. L

Should also be smooth and continuous

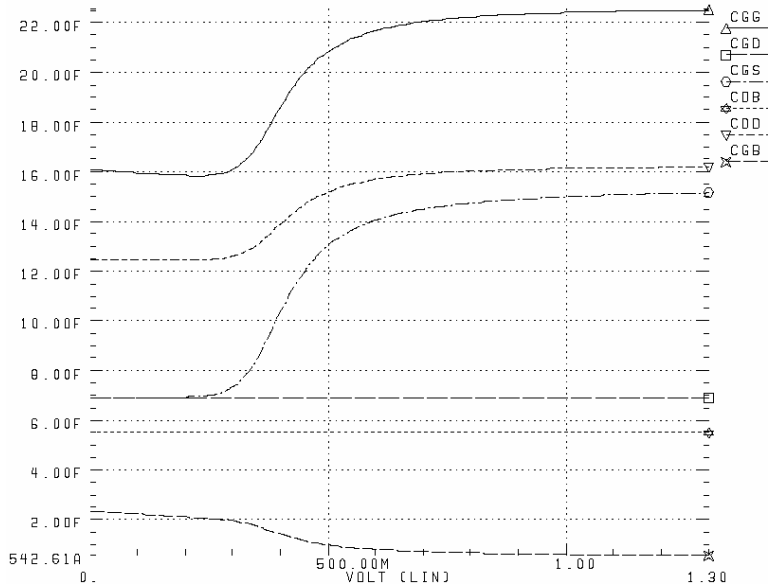


Yikes! Will drive designer crazy

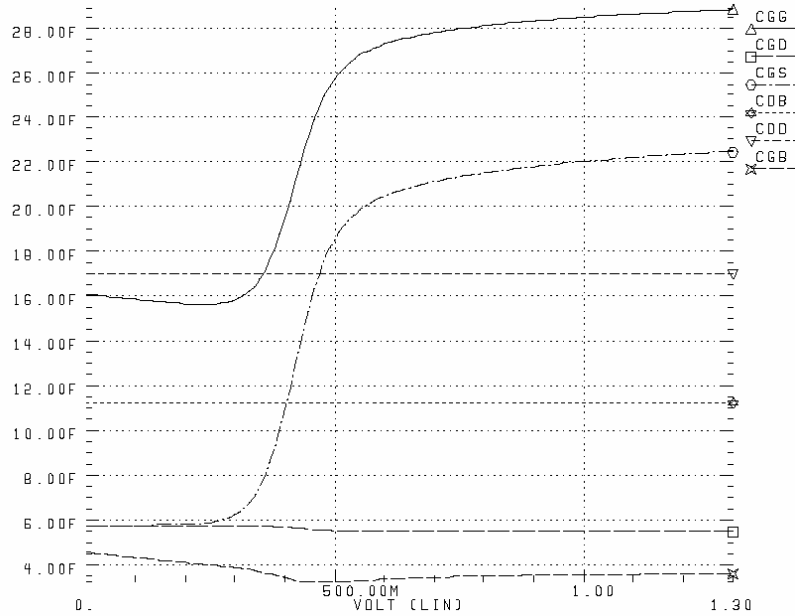
Modeling gotchas: Cgg vs. Vgs - first bsim3



Modeling gotchas: Cgg vs. Vgs - first EKV



Modeling gotchas: Cgg vs. Vgs - bsim3 at +1yr



Overview

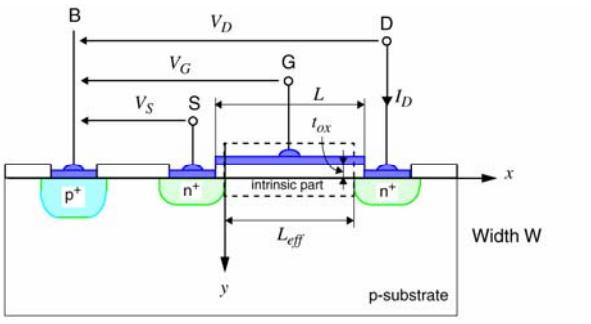
- CMOS technology trends
- Deep submicron & SOI processes
- MOS Modeling & industry gotchas
- Next generation modeling

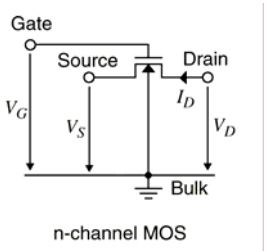
Next Generation Models

- Overly empirical nature of Bsim has led to development of other mos models for analog design
  - Sometimes called ‘compact models’
  - EKV from EPFL in Switzerland
  - SP from U-Penn
    - Surface potential based
    - Uses symmetric linearization ; moves linear point along channel
- In general these models have
  - Bulk reference ; source-drain interchangability
  - Physically based
  - Consistant quasi-static & non quasi-static models
    - Charge is conserved

EKV model : fundamentals

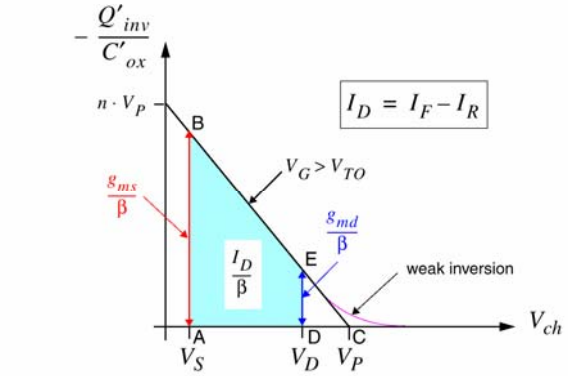
- All voltages referenced to local substrate, not source
  - Takes into account natural symmetry of device





EKV model : V_P, I_F, I_R

$$I_D = \underbrace{\beta \cdot \int_{V_S}^{\infty} \left[\frac{-Q'_{inv}(V_{ch})}{C'_{ox}}\right] \cdot dV_{ch}}_{= \text{forward current } I_F \text{ controlled by } (V_P - V_S)} - \underbrace{\beta \cdot \int_{V_D}^{\infty} \left[\frac{-Q'_{inv}(V_{ch})}{C'_{ox}}\right] \cdot dV_{ch}}_{= \text{reverse current } I_R \text{ controlled by } (V_P - V_D)}$$



$$I_D = I_F - I_R$$

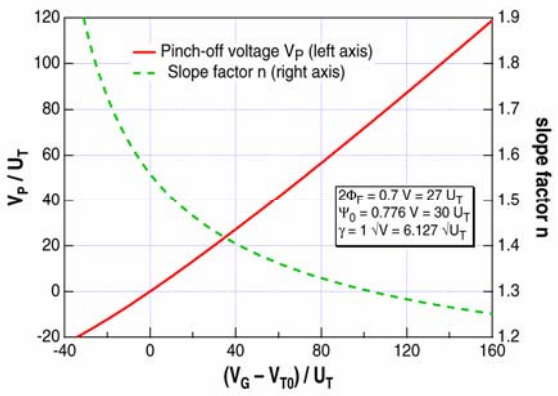
EKV model : Gate sets the pinch-off voltage

V_P represents the voltage that should be applied to the channel to cancel the effect of the gate voltage (Vg > Vt)

- It is where the inversion charge becomes zero

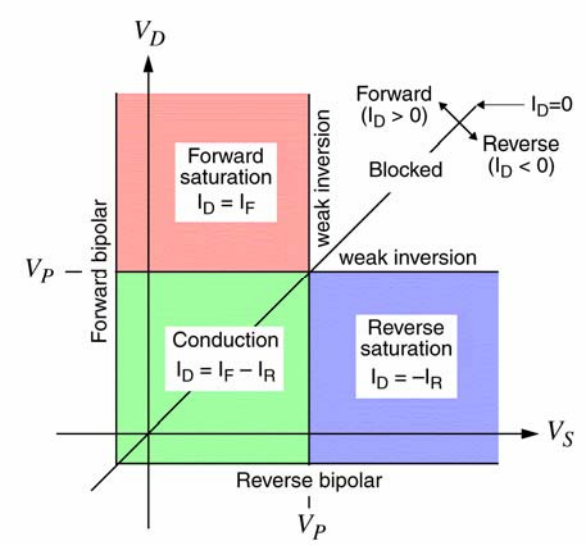
$$V_P = V_G - V_{T0} - \gamma \cdot \left[\sqrt{V_G - V_{T0} + \left(\sqrt{\Psi_0} + \frac{\gamma}{2}\right)^2} - \left(\sqrt{\Psi_0} + \frac{\gamma}{2}\right)\right]$$

$$V_P \cong \frac{V_G - V_{T0}}{n}$$

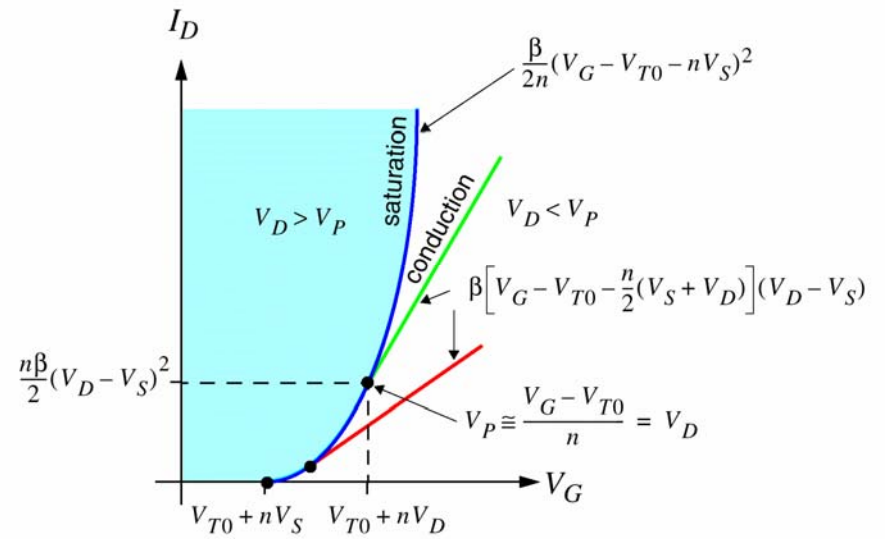


EKV model : Modes of operation

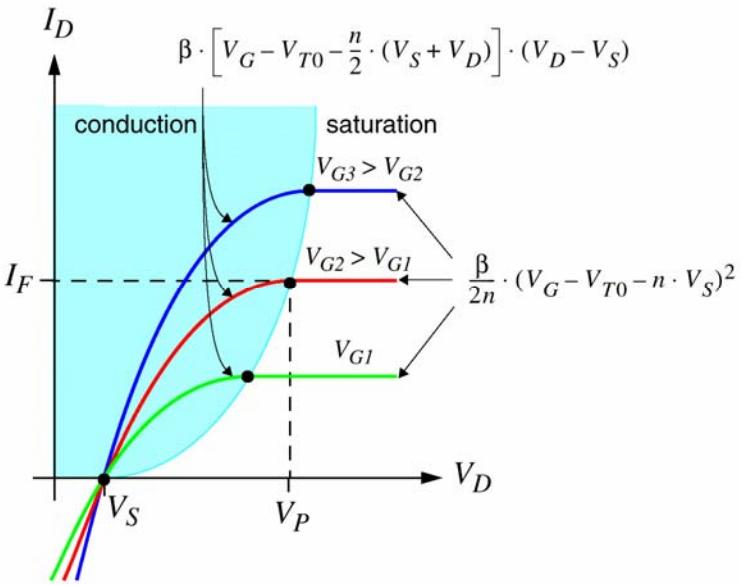
Defined by drain and source voltages w.r.t. V_P



EKV model : Id-Vg characteristics



EKV model : Id-Vd characteristics



MOSFET Modeling : Conclusions

- Big problem of modeling
  - It's in nobody's interest to make sure you have a good model and…
  - There are still disparities between fab & circuit folk
    - What you need as a circuit designer may different than digital
    - This may be uncharacterized
    - Sometimes what you want may be unrealistic!
      "Why is your circuit so sensitive…"
- Result: caveat emptor
  - Examine your models
  - Request reasonable behavior & make your circuits tolerant

References

ITRS (International technology roadmap for semiconductors) website:
http://public.itrs.net/
MEAD Microelectronics (short courses) website:
http://www.mead.ch and http://mead.netgate.net
EKV website:
http://legwww.epfl.ch/ekv/
BSIM website:
http://www-device.eecs.berkeley.edu/~bsim3/
Dan Foty's website:
(author MOSFET Modeling with SPICE principles and practice)
http://www.sover.net/~dfoty
FSA (Fabless semiconductor association) website:
http://www.fsa.org/

Some figures reprinted with permission of C. Enz, M. Bucher, D. Foty