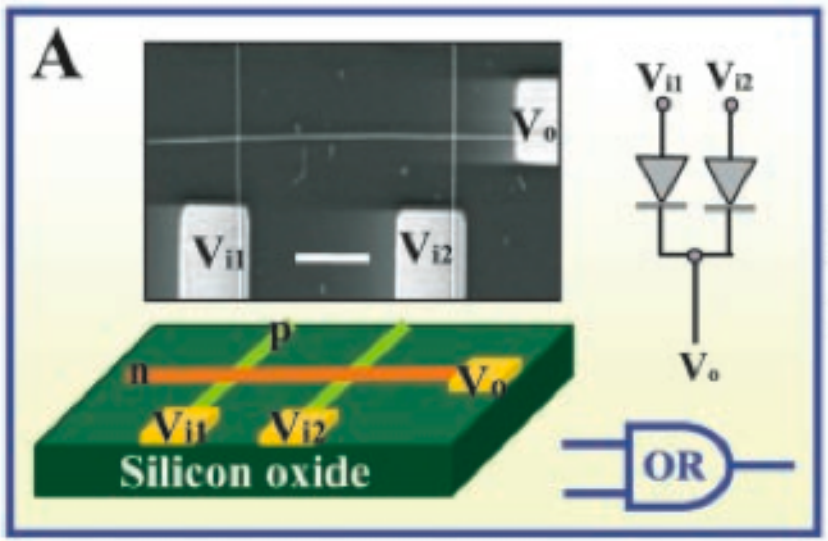
CMOS Transistor Theory (and its effects on scaling)

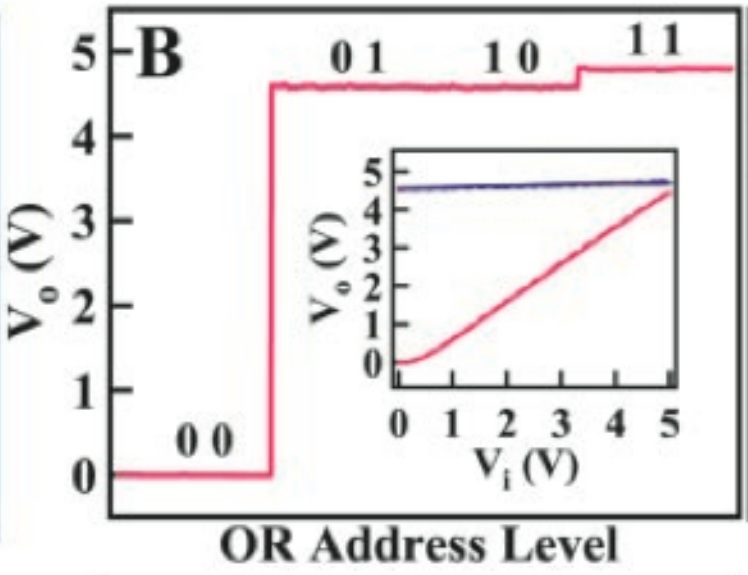
Michael Niemier

(Some slides based on lecture notes by David Harris)

Nanowire-based Gates

Can make *very* small pn junctions and diode based lgoic



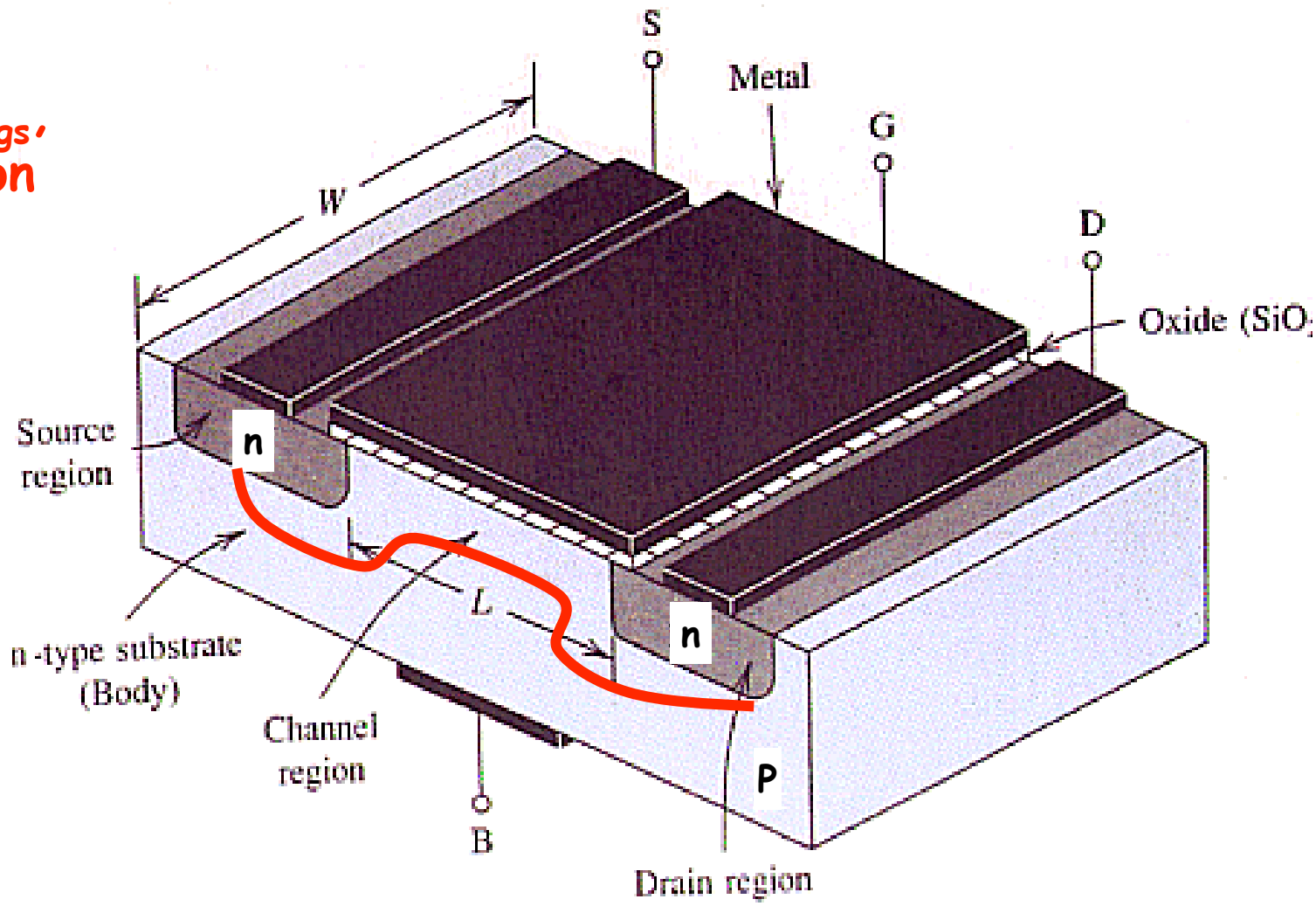


C V_{i1}(V)	V_{i2}(V)	OR V_o(V)
0.0(0)	0.0(0)	0.00(0)
0.0(0)	5.0(1)	4.58(1)
5.0(1)	0.0(0)	4.57(1)
5.0(1)	5.0(1)	4.79(1)

If each wire was just 5 nm in diameter,
would you be excited about this technology?

MOSFET cross section…

With applied V_{gs}, depletion region forms

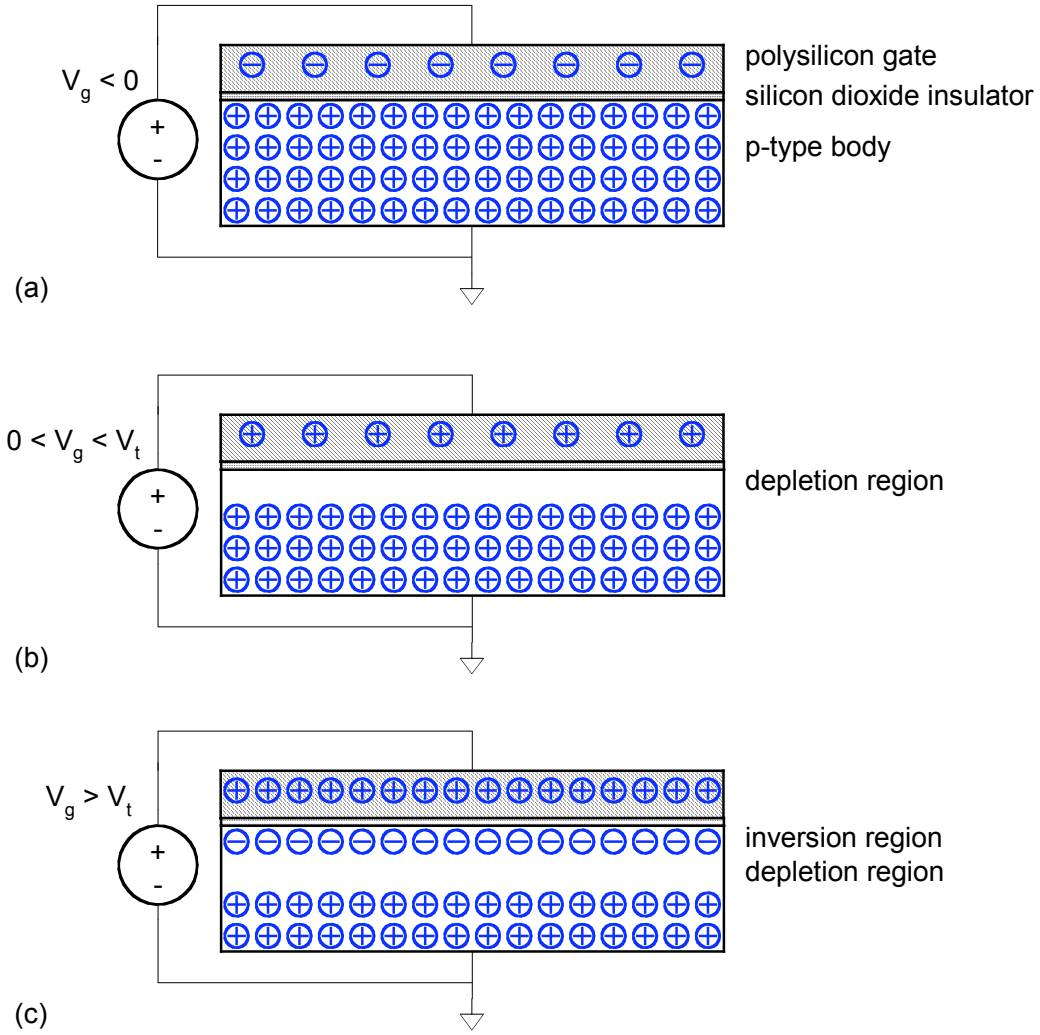


To recap…

- So far, we have treated transistors as ideal switches
- An ON transistor passes a finite amount of current
  - Depends on terminal voltages
  - Derive current-voltage (I-V) relationships
- Transistor gate, source, drain all have capacitance
  - $I = C\,(\Delta V/\Delta t) \rightarrow \Delta t = (C/I)\,\Delta V$
  - Capacitance and current determine speed

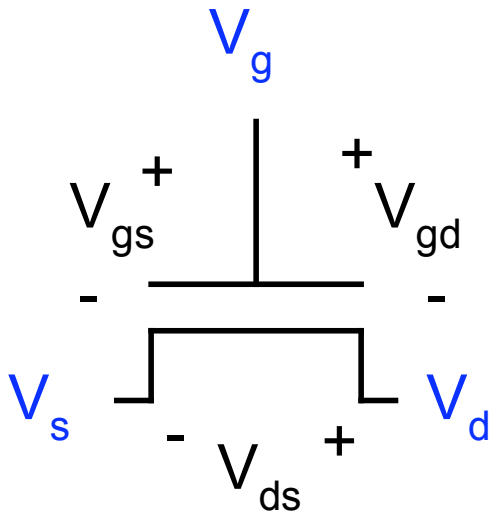
MOS Capacitor

- Gate and body form MOS capacitor
- Operating modes



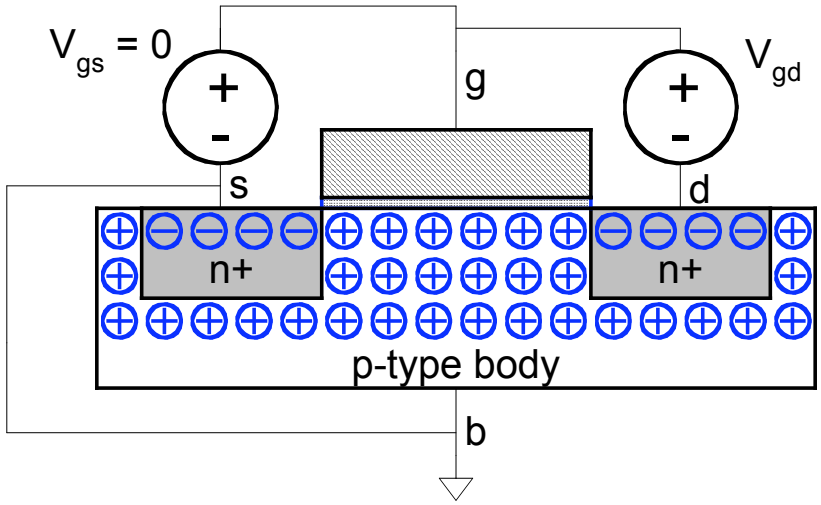
Terminal Voltages

- Mode of operation depends on V_g, V_d, V_s
  - $V_{gs} = V_g - V_s$
  - $V_{gd} = V_g - V_d$
  - $V_{ds} = V_d - V_s = V_{gs} - V_{gd}$
- Source and drain are symmetric diffusion terminals
  - By convention, source is terminal at lower voltage
  - Hence $V_{ds} \geq 0$
- nMOS body is grounded. First assume source is 0 too.
- Three regions of operation
  - *Cutoff*
  - *Linear*
  - *Saturation*



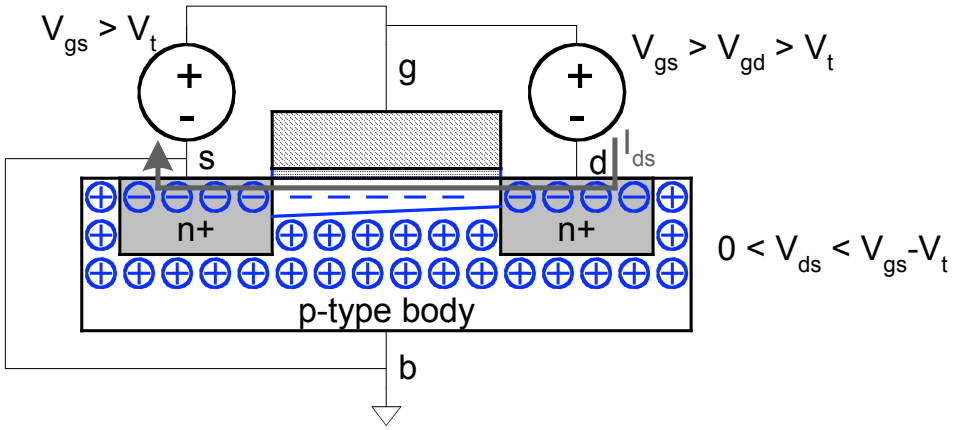
nMOS Cutoff

- No channel formed, so no current flows
- $I_{ds} = 0$



nMOS Linear

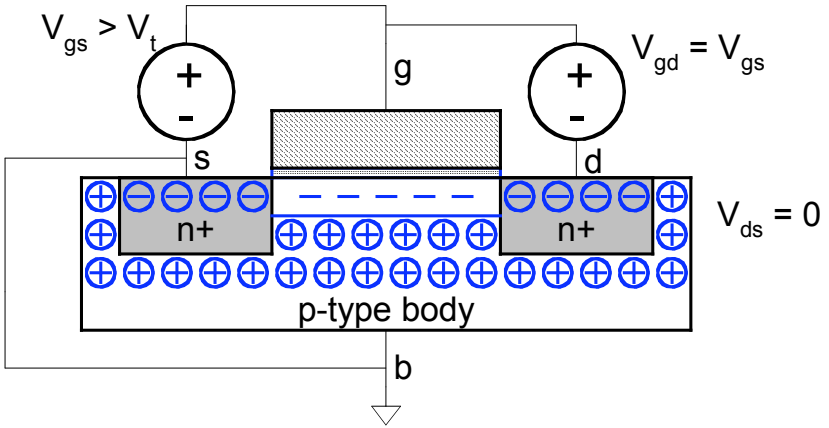
- Channel forms
- Current flows from d to s
  - e^- from s to d
- I_{ds} increases with V_{ds}
- Similar to linear resistor



$V_{gs} > V_t$

$V_{ds} > 0$, but $< (V_{gs} - V_t)$

(current flows)

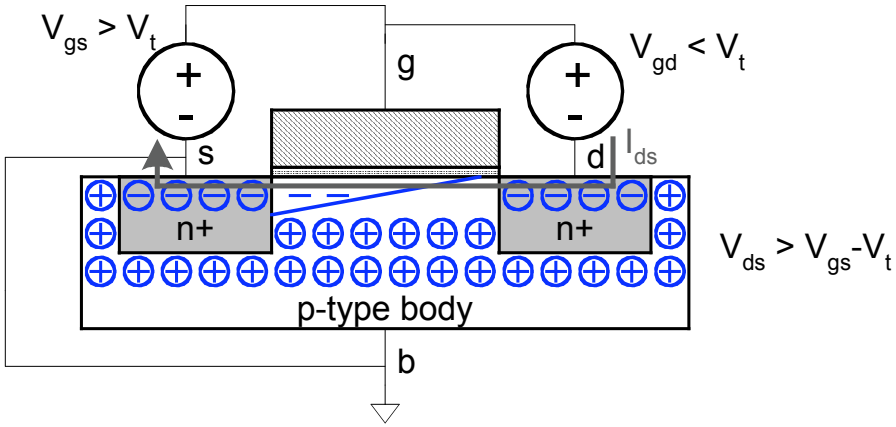


$V_{gs} > V_t$

$V_{ds} = 0$, no current

nMOS Saturation

- Channel pinches off
- I_{ds} independent of V_{ds}
- We say current saturates
- Similar to current source



$$V_{ds} > V_{gs} - V_t$$

Essentially, voltage difference over induced channel fixed at $V_{gs} - V$

(current flows, but saturates)

(or i_{ds} no longer a function of V_{ds})

Outline (part 2)

- Today…
  - nMOS & pMOS I-V characteristics
    - Why (part 1):
      - Quantify - or at least estimate - how we represent & move information
    - Why (part 2):
      - This way we can estimate what happens when we make device smaller -- and in theory, better.
- Possibly today…
  - A very brief discussion of RC delay models
    - Why?
      - Important because delay = one of the 2 performance metrics we care most about.
- Another, "why"
  - Can leverage in 1st HW :-)
    - David Frank talk - starts with 1st principles, extrapolates to practical, chip-level performance

A little bit of foreshadowing

Parameter	Relation	Full Scaling	General Scaling	Fixed-Voltage Scaling
W, L, t_{ox}		1/S	1/S	1/S
V_{dd}, V_t		1/S	1/U	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_{gate}	$C_{ox}WL$	1/S	1/S	1/S
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{sat}	$C_{ox}WV$	1/S	1/U	1
Current Density	I_{sat}/Area	S	S^2/U	S^2
R_{on}	V/I_{sat}	1	1	1
Intrinsic Delay	$R_{on}C_{gate}$	1/S	1/S	1/S
P	$I_{sat}V$	$1/S^2$	$1/U^2$	1
Power Density	P/Area	1	S^2/U^2	S^2

Board digression #0

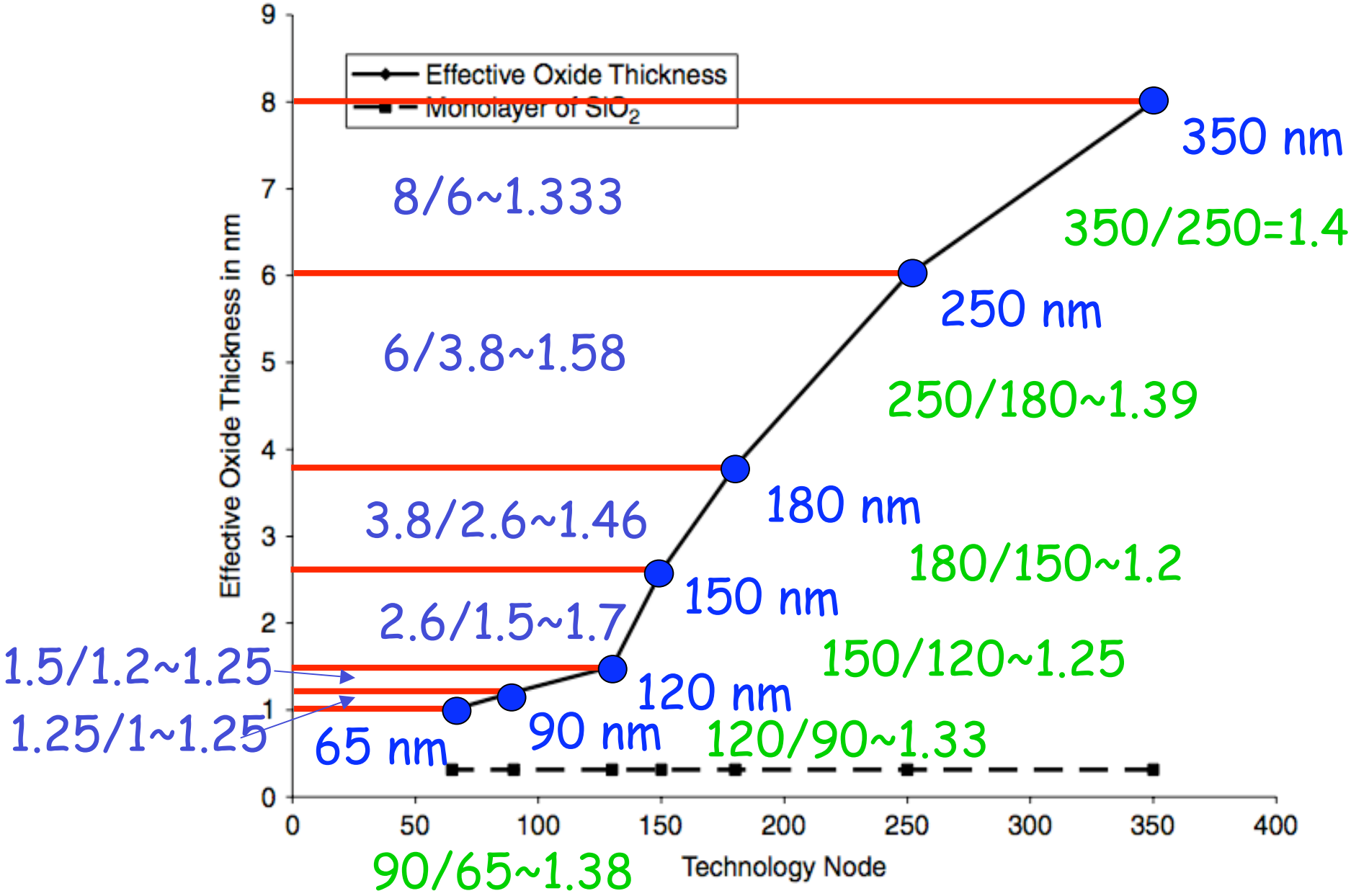
A little bit of foreshadowing

Board digression #0

A little bit of foreshadowing

Board digression #0

t_{ox}


Effective Oxide Thickness
Monolayer of SiO_2
Effective Oxide Thickness in nm
Technology Node
0
1
2
3
4
5
6
7
8
9
0
50
100
150
200
250
300
350
400
350 nm
250 nm
180 nm
150 nm
120 nm
90 nm
65 nm
8/6~1.333
6/3.8~1.58
3.8/2.6~1.46
2.6/1.5~1.7
1.5/1.2~1.25
1.25/1~1.25
350/250=1.4
250/180~1.39
180/150~1.2
150/120~1.25
120/90~1.33
90/65~1.38

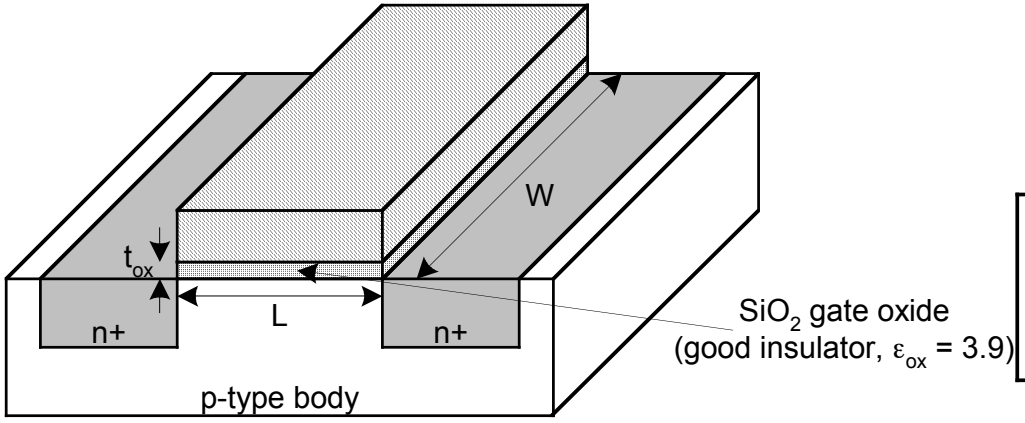
Ok, let's derive some I-V relationships

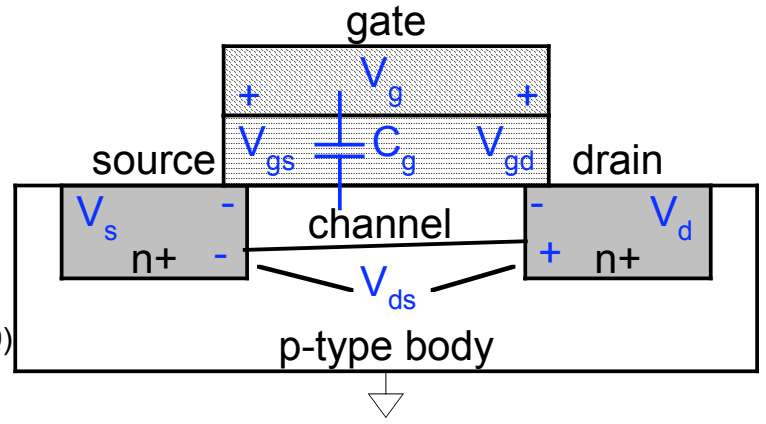
I-V Characteristics

- In Linear region, I_{ds} depends on
  - How much charge is in the channel?
  - How fast is the charge moving?

Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel}$ =

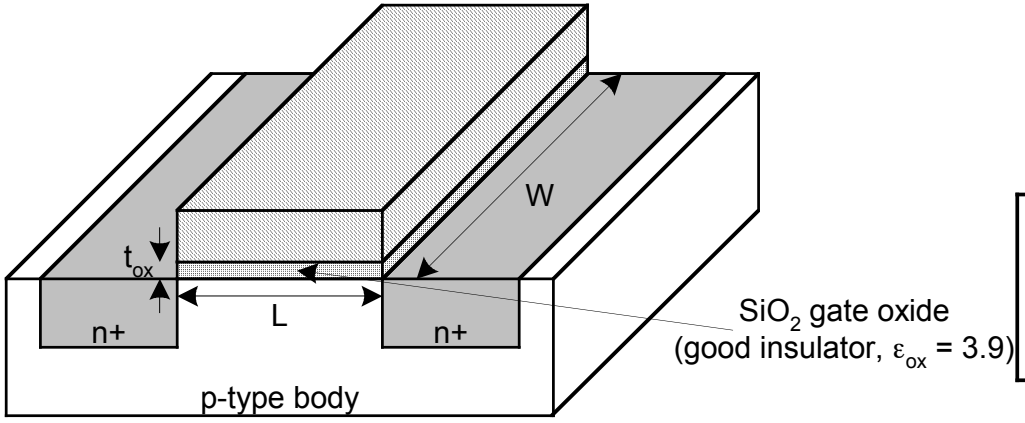


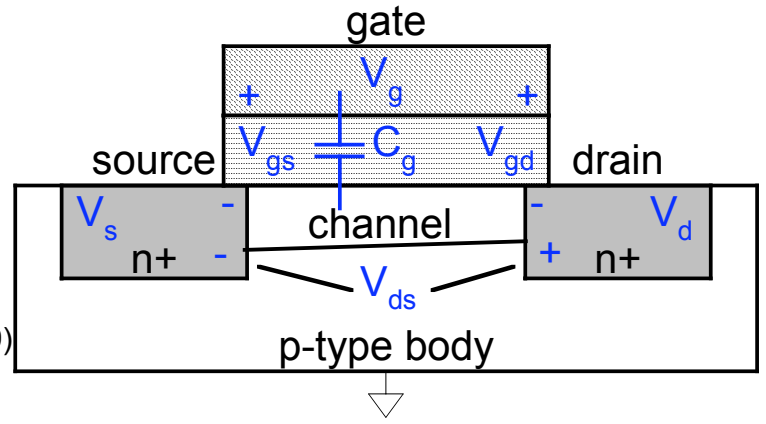


Board digression #1

Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel} = CV$
- $C =$

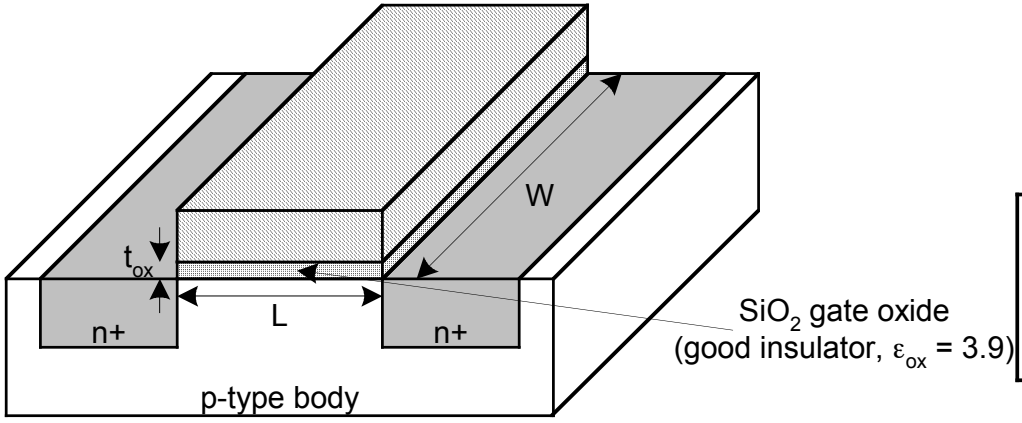


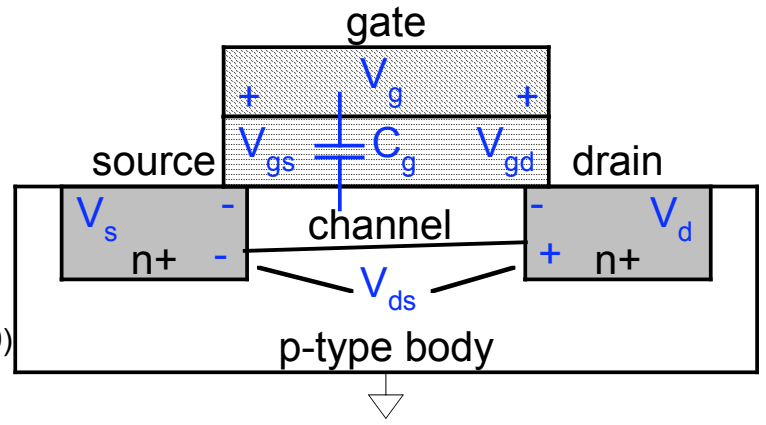


Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel} = CV$
- $C = C_g = \varepsilon_{ox} WL / t_{ox} = C_{ox} WL$
- $V =$

$$C_{ox} = \varepsilon_{ox} / t_{ox}$$

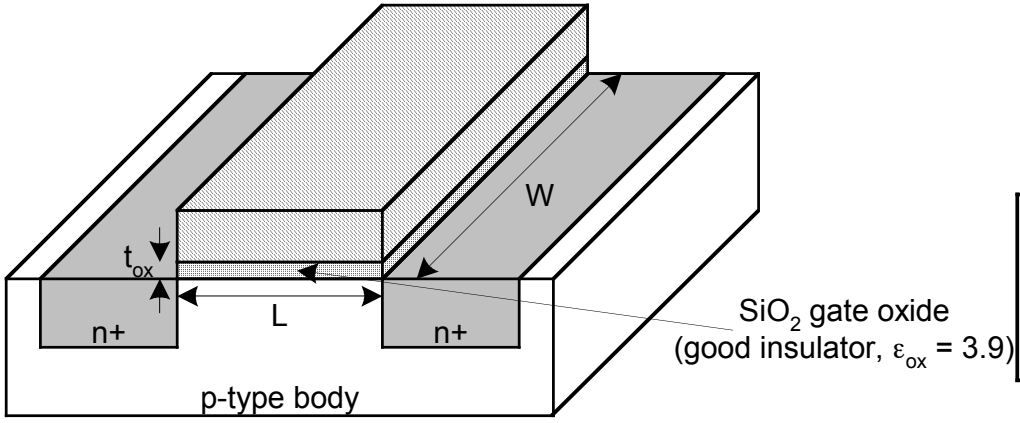


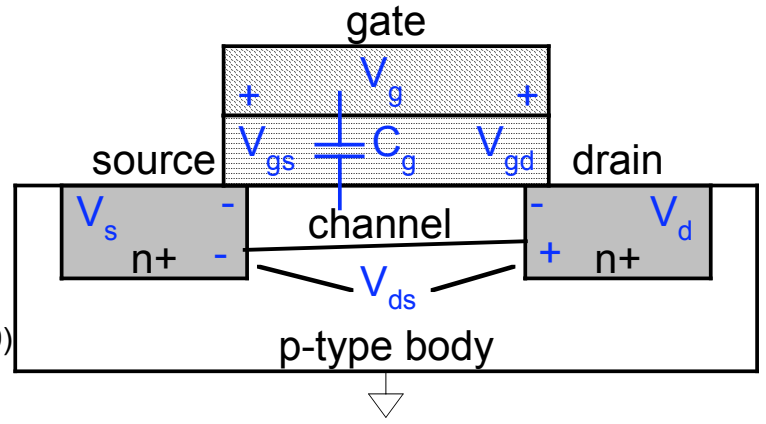


Channel Charge

- MOS structure looks like parallel plate capacitor while operating in inversion
  - Gate – oxide – channel
- $Q_{channel} = CV$
- $C = C_g = \varepsilon_{ox} WL / t_{ox} = C_{ox} WL$
- $V = V_{gc} - V_t = (V_{gs} - V_{ds}/2) - V_t$

$$C_{ox} = \varepsilon_{ox} / t_{ox}$$





Board digression #2

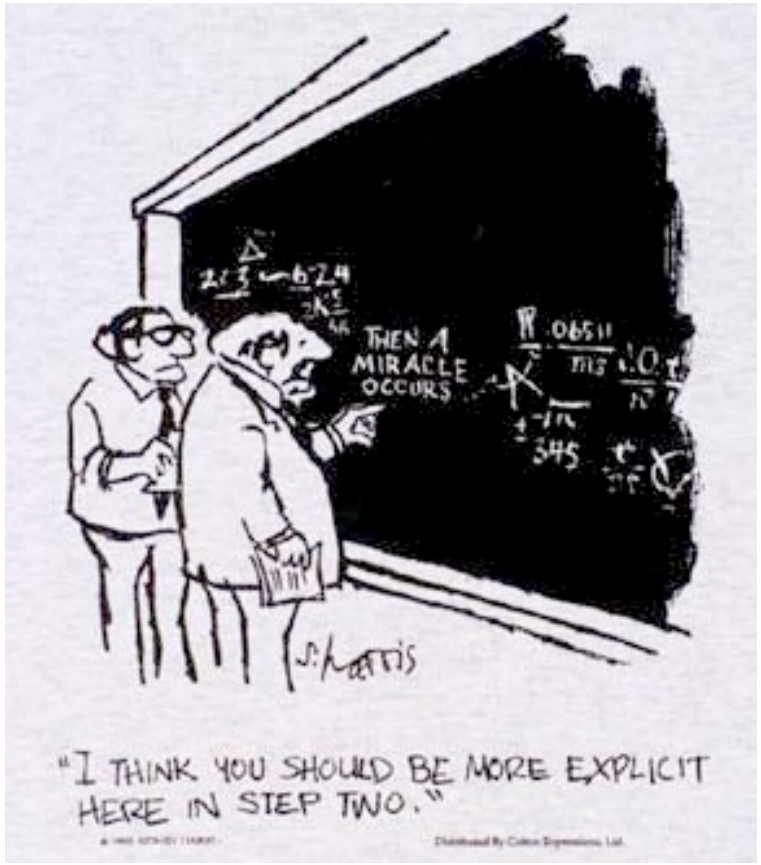
Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- *v* =

Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- $E =$

How I try *not* to teach…



Board digression #3

Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- $E = V_{ds}/L$
- Time for carrier to cross channel:
  - $t =$

Carrier velocity

- Charge is carried by e-
- Carrier velocity *v* proportional to lateral E-field between source and drain
- $v = \mu E$ μ called mobility
- $E = V_{ds}/L$
- Time for carrier to cross channel:
  - $t = L / v$

Board digression #4

nMOS Linear I-V

- Now we know
  - How much charge $Q_{channel}$ is in the channel
  - How much time *t* each carrier takes to cross

$$I_{ds} =$$

nMOS Linear I-V

- Now we know
  - How much charge $Q_{channel}$ is in the channel
  - How much time *t* each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$

$$=$$

nMOS Linear I-V

- Now we know
  - How much charge $Q_{channel}$ is in the channel
  - How much time *t* each carrier takes to cross

$$I_{ds} = \frac{Q_{\text{channel}}}{t}$$

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

$$= \beta\left(V_{gs} - V_t - V_{ds}/2\right)V_{ds}$$

$$\beta = \mu C_{\text{ox}} \frac{W}{L}$$

Board digression #5

Let's go back to…

Parameter	Relation	Full Scaling	General Scaling	Fixed-Voltage Scaling
W, L, t_{ox}		1/S	1/S	1/S
V_{dd}, V_t		1/S	1/U	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_{gate}	$C_{ox}WL$	1/S	1/S	1/S
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{sat}	$C_{ox}WV$	1/S	1/U	1
Current Density	I_{sat}/Area	S	S^2/U	S^2
R_{on}	V/I_{sat}	1	1	1
Intrinsic Delay	$R_{on}C_{gate}$	1/S	1/S	1/S
P	$I_{sat}V$	$1/S^2$	$1/U^2$	1
Power Density	P/Area	1	S^2/U^2	S^2

Board digression #6

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} =$$

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} = \beta \left(V_{gs} - V_t - V_{dsat}/2 \right) V_{dsat}$$

nMOS Saturation I-V

- If $V_{gd} < V_t$, channel pinches off near drain
  - When $V_{ds} > V_{dsat} = V_{gs} - V_t$
- Now drain voltage no longer increases current

$$I_{ds} = \beta\left(V_{gs} - V_t - V_{dsat}/2\right)V_{dsat}$$

$$= \frac{\beta}{2}\left(V_{gs} - V_t\right)^2$$

Board digression #7

nMOS I-V Summary

- ***Shockley* 1st order transistor models**

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ \beta\left(V_{gs} - V_t - V_{ds}/2\right)V_{ds} & V_{ds} < V_{dsat} \quad \text{linear} \\ \frac{\beta}{2}\left(V_{gs} - V_t\right)^2 & V_{ds} > V_{dsat} \quad \text{saturation} \end{cases}$$

Again, let's go back to…

Parameter	Relation	Full Scaling	General Scaling	Fixed-Voltage Scaling
W, L, t_{ox}		1/S	1/S	1/S
V_{dd}, V_t		1/S	1/U	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_{gate}	$C_{ox}WL$	1/S	1/S	1/S
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{sat}	$C_{ox}WV$	1/S	1/U	1
Current Density	I_{sat}/Area	S	S^2/U	S^2
R_{on}	V/I_{sat}	1	1	1
Intrinsic Delay	$R_{on}C_{gate}$	1/S	1/S	1/S
P	$I_{sat}V$	$1/S^2$	$1/U^2$	1
Power Density	P/Area	1	S^2/U^2	S^2

Board digression #8

Look at I_{DS} in context of scaling

Board digression #8

Board digression #8