ESE3700: Circuit-Level Modeling, Design and Optimization for Digital Systems

Lec 6: February 12, 2024

MOS Transistor Operating Regions

Part 2, Parasitics

Today

- Operating Regions
  - Resistive
  - Saturation
  - Subthreshold
  - Velocity Saturation
- Short Channel Effects
  - V_{th}
  - Drain Induced Barrier Lowering
- Capacitance

Velocity Saturation

Carrier Velocity

- Model assumes carrier velocity increases with field
  - Increases with voltage proportionally to mobility

$$v = \mu_n \cdot \frac{V_{DS}}{L_{eff}} = \left(\frac{\mu_n}{L_{eff}} \right) V_{DS}$$

Carrier Velocity

- Model assumes carrier velocity increases with field
  - Increases with voltage proportionally to mobility

$$v = \mu_n \cdot \frac{V_{DS}}{L_{eff}} = \left(\frac{\mu_n}{L_{eff}} \right) V_{DS}$$

(3×10^8 m/sec)

Preclass 1

- (a) What is the electrical field in the channel?

$$L_{eff} = 25nm, V_{DS} = 1V$$

$$\text{Uniform Field} = \frac{V_{DS}}{L_{eff}}$$

- Velocity:

$$v = F \cdot \mu_n$$

- Electron mobility: $\mu_n = 500 cm^2 / (V \cdot s)$
- (b) What is the electron velocity?

Moving Charge

$$I = \left(\frac{1}{R}\right)V$$

- I increases
  linearly in V

- What's I?

Moving Charge

$$I = \left(\frac{1}{R}\right)V$$

- I increases
  linearly in V

- What's I?
  - $\Delta Q/\Delta t$
  - Speed at which charge moves

Moving Charge

$$I = \left(\frac{1}{R}\right)V$$

- I increases linearly in V

- What's I?
  - ΔQ/Δt
  - Speed at which charge moves

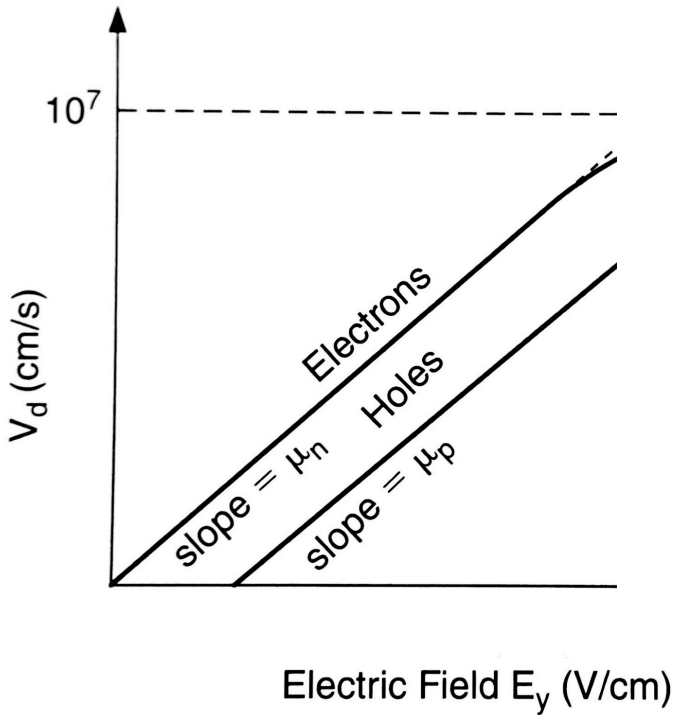
$$Field = \frac{V_{DS}}{L_{eff}}, v = \mu_n \cdot F$$

$$v = \mu_n \cdot \frac{V_{DS}}{L_{eff}} = \left(\frac{\mu_n}{L_{eff}}\right)V_{DS}$$

- Velocity increases linearly in V

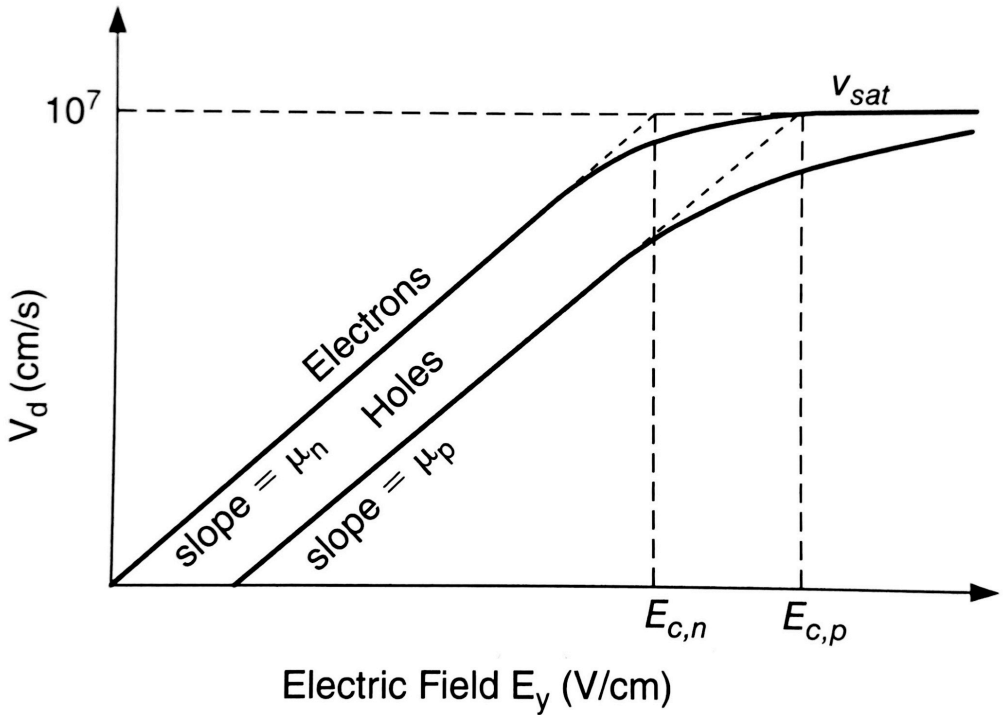
- What's a moving electron?

Carrier Velocity



- Velocity –
  - increases for increasing field with slope of mobility

Carrier Velocity



- Velocity –
  - increases for increasing field with slope of mobility
  - saturates for increasing field
    - More likely to hit the critical field in short channel

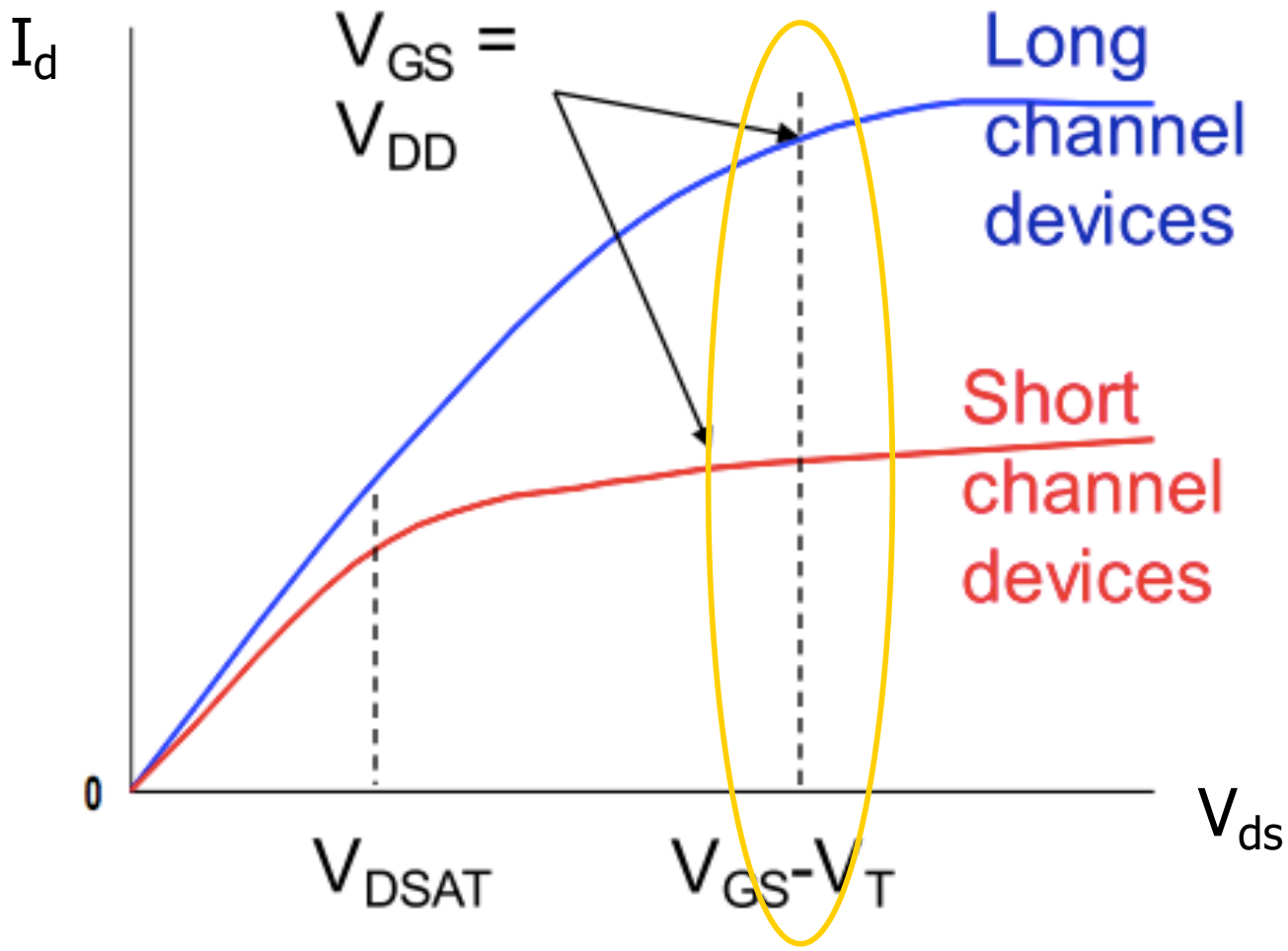
Short Channel

- Model assumes carrier velocity increases with field
  - Increases with voltage proportionally to mobility
- There is a limit to how fast carriers can move
  - Limited by scattering effects
    - ~ 10^5m/s
- Encounter *velocity saturation* when channel short
  - Modern processes, L is short enough to reach this region of operation

Velocity Saturation (Preclass 1)

- (c) At what voltage do we hit the speed limit 10^5m/s?
  - L_{eff}=25nm, V_{ds}=1V
  - V_{DSAT} = voltage at which velocity (current) saturates

Velocity Saturation



Velocity Saturation

- Our current model equation:

$$I_{DS} = \mu_n C_{OX} \left(\frac{W}{L} \right) \left[\left(V_{GS} - V_{th} \right) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

- Once velocity saturates:

Velocity Saturation

- Our current model equation:

$$I_{DS} = \mu_n C_{OX} \left(\frac{W}{L}\right)\left[\left(V_{GS} - V_{th}\right)V_{DS} - \frac{V_{DS}^2}{2}\right]$$

- Once velocity saturates:

$$V_{DS} = V_{DSAT} \Rightarrow I_{DS} = \mu_n C_{OX} \left(\frac{W}{L}\right)\left[\left(V_{GS} - V_{th}\right)V_{DSAT} - \frac{V_{DSAT}^2}{2}\right]$$

Velocity Saturation

- Our current model equation:

$$I_{DS} = \mu_n C_{OX} \left(\frac{W}{L} \right) \left[\left(V_{GS} - V_{th} \right) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

- Once velocity saturates:

$$V_{DS} = V_{DSAT} \Rightarrow I_{DS} = \mu_n C_{OX} \left(\frac{W}{L} \right) \left[\left(V_{GS} - V_{th} \right) V_{DSAT} - \frac{V_{DSAT}^2}{2} \right]$$

$$I_{DS} = \left(\mu_n \frac{V_{DSAT}}{L} \right) C_{OX} W \left[\left(V_{GS} - V_{th} \right) - \frac{V_{DSAT}}{2} \right]$$

Velocity Saturation

- Our current model equation:

$$I_{DS} = \mu_n C_{OX} \left(\frac{W}{L} \right) \left[\left(V_{GS} - V_{th} \right) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

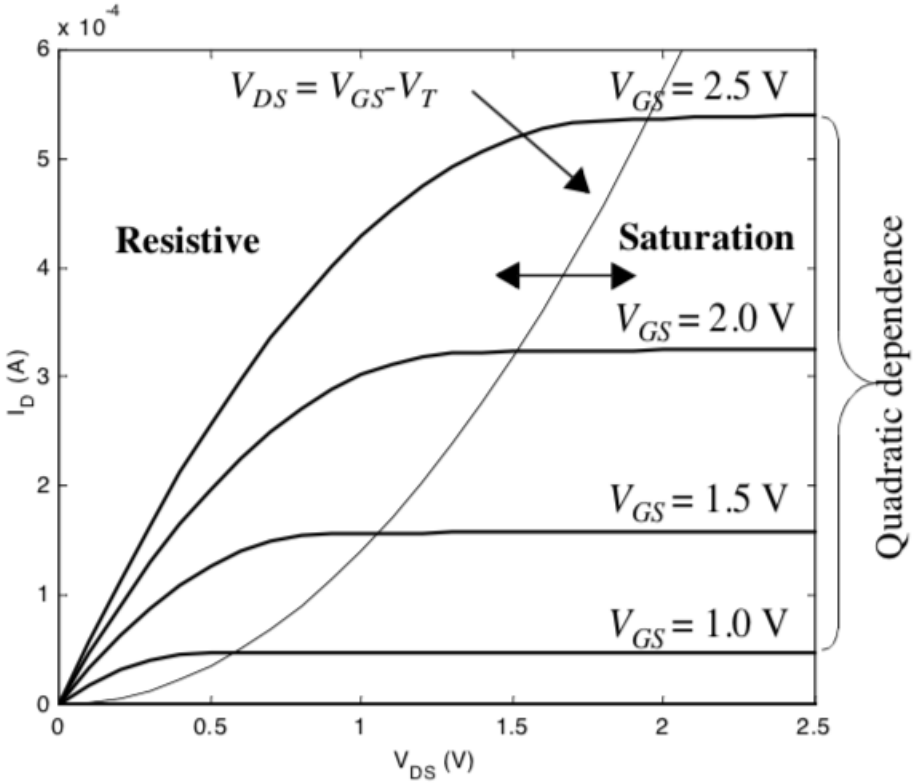
- Once velocity saturates:

$$V_{DS} = V_{DSAT} \Rightarrow I_{DS} = \mu_n C_{OX} \left(\frac{W}{L} \right) \left[\left(V_{GS} - V_{th} \right) V_{DSAT} - \frac{V_{DSAT}^2}{2} \right]$$

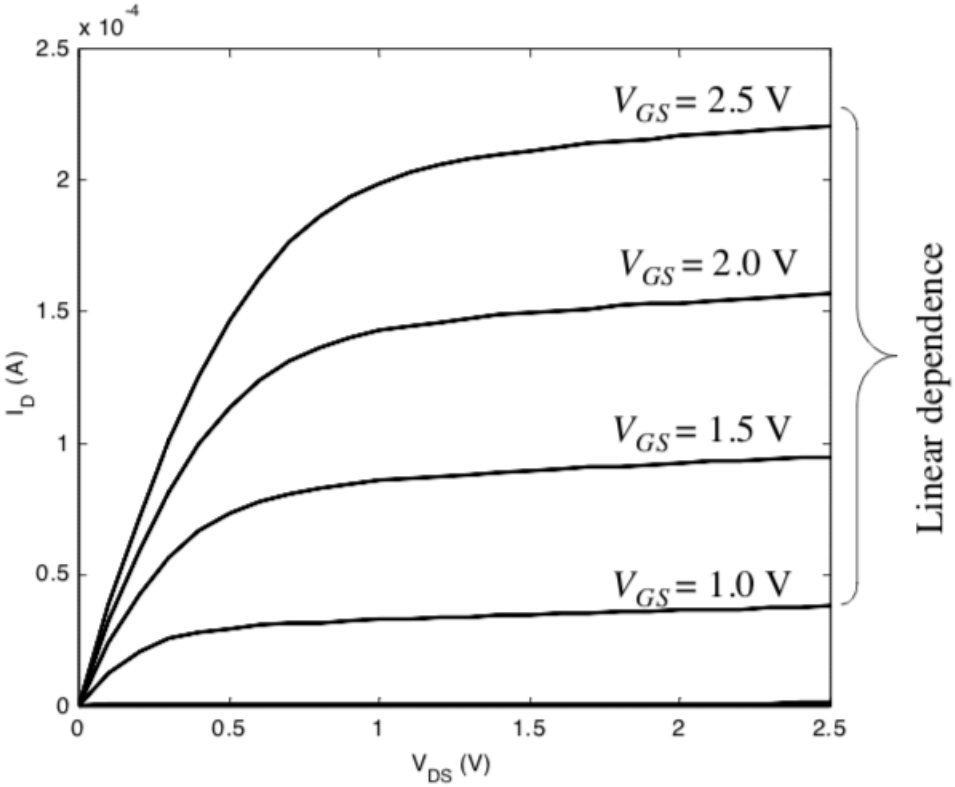
$$I_{DS} = \left(\mu_n \frac{V_{DSAT}}{L} \right) C_{OX} W \left[\left(V_{GS} - V_{th} \right) - \frac{V_{DSAT}}{2} \right]$$

$$I_{DS} \approx v_{sat} C_{OX} W \left[\left(V_{GS} - V_{th} \right) - \frac{V_{DSAT}}{2} \right]$$

Velocity Saturation



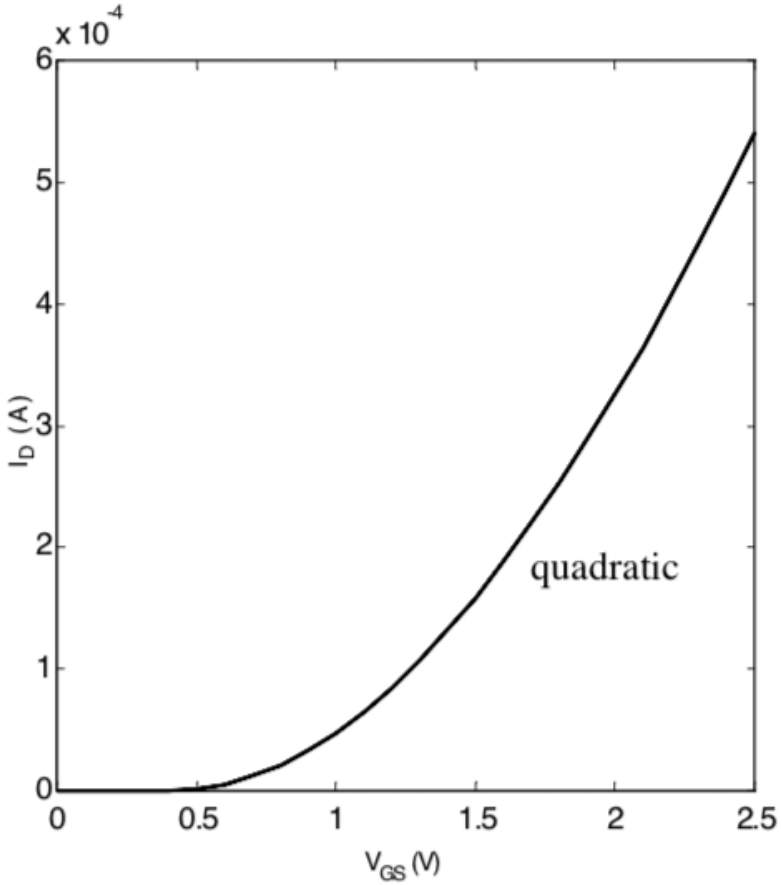
(a) Long-channel transistor ($L_d = 10$ μm)



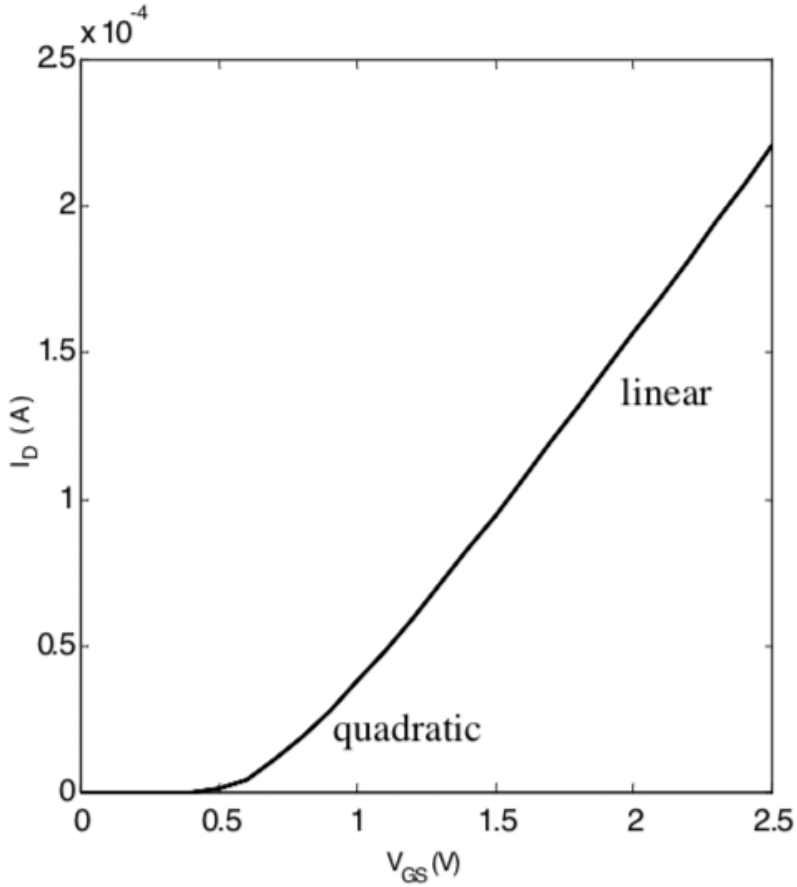
(b) Short-channel transistor ($L_d = 0.25$ μm)

- Long Channel
- Short Channel

Velocity Saturation



(a) Long-channel device (L_d = 10 μm)



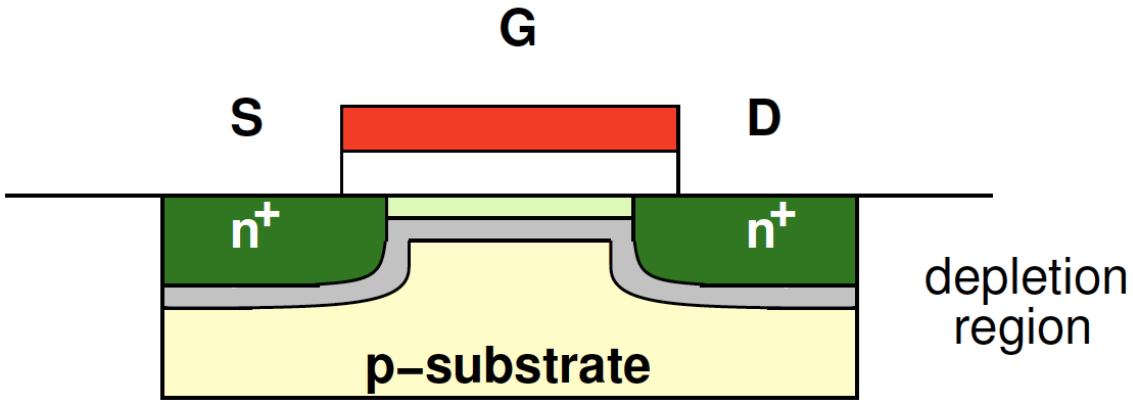
(b) Short-channel device (L_d = 0.25 μm)

- Long Channel
- Short Channel

Velocity Saturation

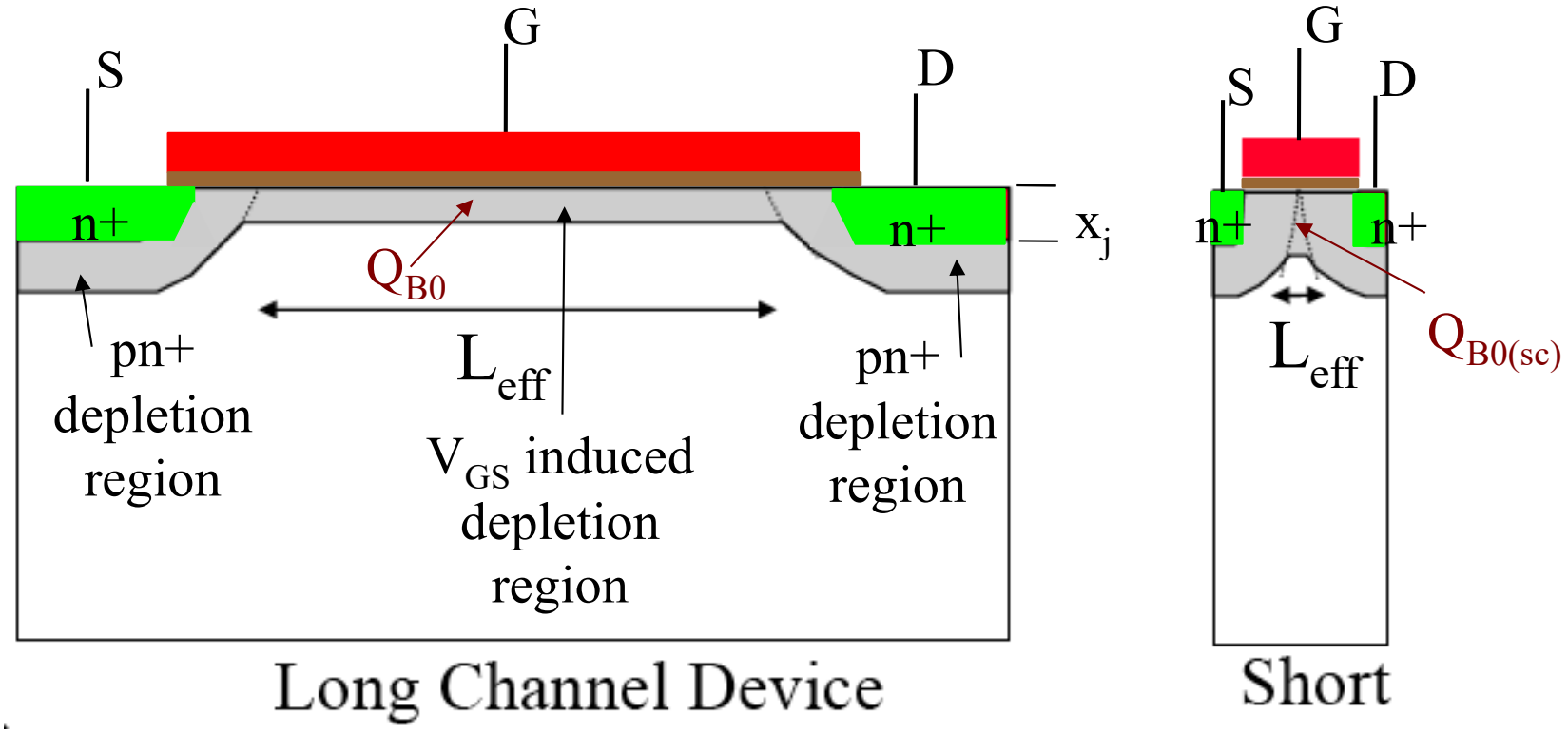
- Once velocity saturates we can still increase current with parallelism
  - Effectively make a wider device

$$I_{DS} \approx v_{sat} C_{OX} W \left[\left(V_{GS} - V_{th} \right) - \frac{V_{DSAT}}{2} \right]$$



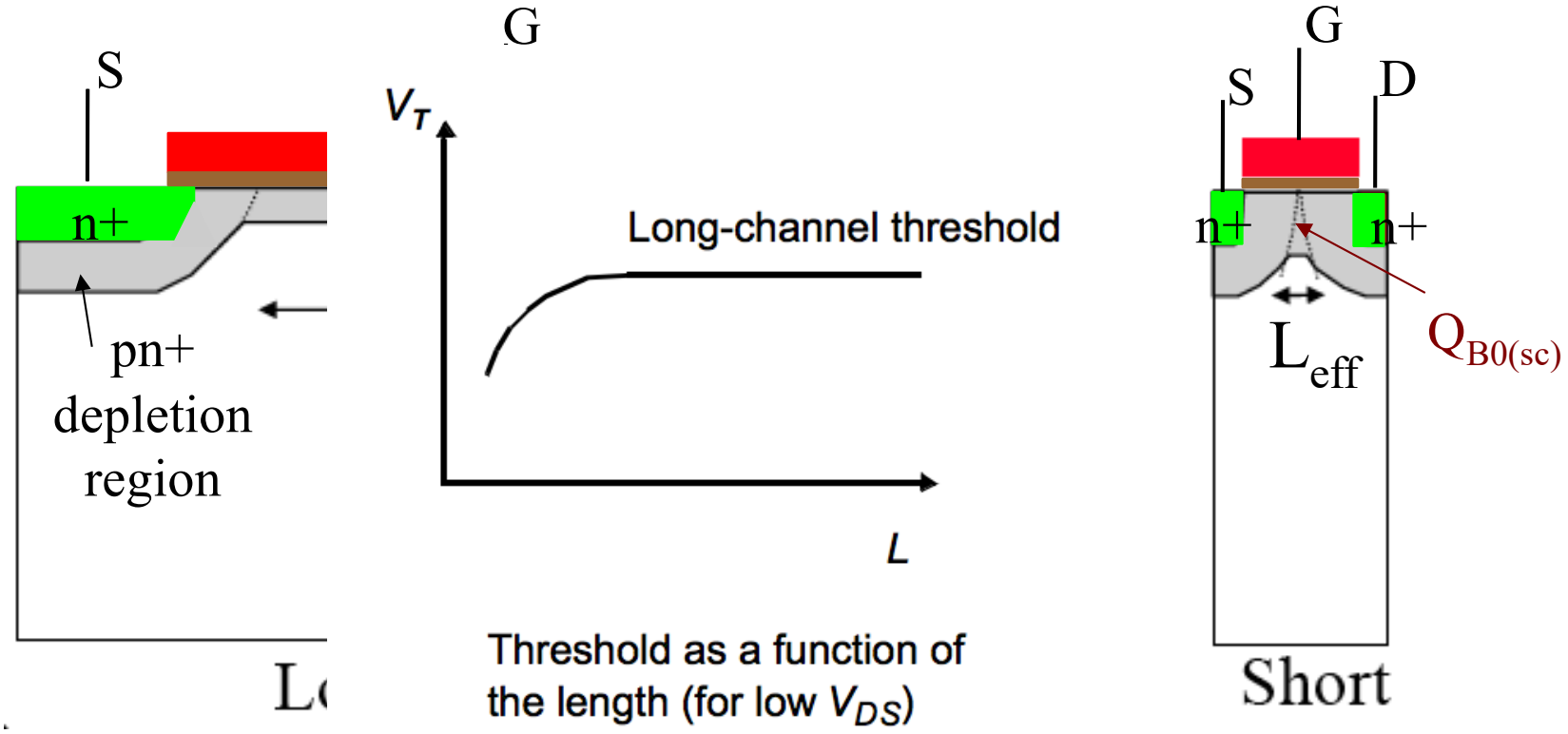
Threshold

Short Channel Effects – V_T Reduction



V_{T0} (short channel) = V_{T0} - ΔV_{T0}

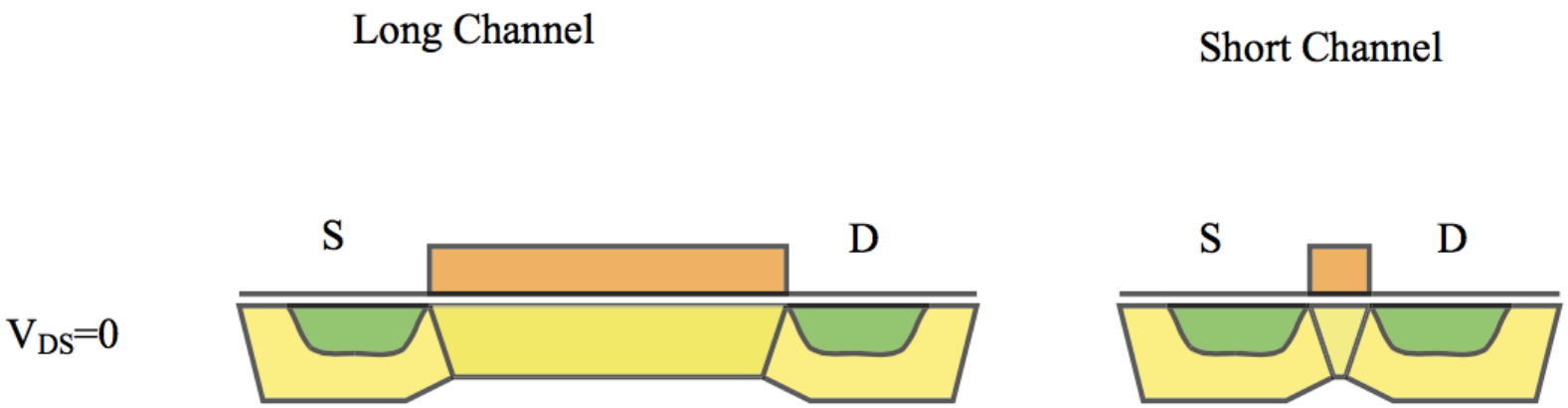
Short Channel Effects – V_T Reduction



V_{T0} (short channel) = V_{T0} - ΔV_{T0}

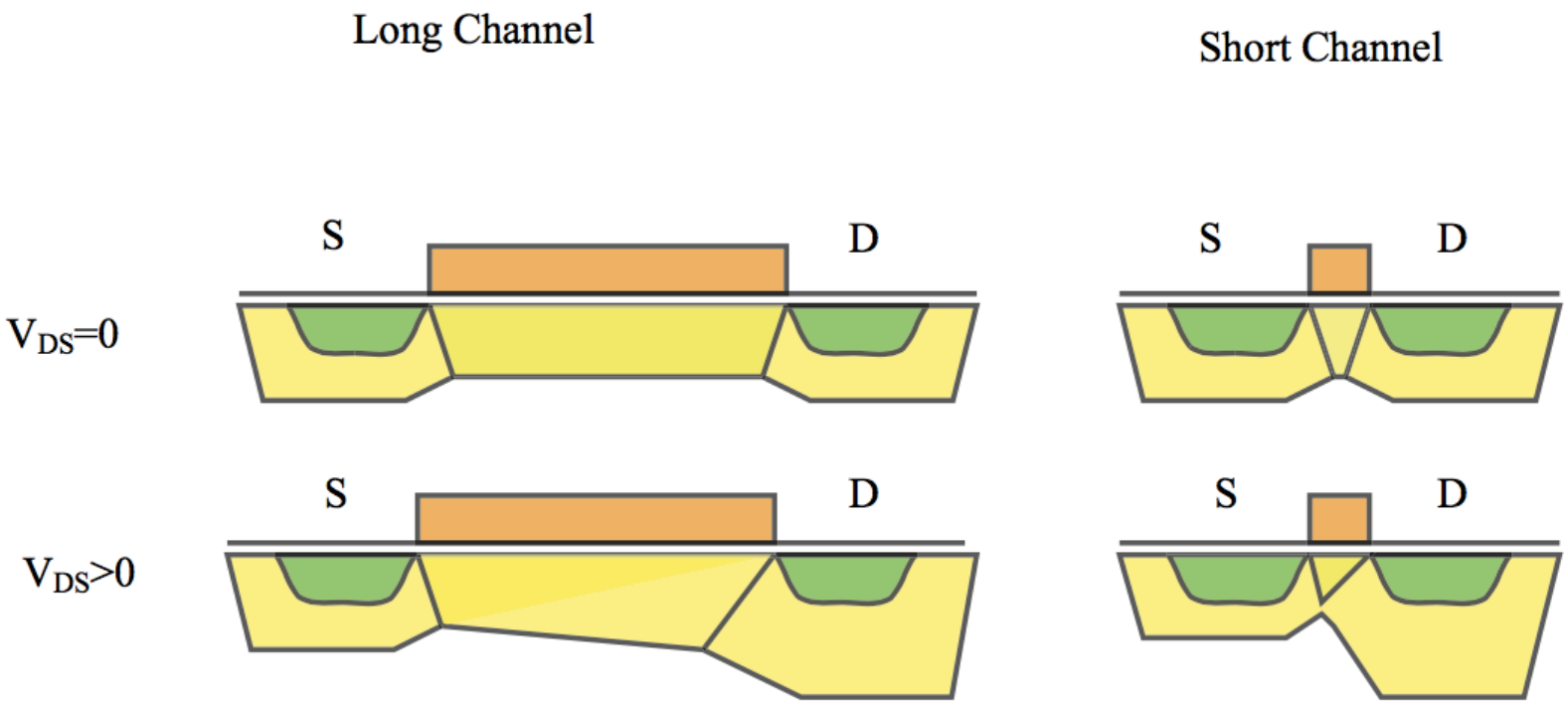
Short Channel Effects - DIBL

- Drain Induced Barrier Lowering
  - V_T Reduction with Drain Bias



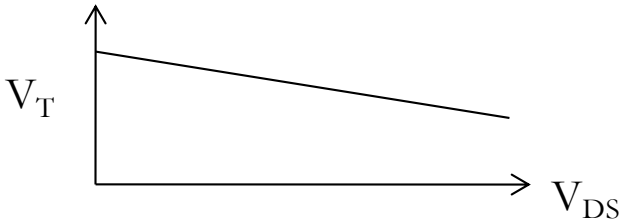
Short Channel Effects - DIBL

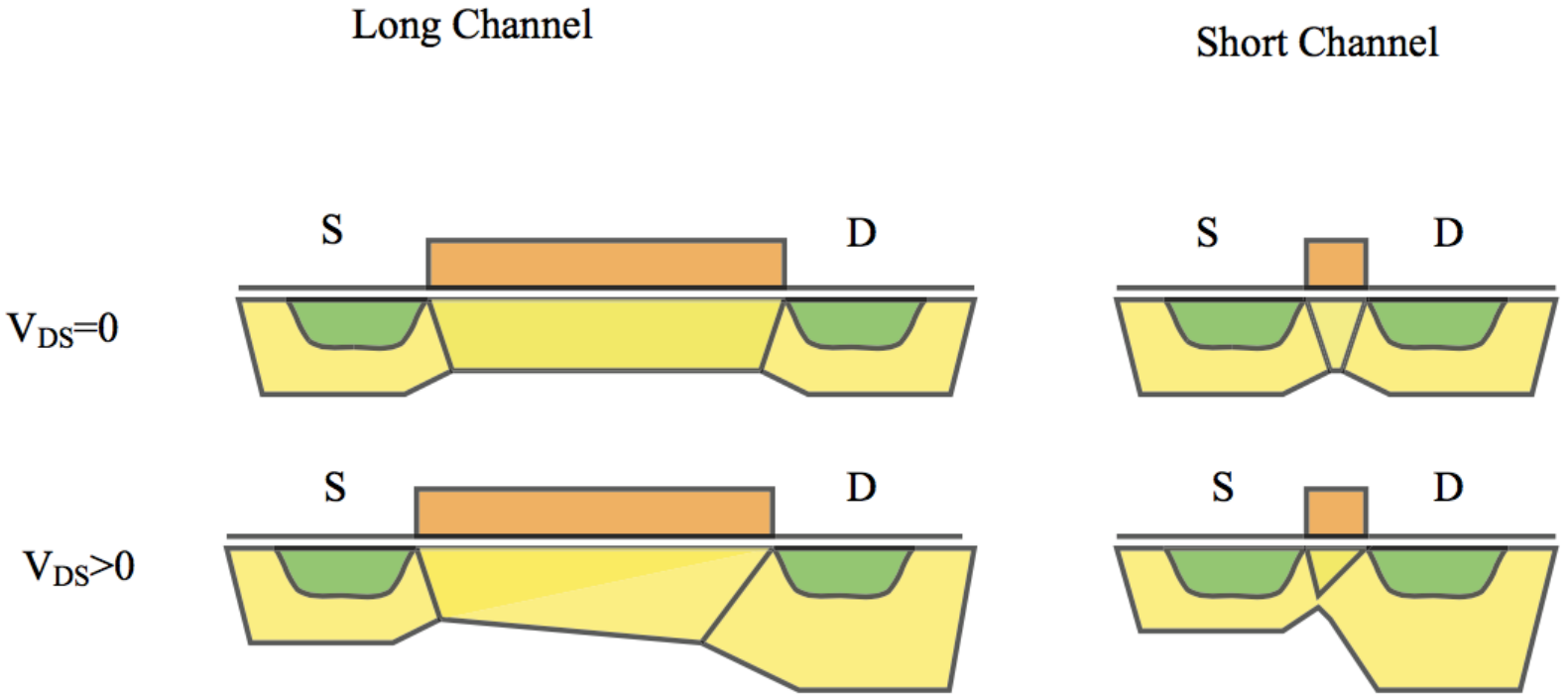
- Drain Induced Barrier Lowering
  - V_T Reduction with Drain Bias



Short Channel Effects - DIBL

- Drain Induced Barrier Lowering
  - V_T Reduction with Drain Bias

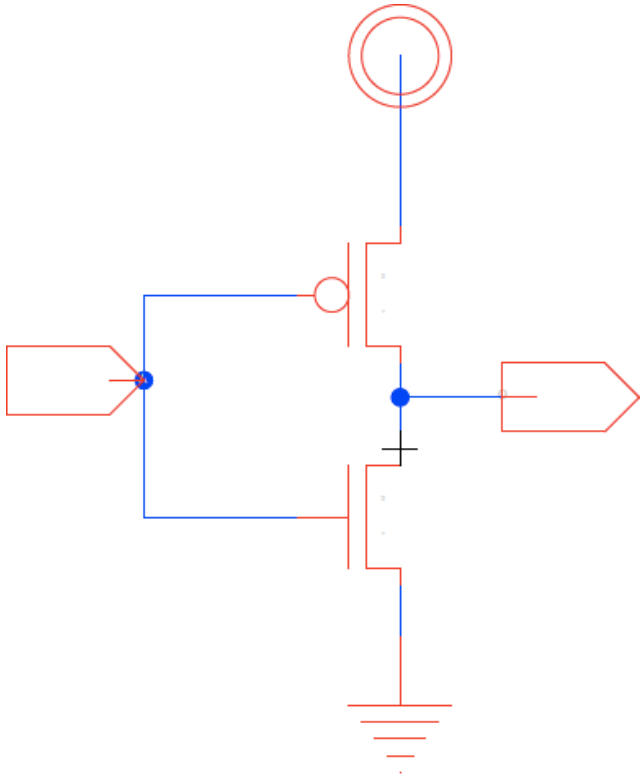




Threshold Reduction Impact

In a Gate?

- What does it impact most?
  - With Vin=Vdd, which device has large V_{ds}?
  - How does this effect operation?
    - Leakage?
    - Speed of switching?



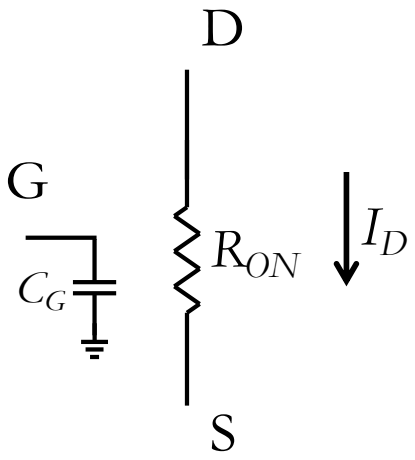
Capacitance

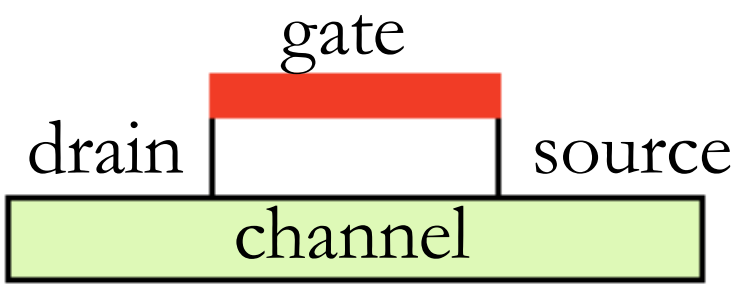
Simplified Design Flow

- Design a circuit to perform a function with specified minimum speed and optimized power (minimized with an upper bound)
  - Zero order model to design topology
  - First order model to meet speed spec
    - Rise/fall times, propagation delay, gate capacitance, output stage equivalent resistance
  - Transistor IV curves
    - Iterative SPICE simulation – tweak knobs to optimize for power (switching (dynamic), leakage (static), etc.)

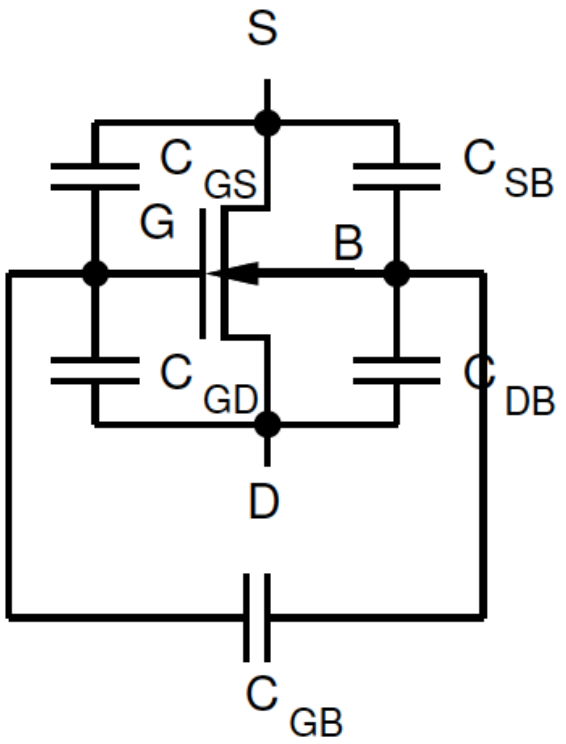
Capacitance

- First order: gate input looks like a capacitor





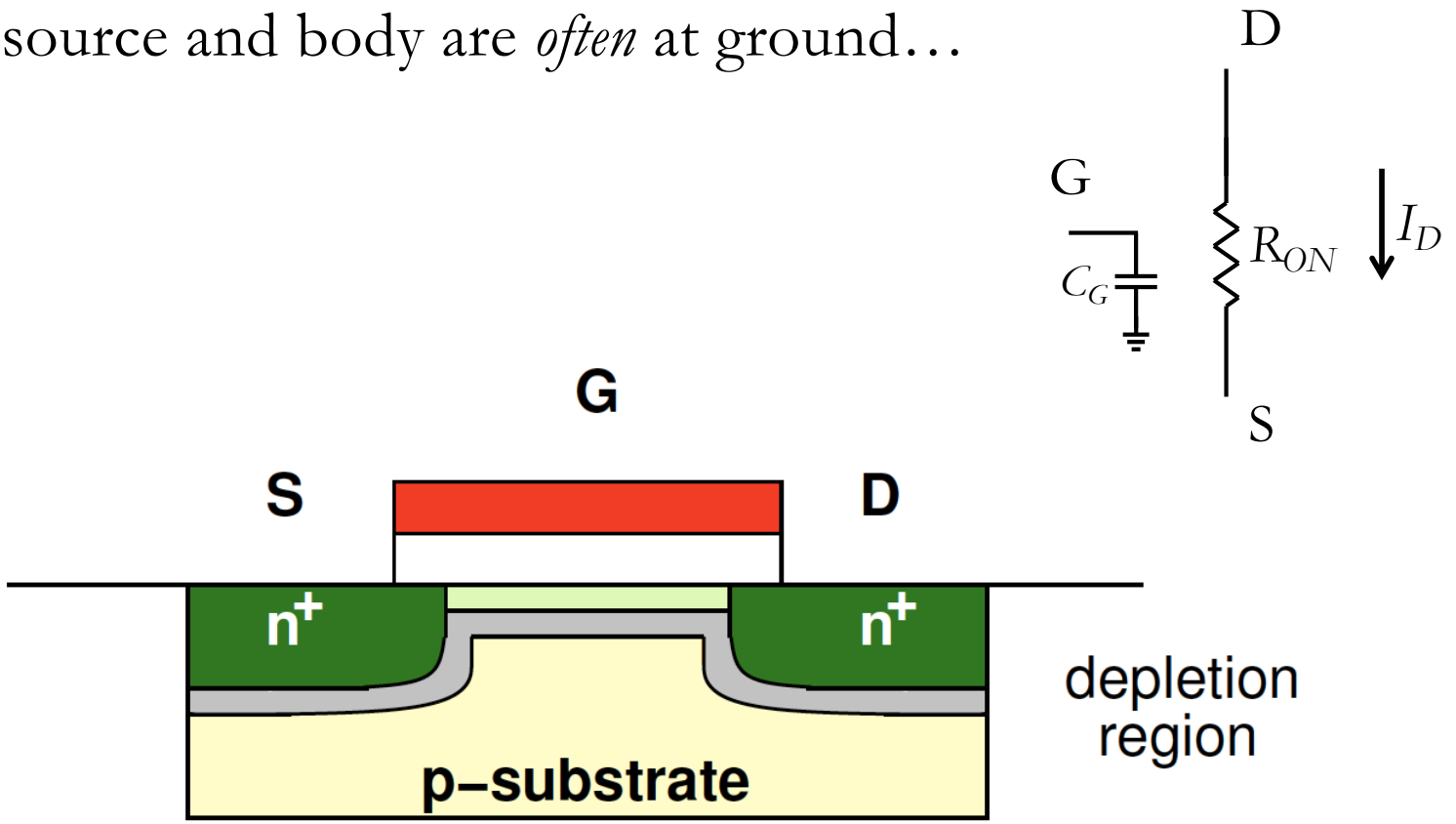
- Today:
  - Capacitance is not constant
  - Capacitance not physically to gnd
    - Modeled as such



Capacitance Setup

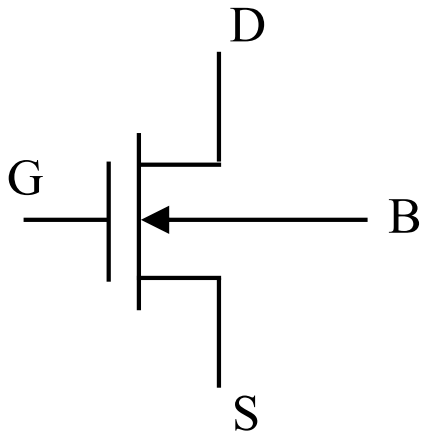
Capacitance

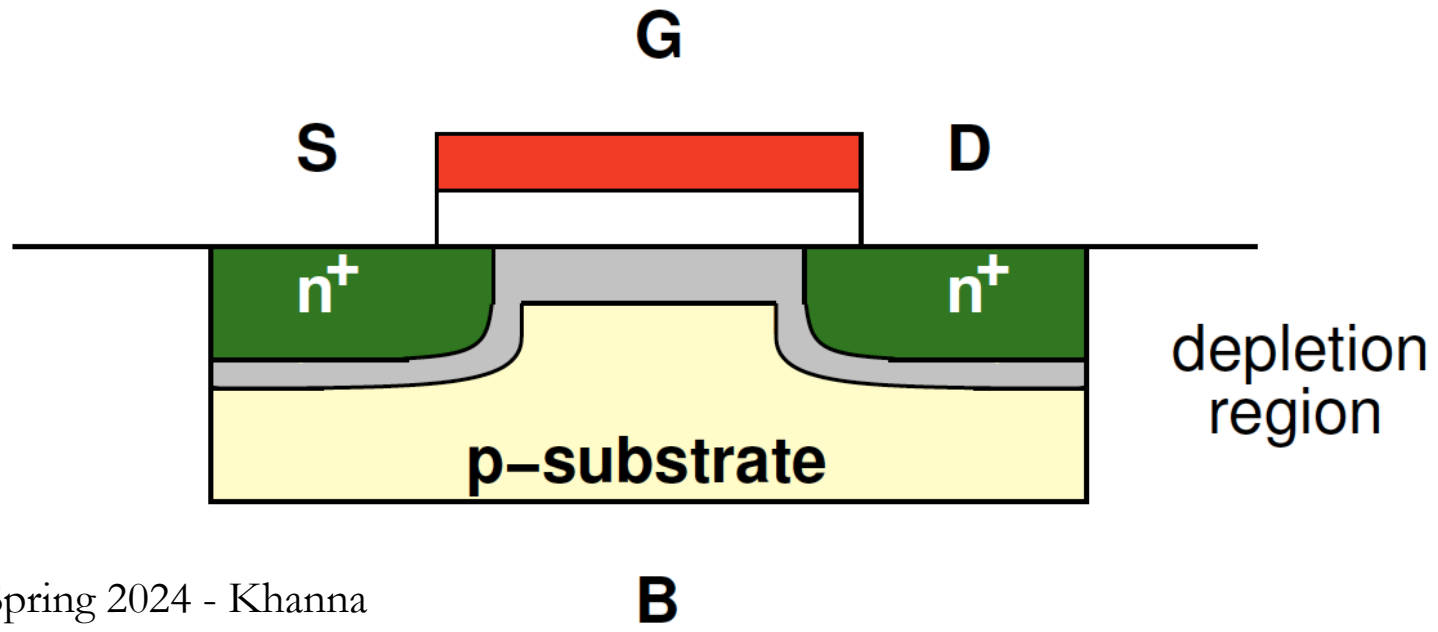
- Modeled gate with a capacitor to ground
- …but ground isn't really one of our terminals
  - Don't connect directly to it
  - …source and body are *often* at ground…



Capacitance (Preclass 2)

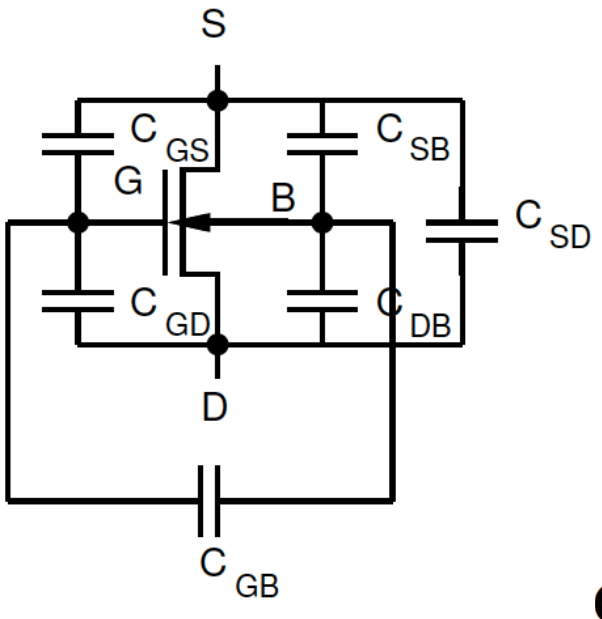
- Four Terminals
- How many combinations?
  - 4 things taken 2 at a time?

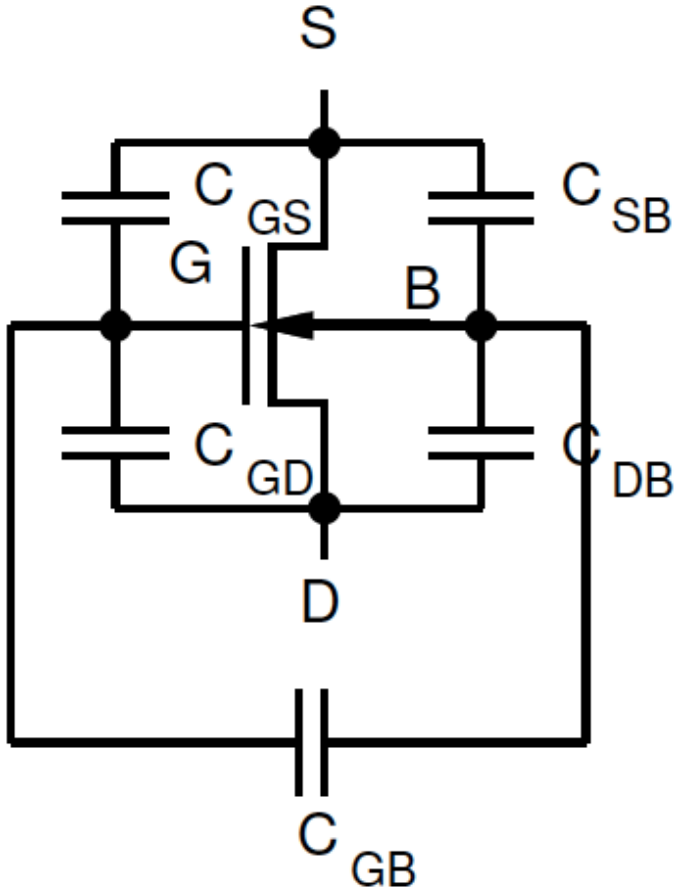


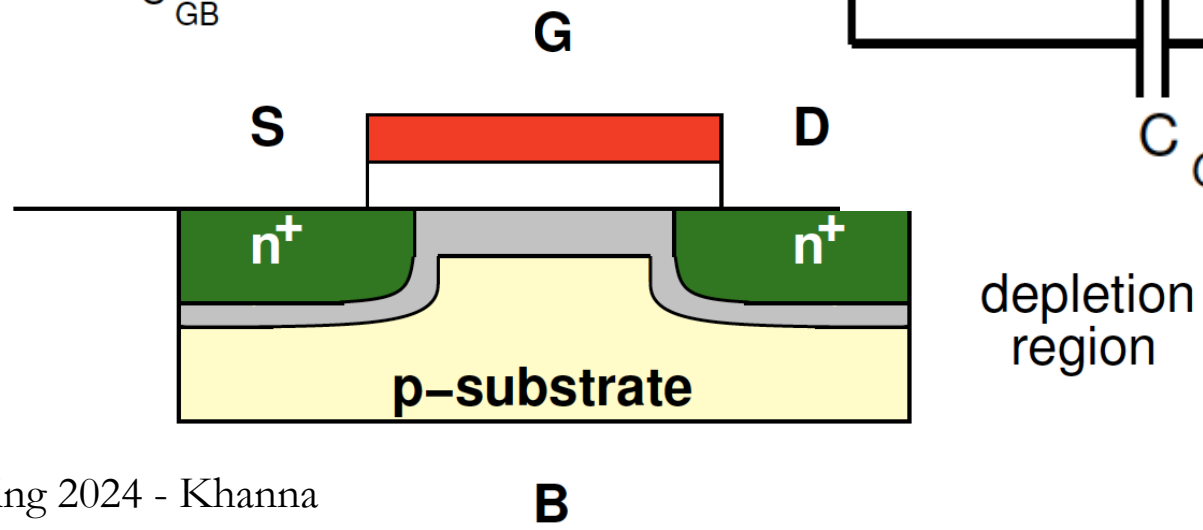


Capacitances

- GS, GB, GD, SB, DB, SD

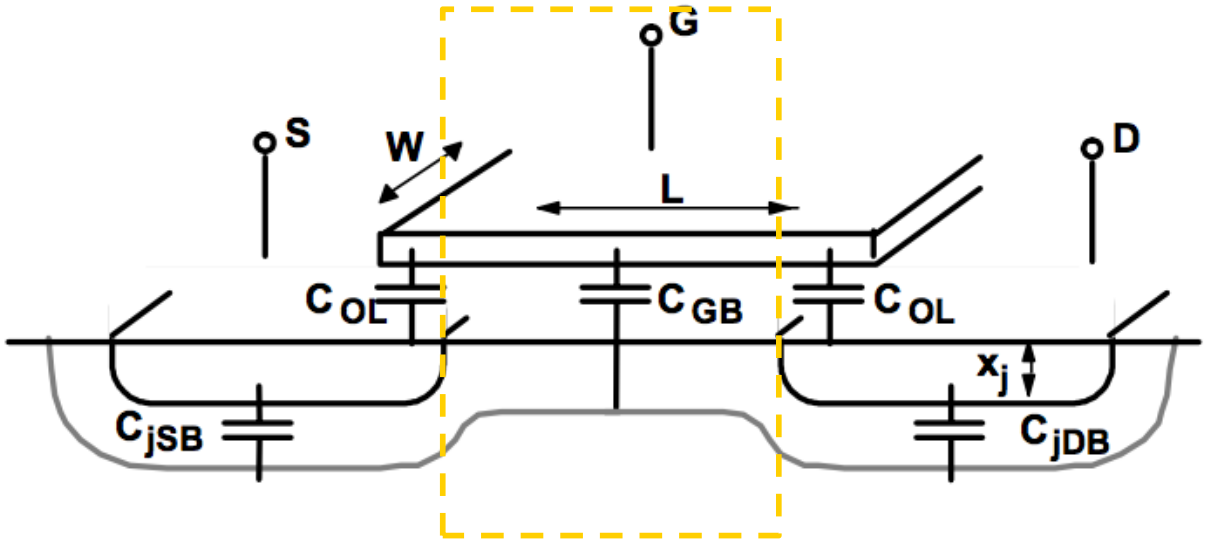






Capacitance Decomposition

MOSFET Parasitic Capacitance

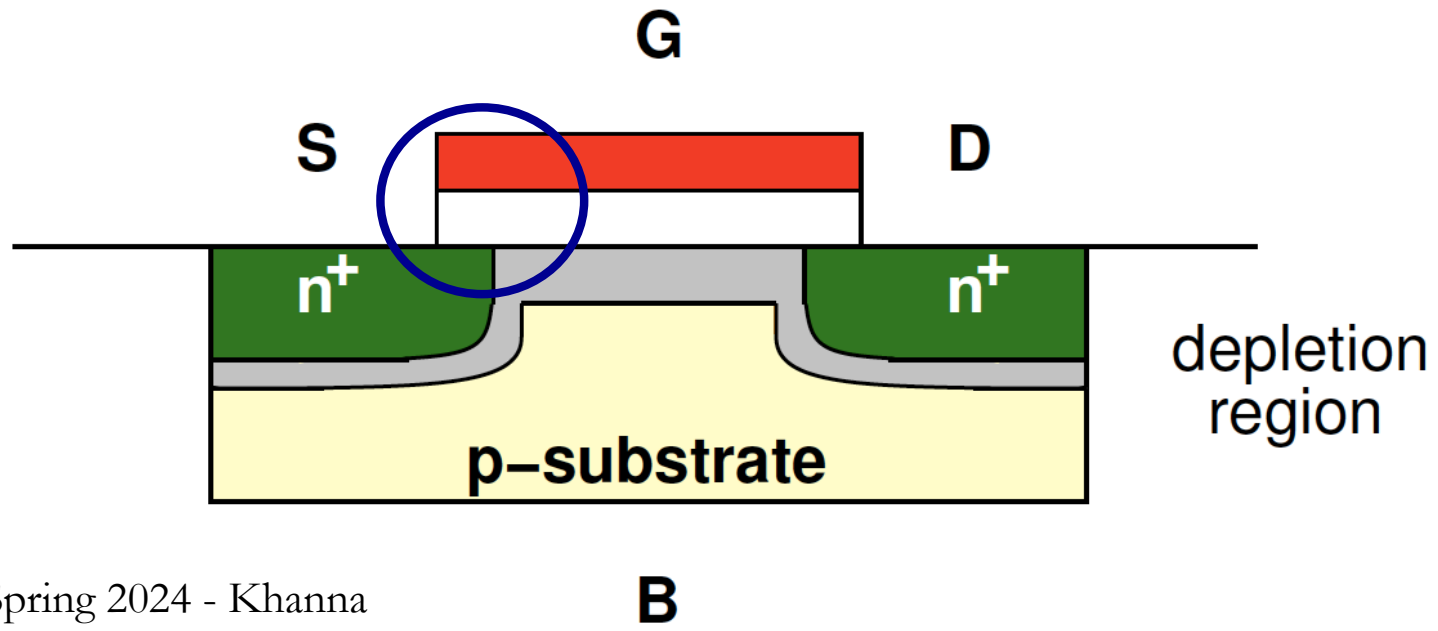


- Any two conductors separated by an insulator form a parallel-plate capacitor
- Two types
  - Extrinsic – Outside the box (e.g. junction, overlap)
  - Intrinsic – Inside the box (e.g. gate-to-channel)

Overlap Capacitance

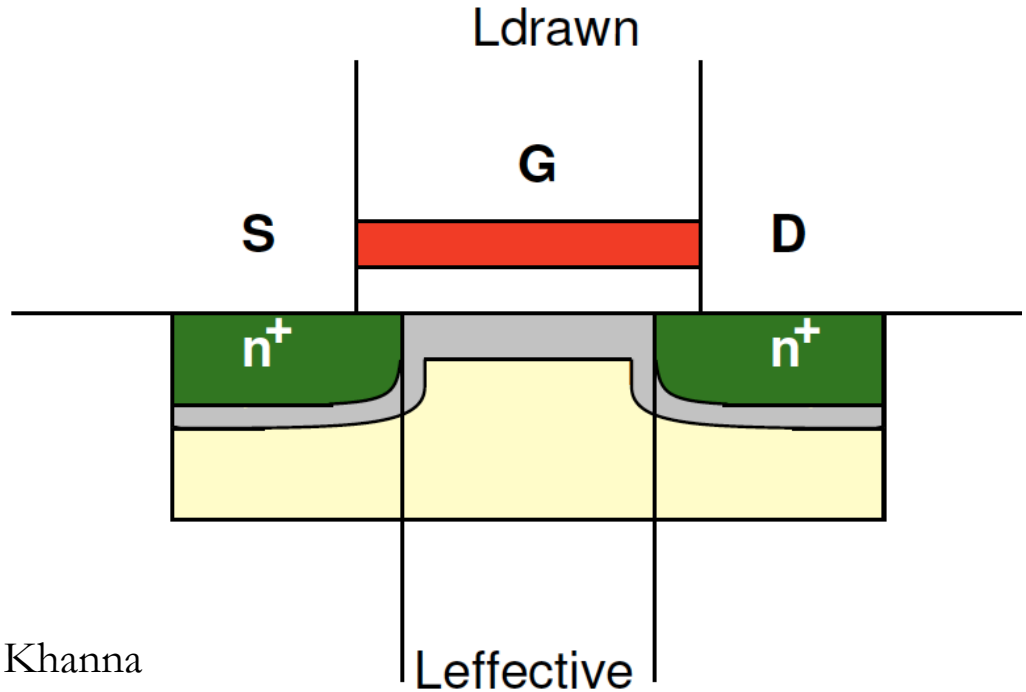
Overlap

- What is the capacitive implication of gate/source and gate/drain overlap?

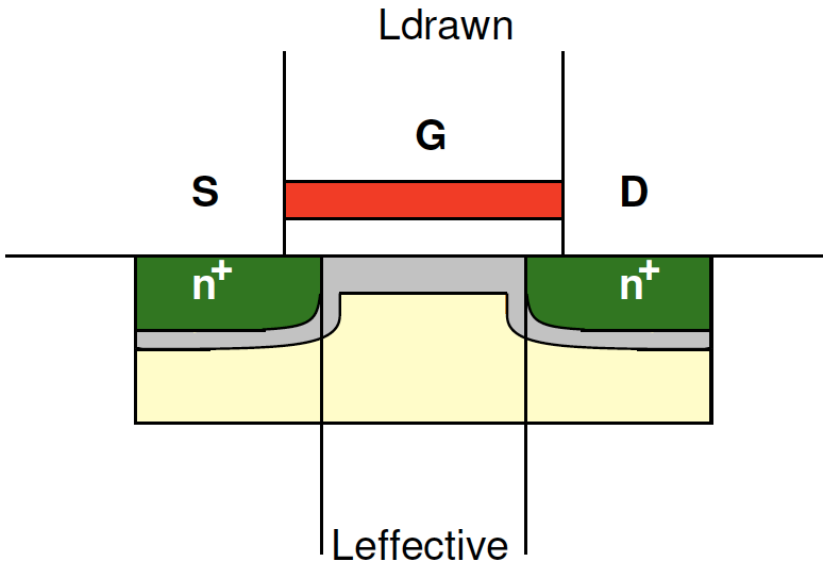


Overlap

- Length of overlap?



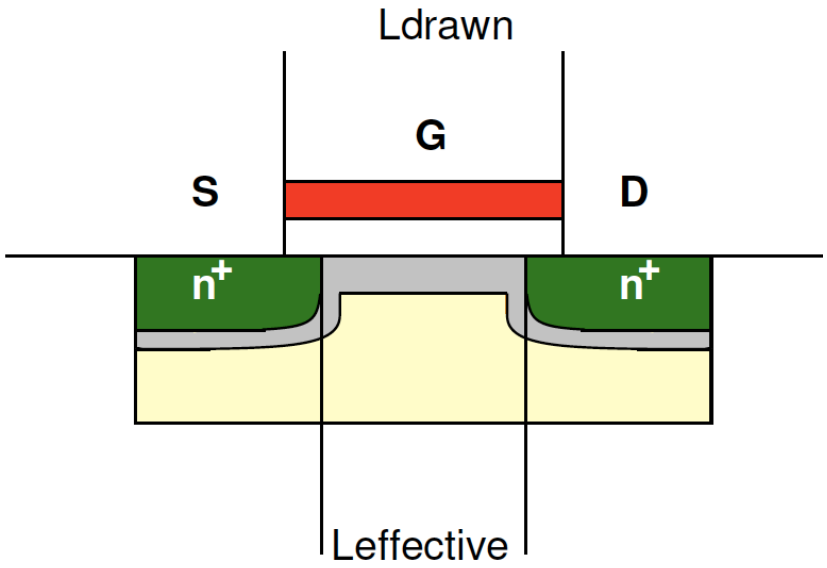
Overlap Capacitance



$$C = \varepsilon_r \varepsilon_0 \frac{A}{d}$$

$$C_o = \varepsilon_{ox} \frac{W\left(L_{drawn} - L_{effective}\right)/2}{t_{ox}}$$

Overlap Capacitance



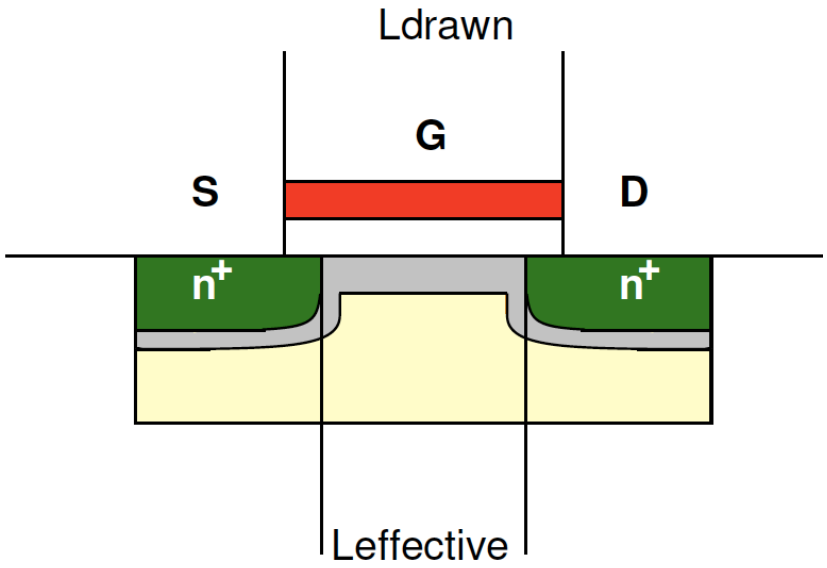
$$C = \varepsilon_r \varepsilon_0 \frac{A}{d}$$

$$C_{OX} = \frac{\varepsilon_{OX}}{t_{OX}}$$

$$C_o = \varepsilon_{ox} \frac{W\left(L_{drawn} - L_{effective}\right)/2}{t_{ox}}$$

$$C_o = \frac{1}{2} C_{ox} W \left(L_{drawn} - L_{effective}\right)$$

Overlap Capacitance



$$C = \varepsilon_r \varepsilon_0 \frac{A}{d}$$

$$C_{OX} = \frac{\varepsilon_{OX}}{t_{OX}}$$

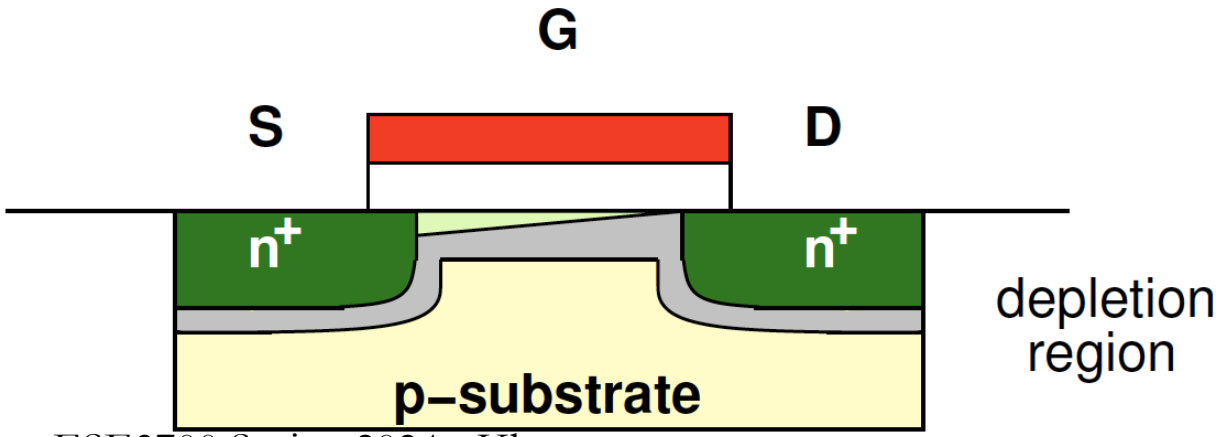
$$C_o = \varepsilon_{ox} \frac{W\left(L_{drawn} - L_{effective}\right)/2}{t_{ox}}$$

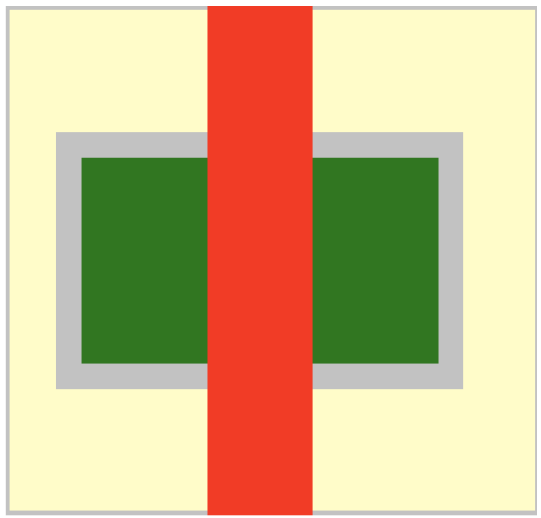
$$C_o = \frac{1}{2} C_{ox} W \left(L_{drawn} - L_{effective}\right) = C_{GSO} = C_{GDO}$$

Junction Capacitances

Junction (diffusion) Capacitance

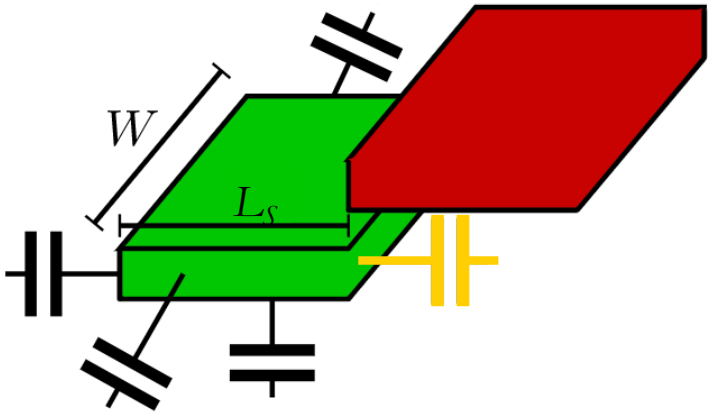
- n^+ contacts are formed by doping = diffusion
- Depletion under diffusion region (bottom-plate)
  - Due to reverse biased PN junction
  - Bottom-plate junction capacitance, C_j
- Depletion around perimeter (sidewall) of diffusion region
  - Sidewall junction capacitance, C_{jsw}

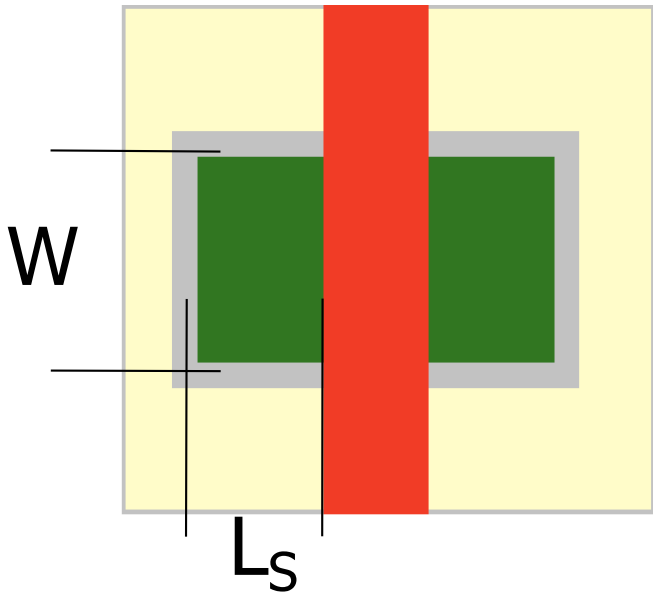




Junction (Diffusion) Capacitance

- C_j – Bottom-plate junction capacitance (F/Area)
- C_{jsw} – Sidewall junction capacitance (F/Length)
- L_S – length of diffusion region

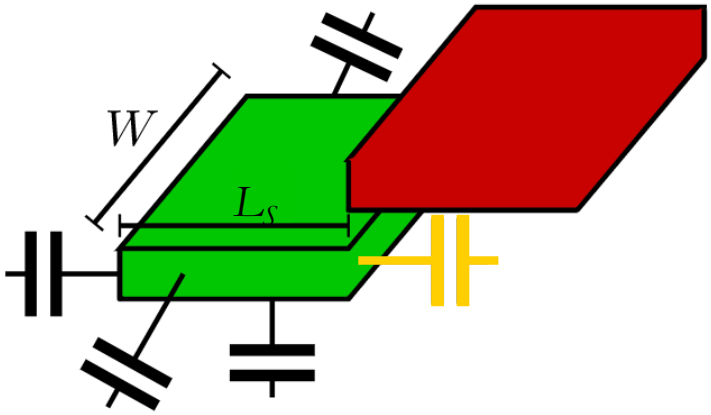


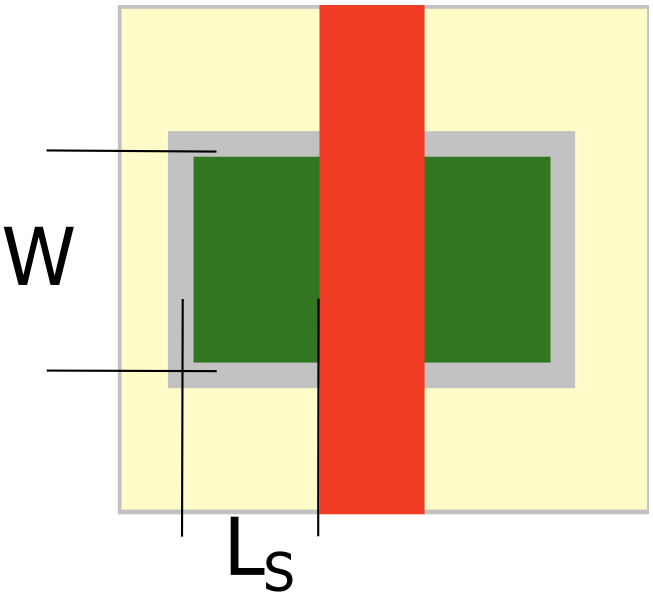


$$C_{diff} = C_j L_S W +$$

Junction (Diffusion) Capacitance

- C_j – Bottom-plate junction capacitance (F/Area)
- C_{jsw} – Sidewall junction capacitance (F/Length)
- L_S – length of diffusion region

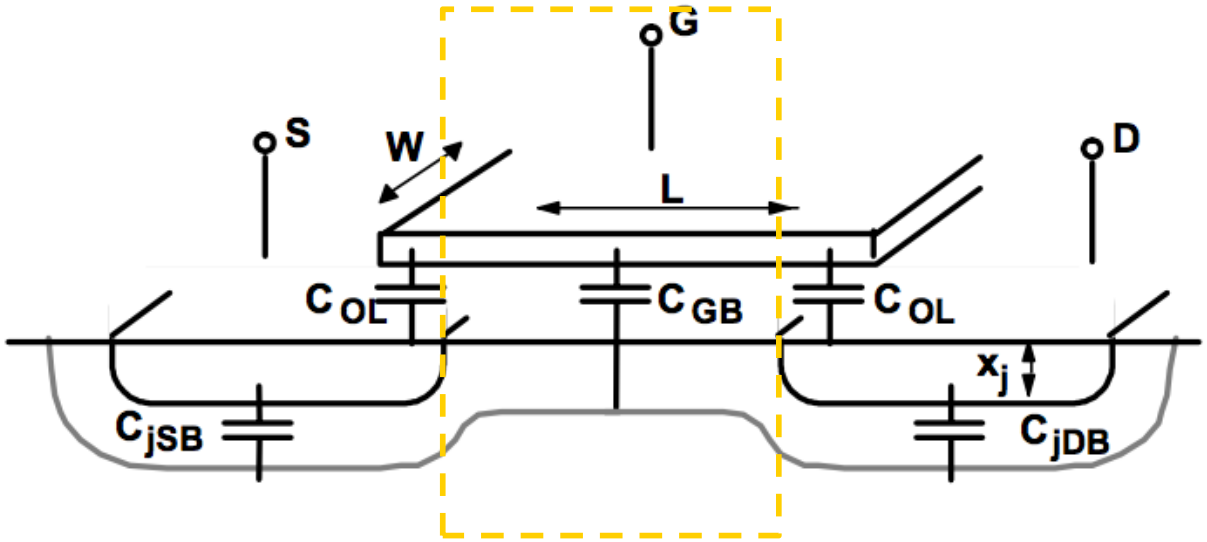




$$C_{diff} = C_j L_S W + C_{jsw}\left(2L_S + W\right)$$

Gate-to-channel

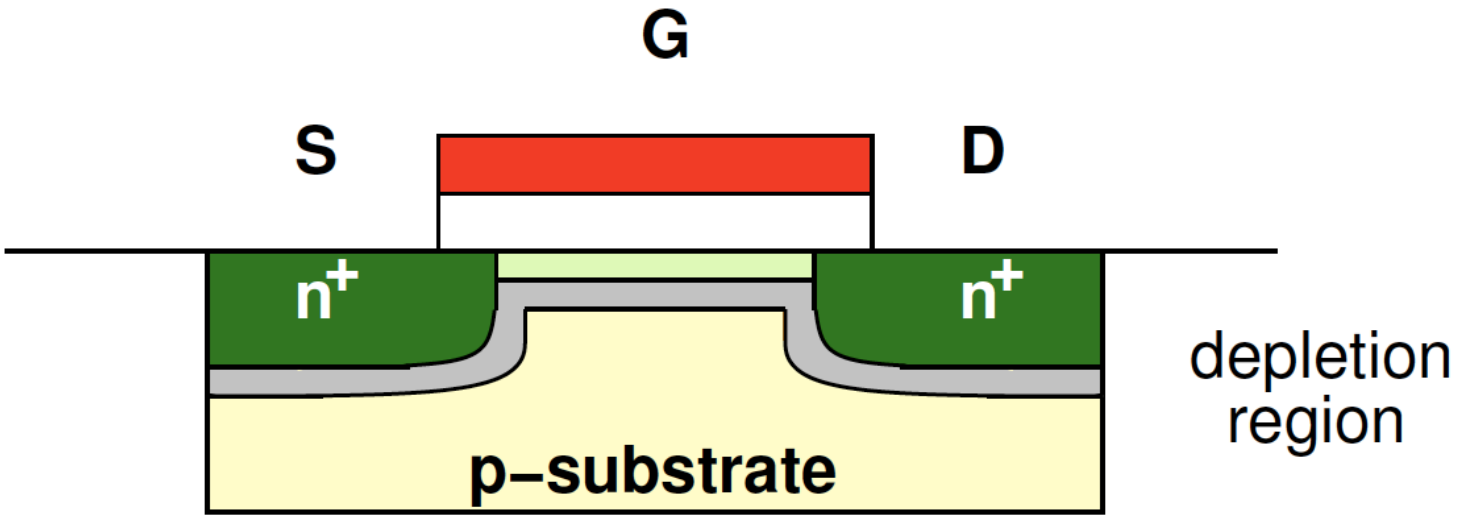
MOSFET Parasitic Capacitance



- Any two conductors separated by an insulator form a parallel-plate capacitor
- Two types
  - Extrinsic – Outside the box (e.g. junction, overlap)
  - Intrinsic – Inside the box (e.g. gate-to-channel)

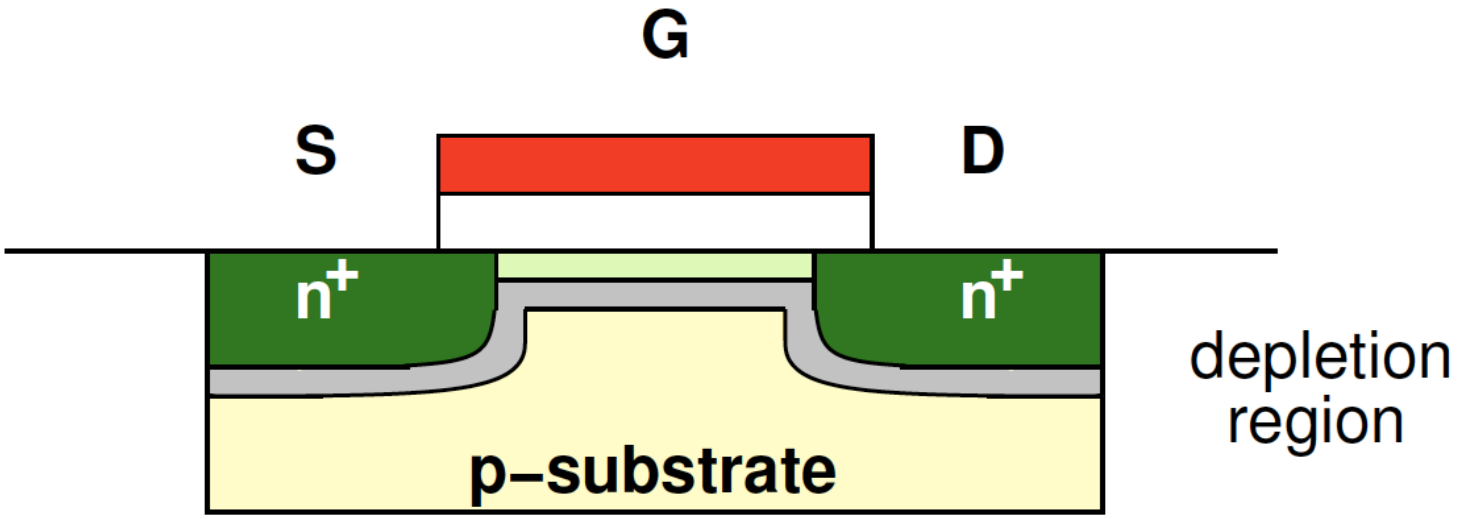
Gate-to-Bulk Capacitance

- Looks like parallel plate capacitance
- Two components:
  - What is C_{GC}? (C_{GCS}, C_{GCD})
  - What is C_{GCB}?



Gate-to-Channel Capacitance

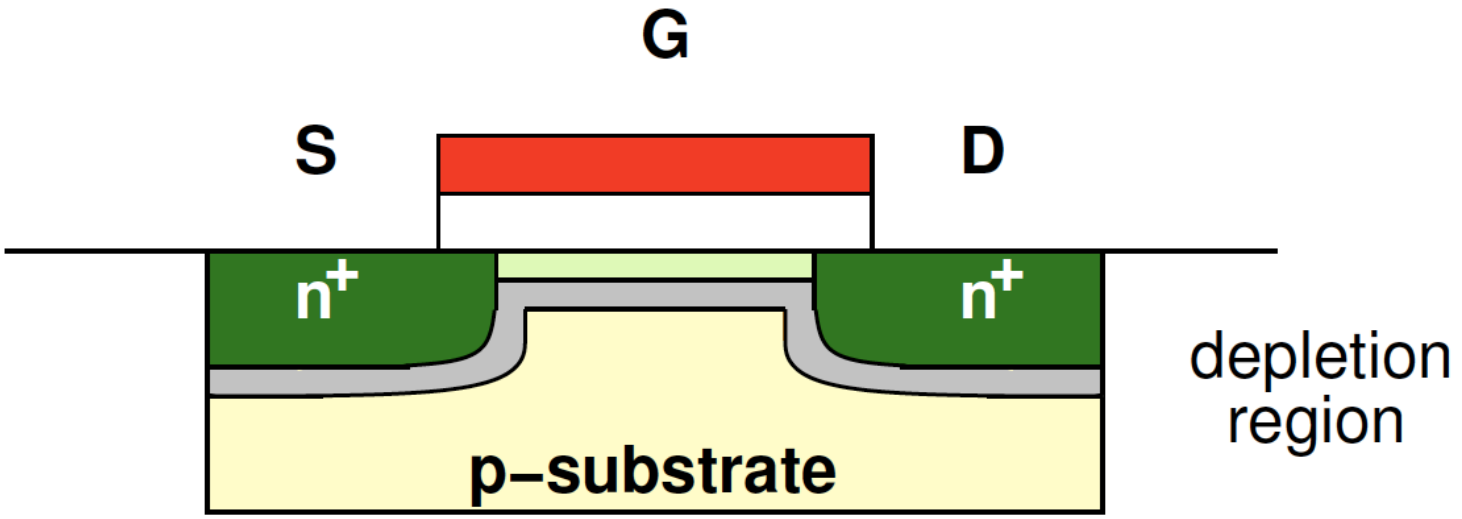
- Looks like parallel plate capacitance
- Two components: Case: Strong Inversion (small Vds)
  - C_{GC}
  - C_{GCB}



Gate-to-Channel Capacitance

- Looks like parallel plate capacitance
- Two components: Case: Strong Inversion
  - C_{GC}
  - $C_{GCB}=0$

$$C_{GC} = C_{ox} W L_{effective}$$

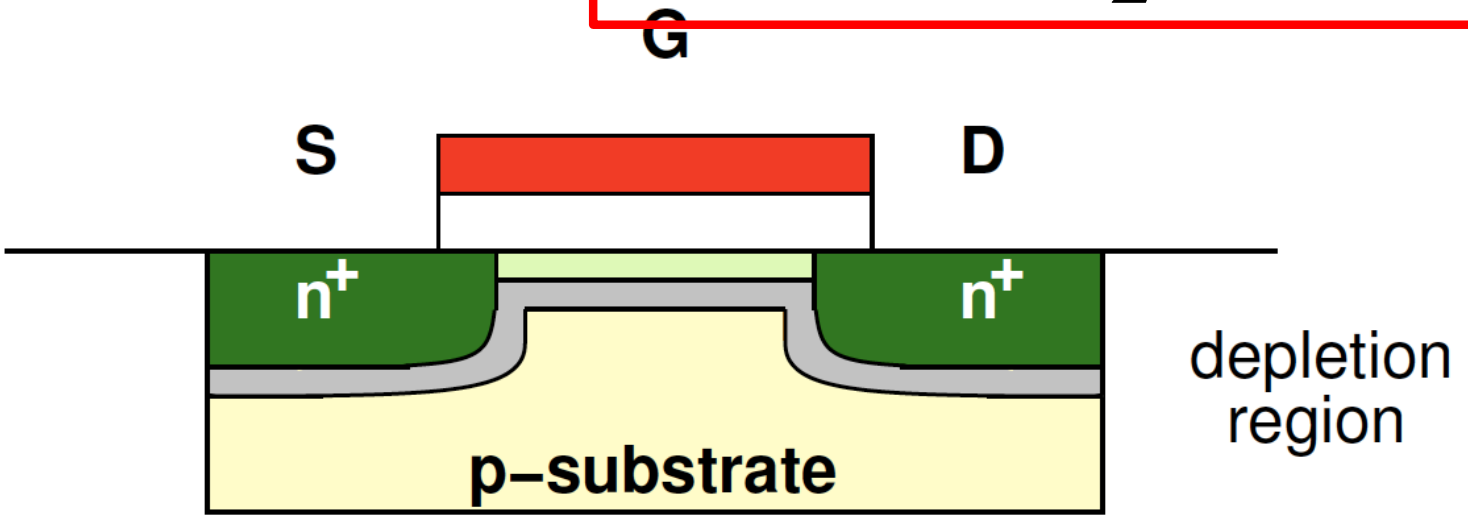


Gate-to-Channel Capacitance

- Looks like parallel plate capacitance
- Two components: Case: Strong Inversion
  - C_{GC} – Split evenly between S and D
  - C_{GB}=0

$$C_{GC} = C_{ox} WL_{effective}$$

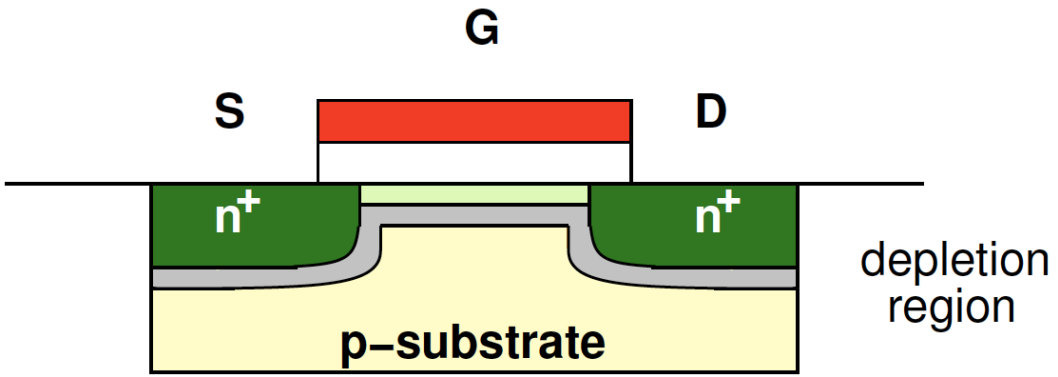
$$C_{GCS} = C_{GCD} = \frac{1}{2} C_{ox} WL_{effective}$$



Gate-to-Source Capacitance

- Channel + Overlap

$$C_{GS} = C_{GCS} + C_{GSO}$$



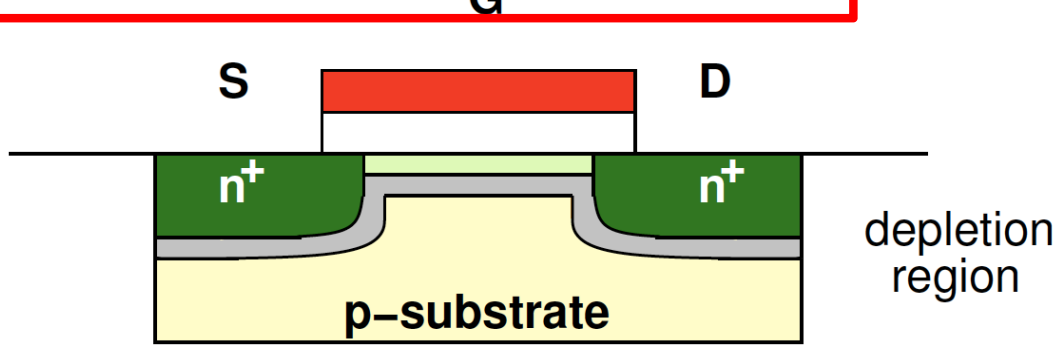
Gate-to-Source Capacitance

- Channel + Overlap

$$C_{GS} = C_{GCS} + C_{GSO}$$

$$C_{GS} = \frac{1}{2} C_{OX} W (L_{drawn} - L_{effective}) + \frac{1}{2} C_{OX} W L_{effective}$$

$$C_{GS} = \frac{1}{2} C_{OX} W L_{drawn}$$



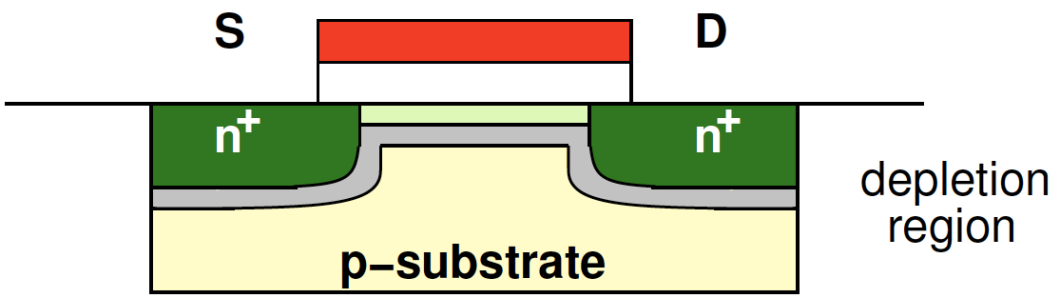
Gate-to-Drain Capacitance

- Channel + Overlap

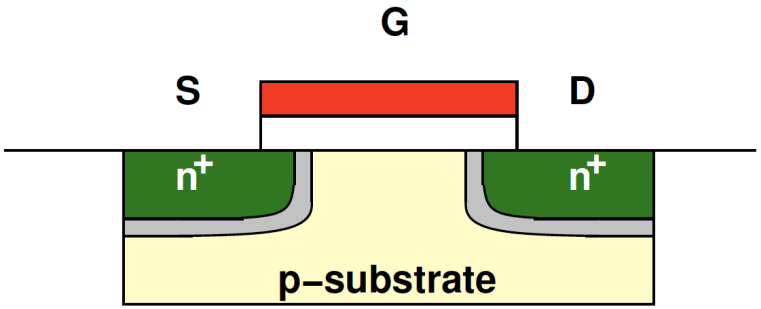
$$C_{GD} = C_{GCD} + C_{GDO}$$

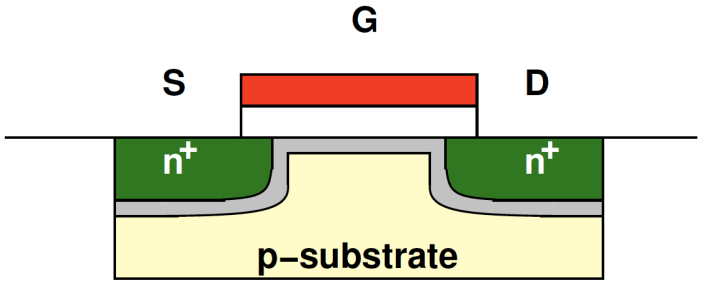
$$C_{GD} = \frac{1}{2} C_{OX} W (L_{drawn} - L_{effective}) + \frac{1}{2} C_{OX} W L_{effective}$$

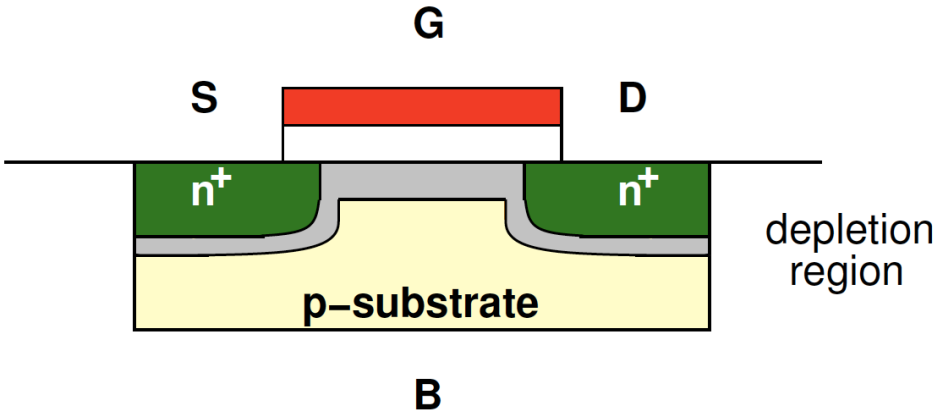
$$C_{GD} = \frac{1}{2} C_{OX} W L_{drawn}$$

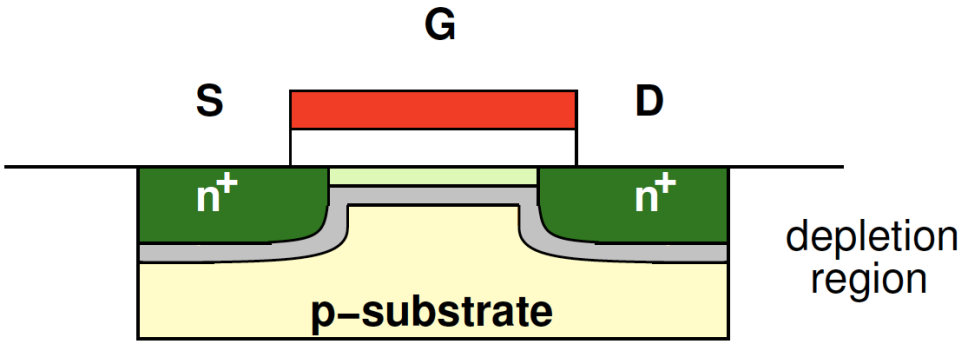


Channel Evolution: Weak Inversion



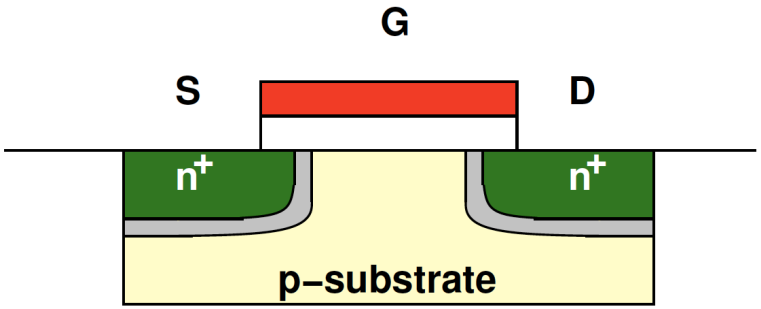


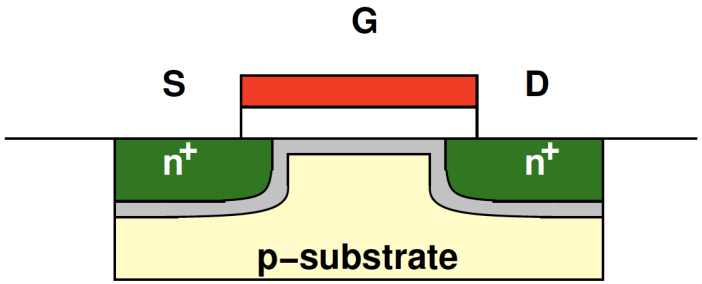


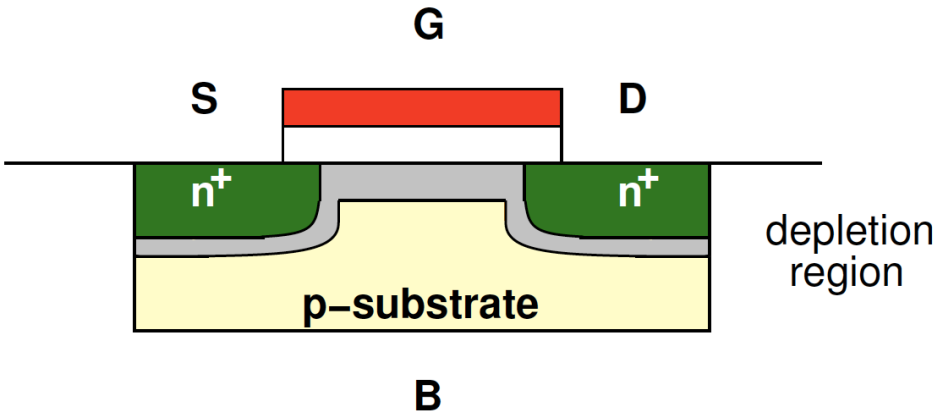


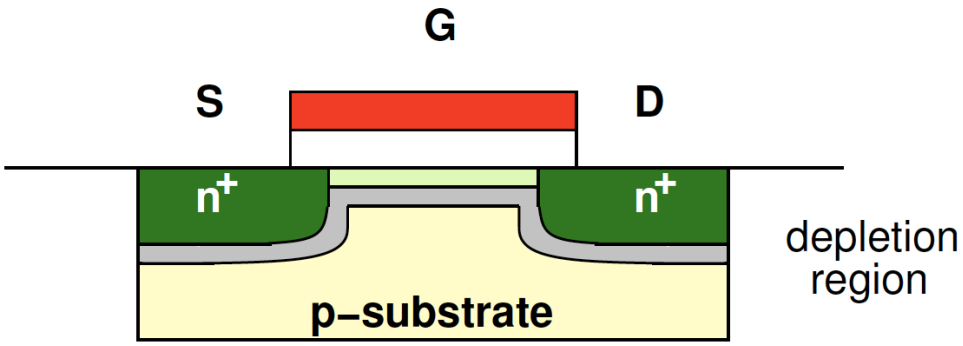
Channel Evolution: Weak Inversion

- $V_{GS} = 0 \rightarrow C_{GC}=0, C_{GCB}=WLC_{ox}$



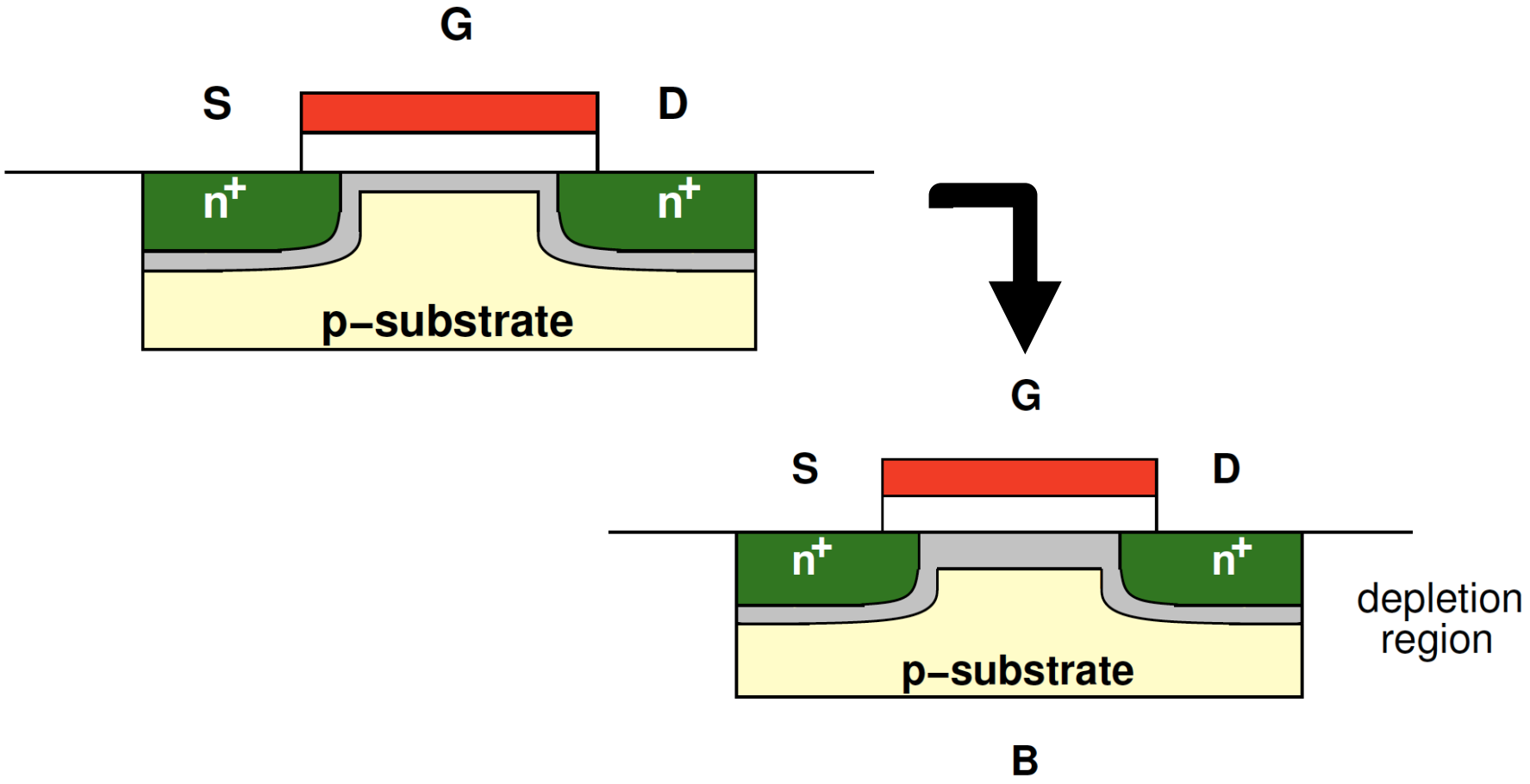






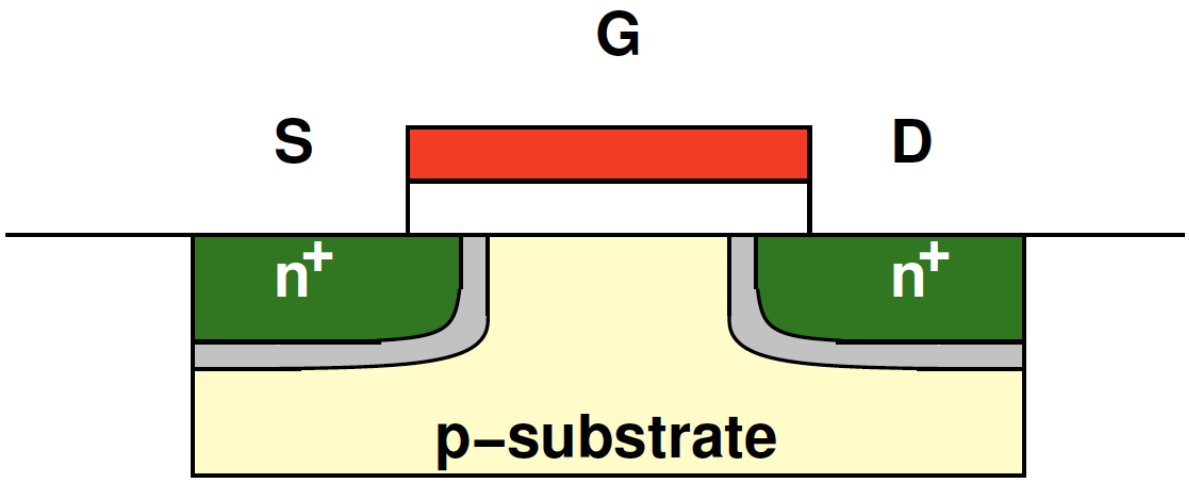
Channel Evolution: Weak Inversion

- What happens to capacitance here as V_{GS} increases?
  - Capacitor plate distance?

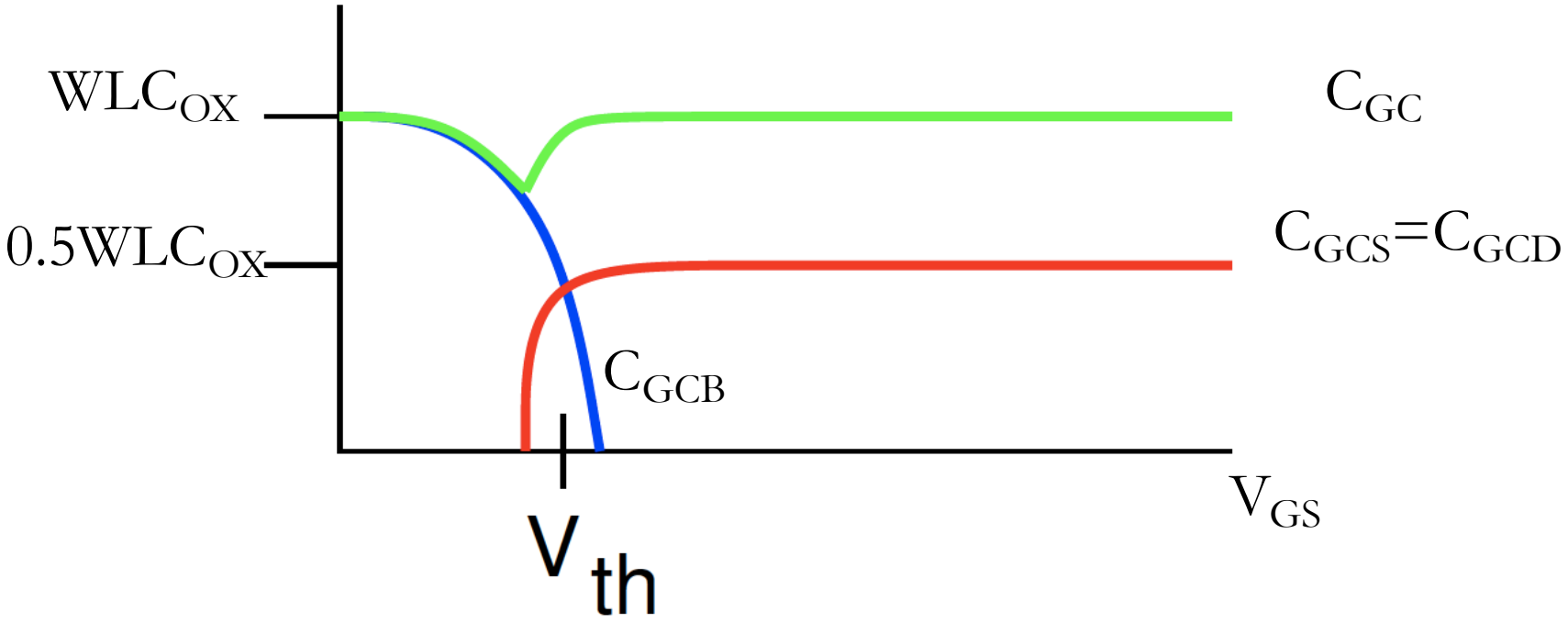


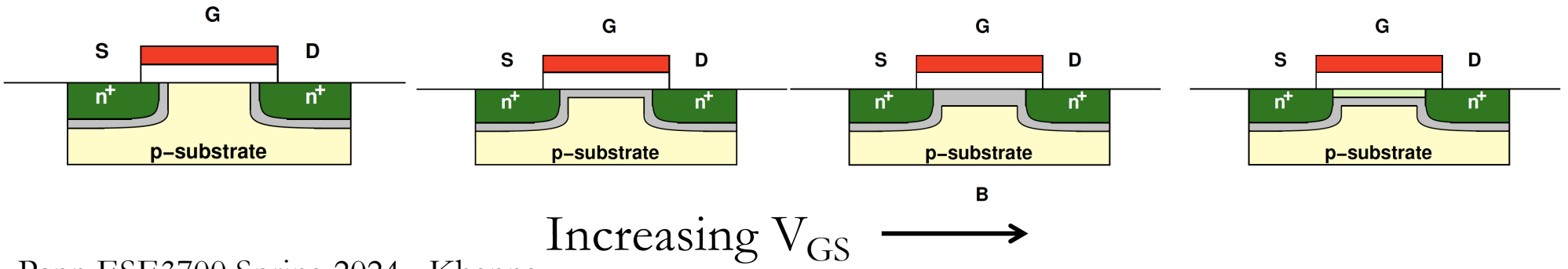
Channel Evolution: Weak Inversion

- Capacitance is initially dominated by Gate-to-bulk capacitance ($C_{GCS,D}=0$)
- Gate-to-bulk capacitance drops as V_{GS} increases toward V_{th}

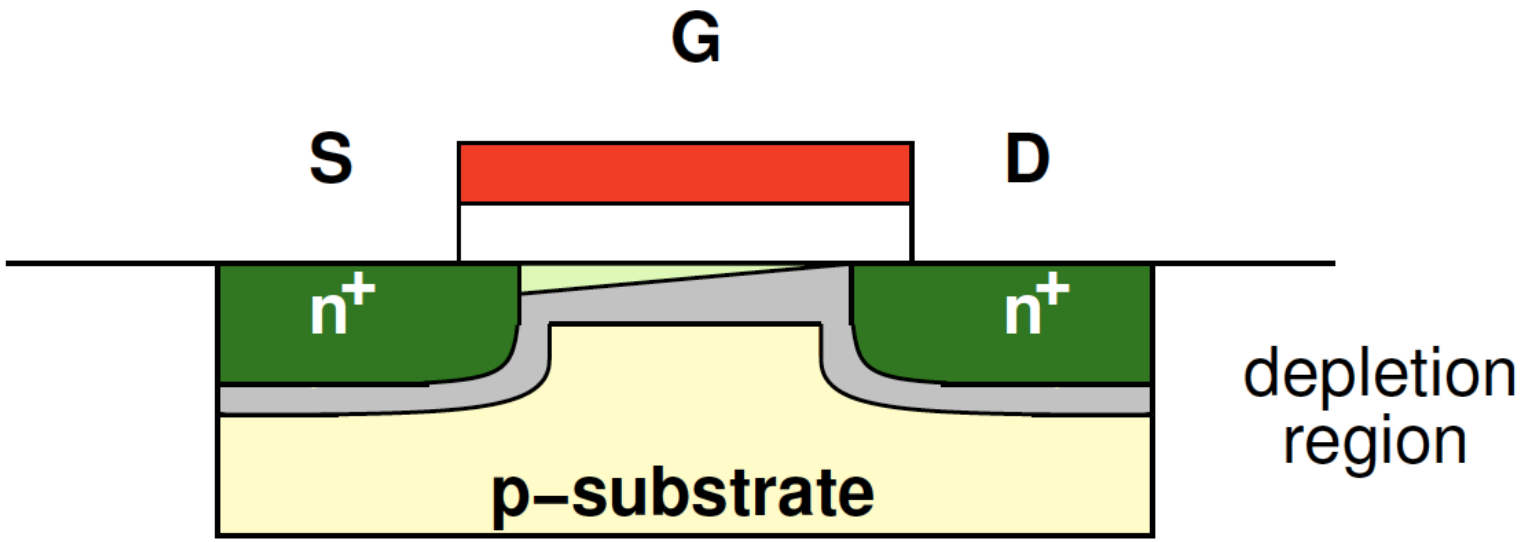


Capacitance vs V_{GS} (V_{DS}=0)



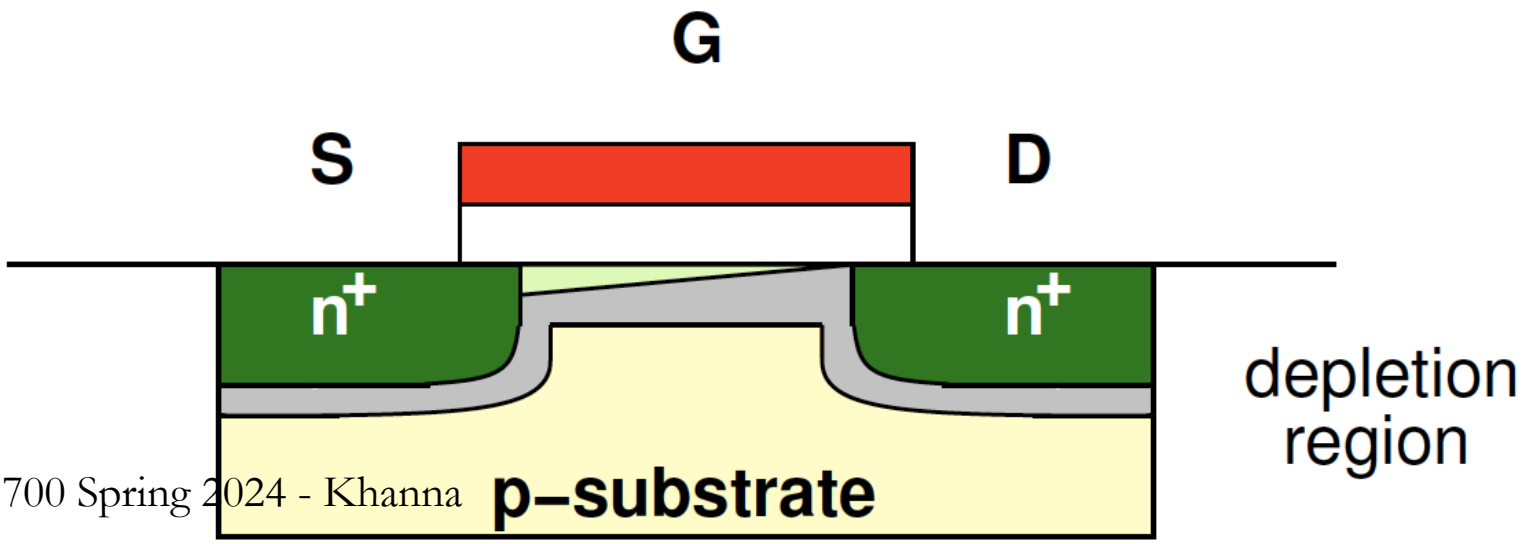


Saturation Capacitance?

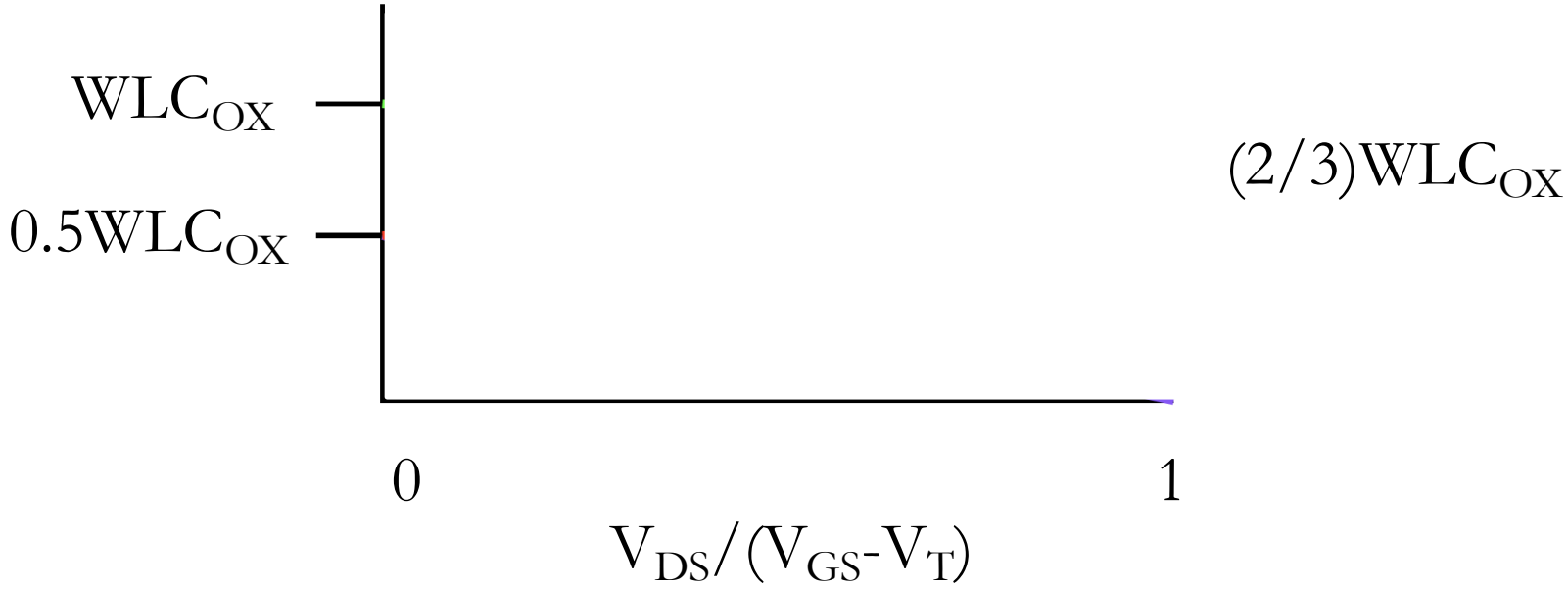


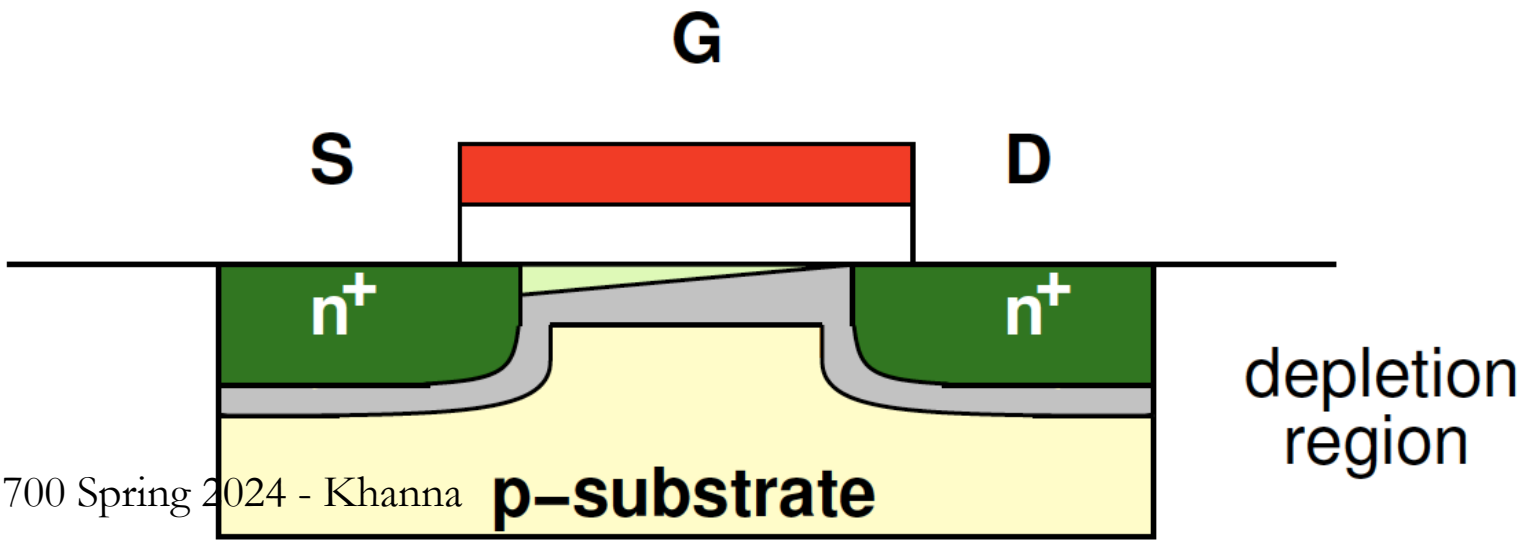
Saturation Capacitance?

- Source end of channel in inversion
- Voltage at drain end of channel at or below threshold
- Capacitance shifts to source
  - Total capacitance reduced

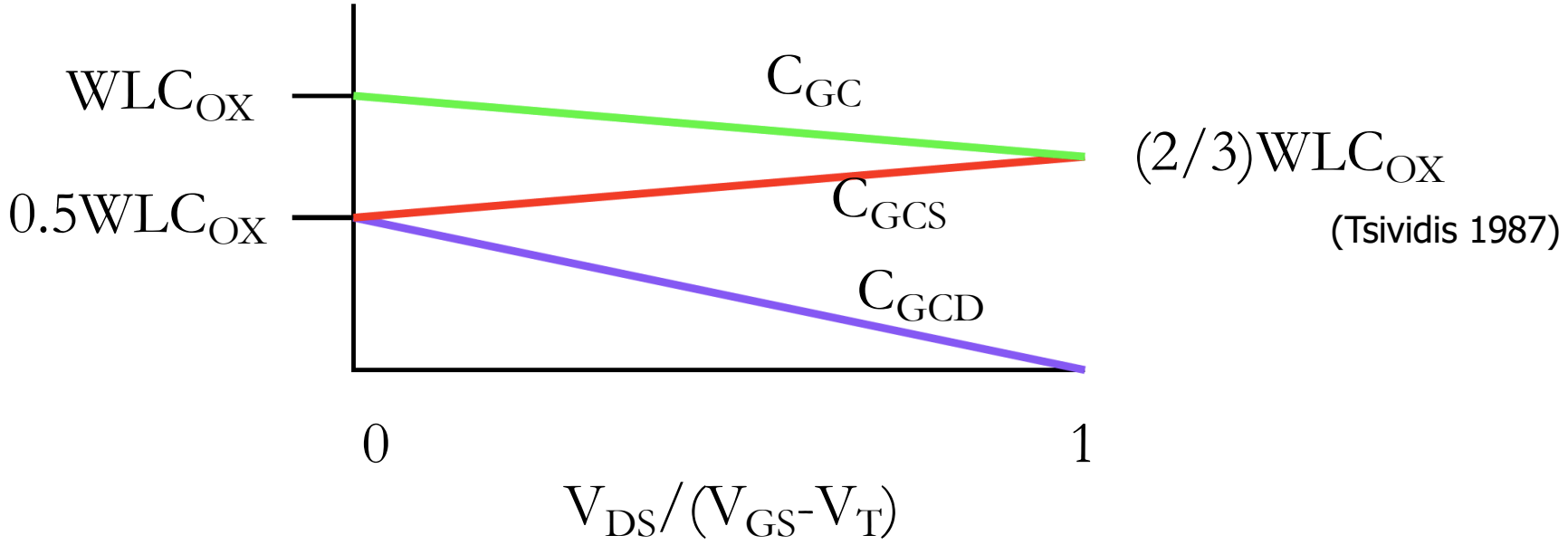


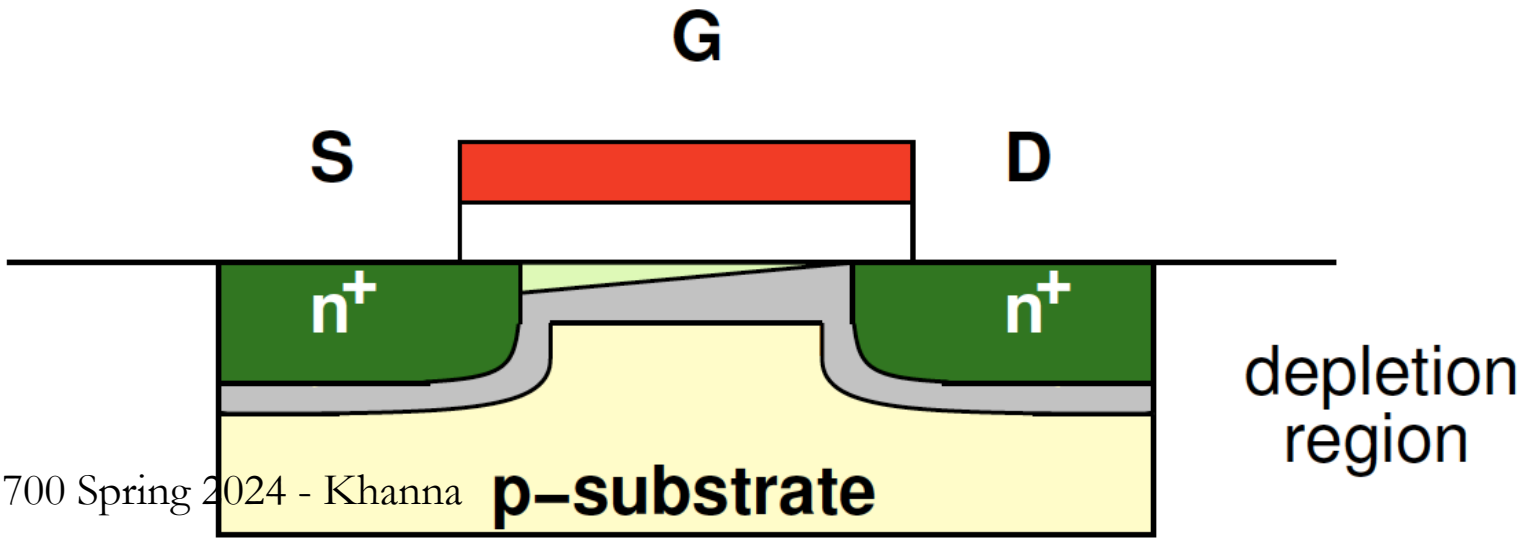
Saturation Capacitance





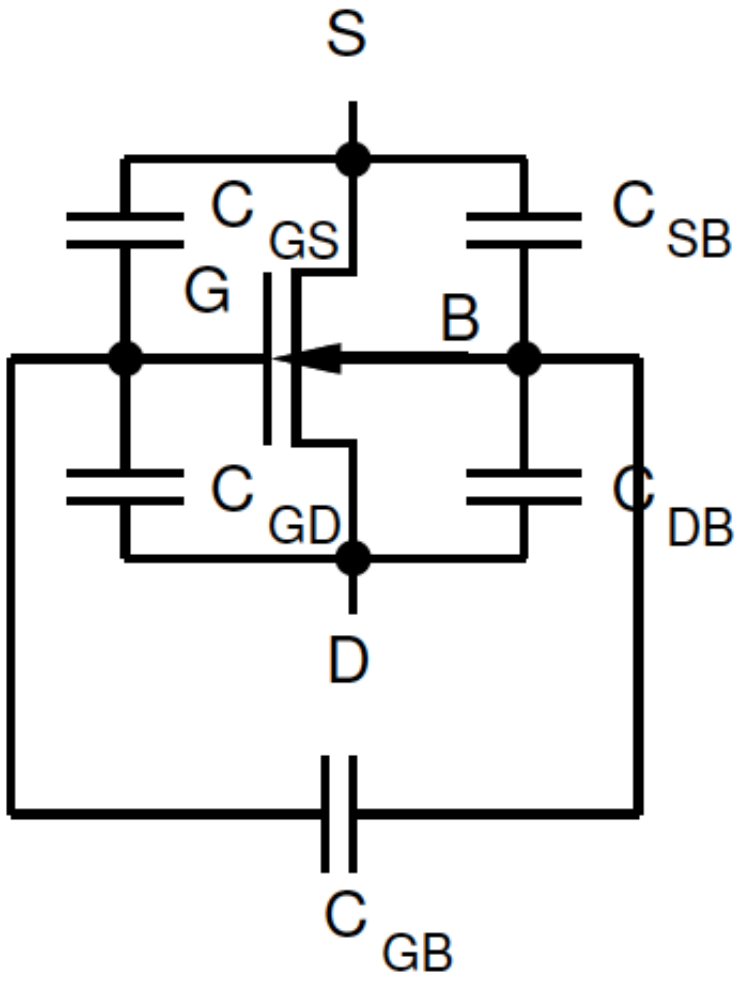
Saturation Capacitance





Capacitance Roundup

- $C_{GS}=C_{GCS}+C_{GSO}$
- $C_{GD}=C_{GCD}+C_{GDO}$
- $C_{GB}=C_{GCB}$
- $C_{SB}=C_{diff}$
- $C_{DB}=C_{diff}$



First Order Capacitance Summary

Operation Region	C_{GCB}	C_{GCS}	C_{GCD}	C_{GC}	C_G
Subthreshold					
Linear					
Saturation					

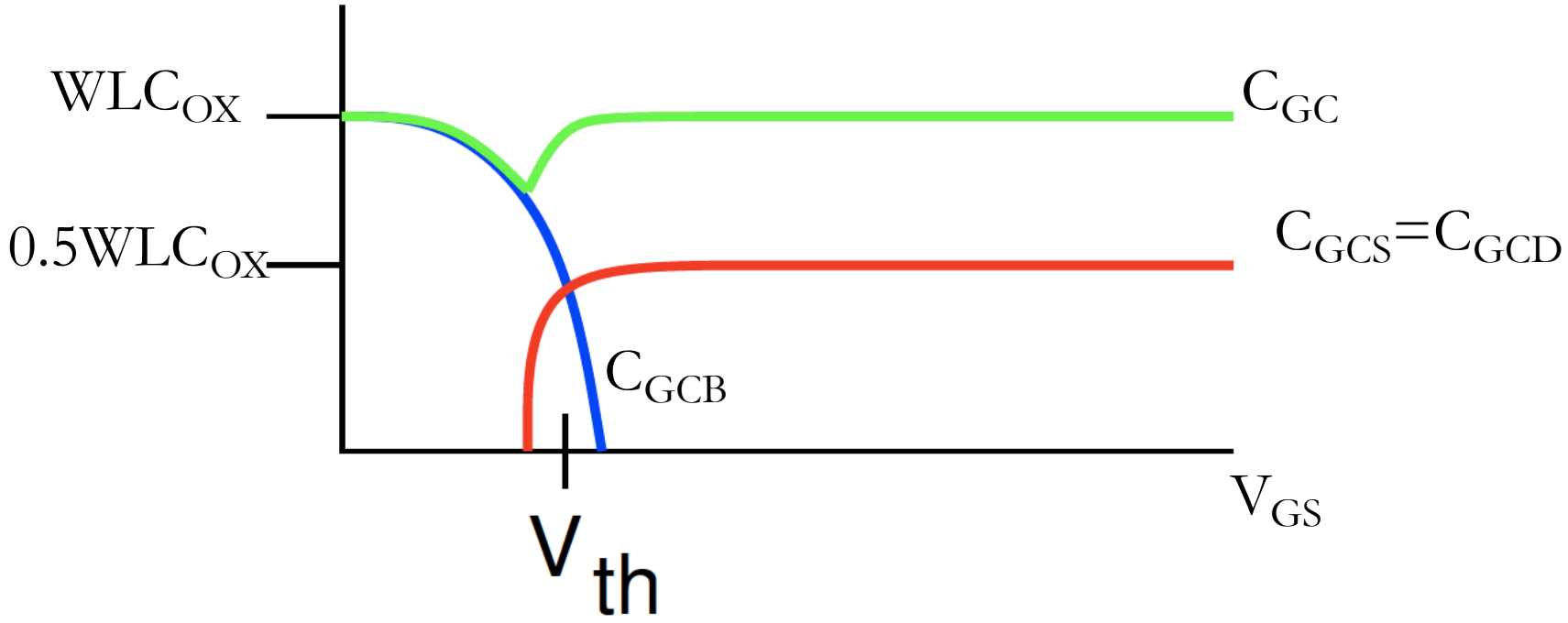
First Order Capacitance Summary

Operation Region	C_{GCB}	C_{GCS}	C_{GCD}	C_{GC}	C_G
Subthreshold					
Linear	0	$C_{OX}WL/2$	$C_{OX}WL/2$		
Saturation					

$$C_{GCS} = C_{GCD} = \frac{1}{2} C_{ox} W L_{effective}$$

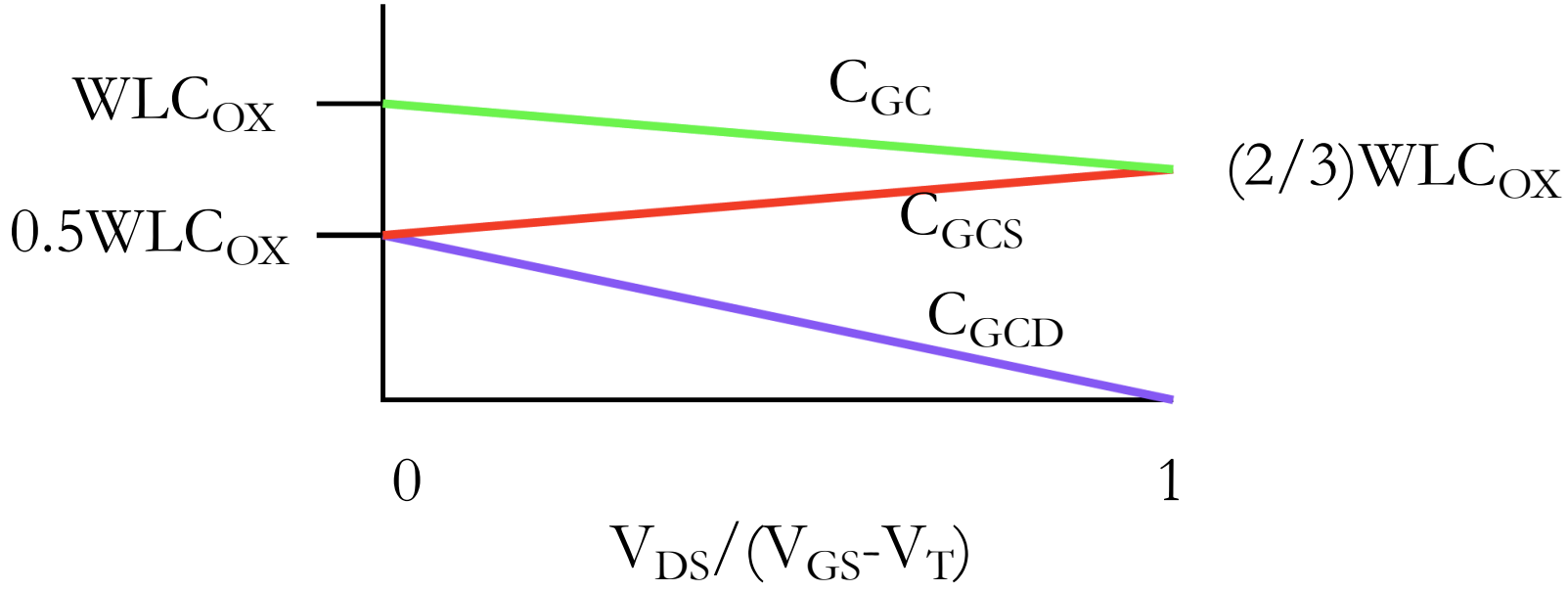
First Order Capacitance Summary

Operation Region	C_{GCB}	C_{GCS}	C_{GCD}	C_{GC}	C_G
Subthreshold	$C_{OX}WL$	0	0		
Linear	0	$C_{OX}WL/2$	$C_{OX}WL/2$		
Saturation					



First Order Capacitance Summary

Operation Region	C_{GCB}	C_{GCS}	C_{GCD}	C_{GC}	C_G
Subthreshold	$C_{OX}WL$	0	0		
Linear	0	$C_{OX}WL/2$	$C_{OX}WL/2$		
Saturation	0	$(2/3)C_{OX}WL$	0		



First Order Capacitance Summary

Operation Region	C_{GCB}	+	C_{GCS}	+	C_{GCD}	=	C_{GC}	C_G
Subthreshold	$C_{OX}WL$		0		0		$C_{OX}WL$	
Linear	0		$C_{OX}WL/2$		$C_{OX}WL/2$		$C_{OX}WL$	
Saturation	0		$(2/3)C_{OX}WL$		0		$(2/3)C_{OX}WL$	

First Order Capacitance Summary

Operation Region	C_{GCB}	C_{GCS}	C_{GCD}	C_{GC}	C_G
Subthreshold	$C_{OX}WL$	0	0	$C_{OX}WL$	$C_{OX}WL+2C_O$
Linear	0	$C_{OX}WL/2$	$C_{OX}WL/2$	$C_{OX}WL$	$C_{OX}WL+2C_O$
Saturation	0	$(2/3)C_{OX}WL$	0	$(2/3)C_{OX}WL$	$(2/3)C_{OX}WL+2C_O$

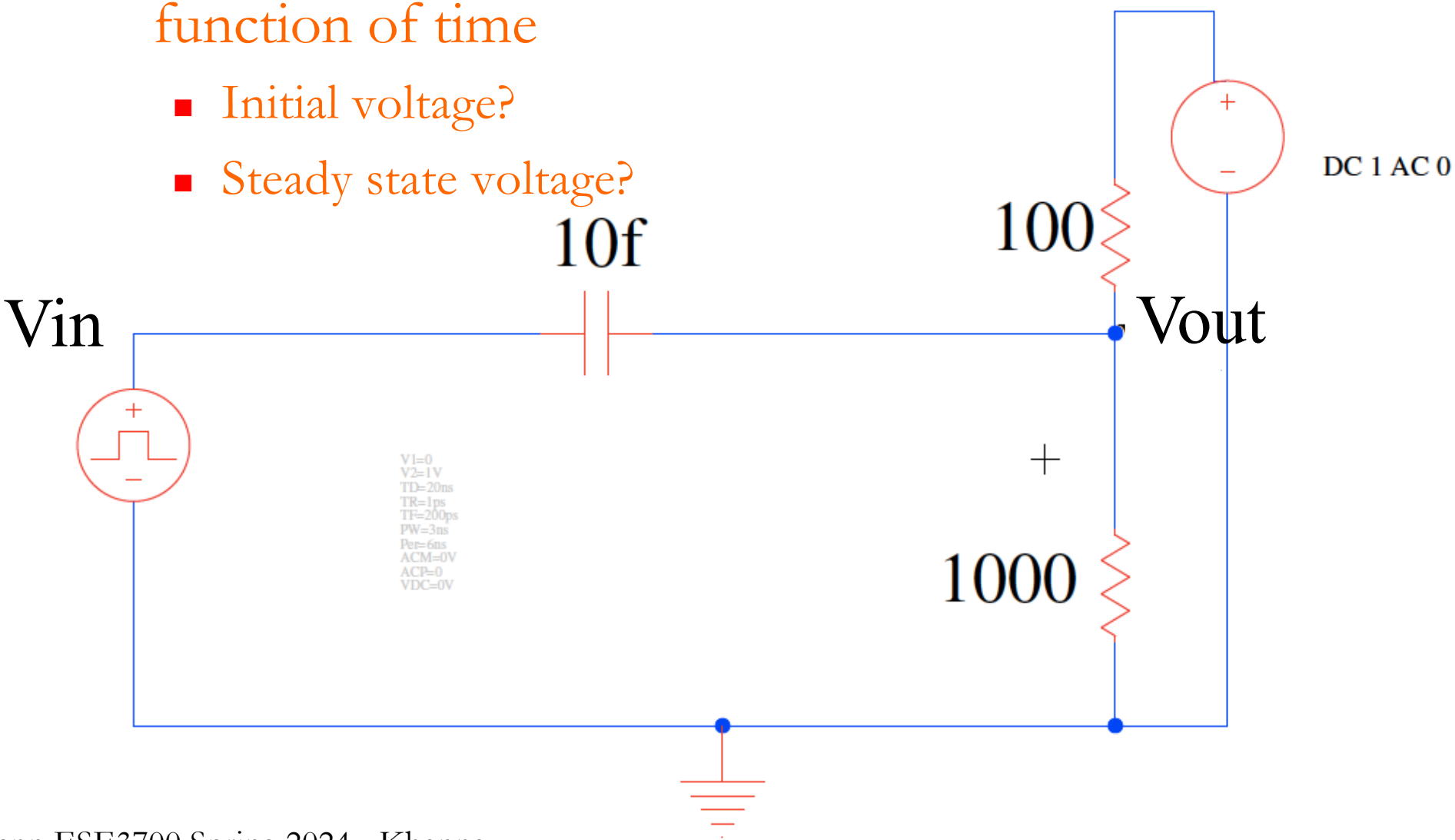
$$C_o = \frac{1}{2} C_{ox} W \left(L_{drawn} - L_{effective} \right) = C_{GSO} = C_{GDO}$$

One Implication

Feedback Capacitance C_{gd}

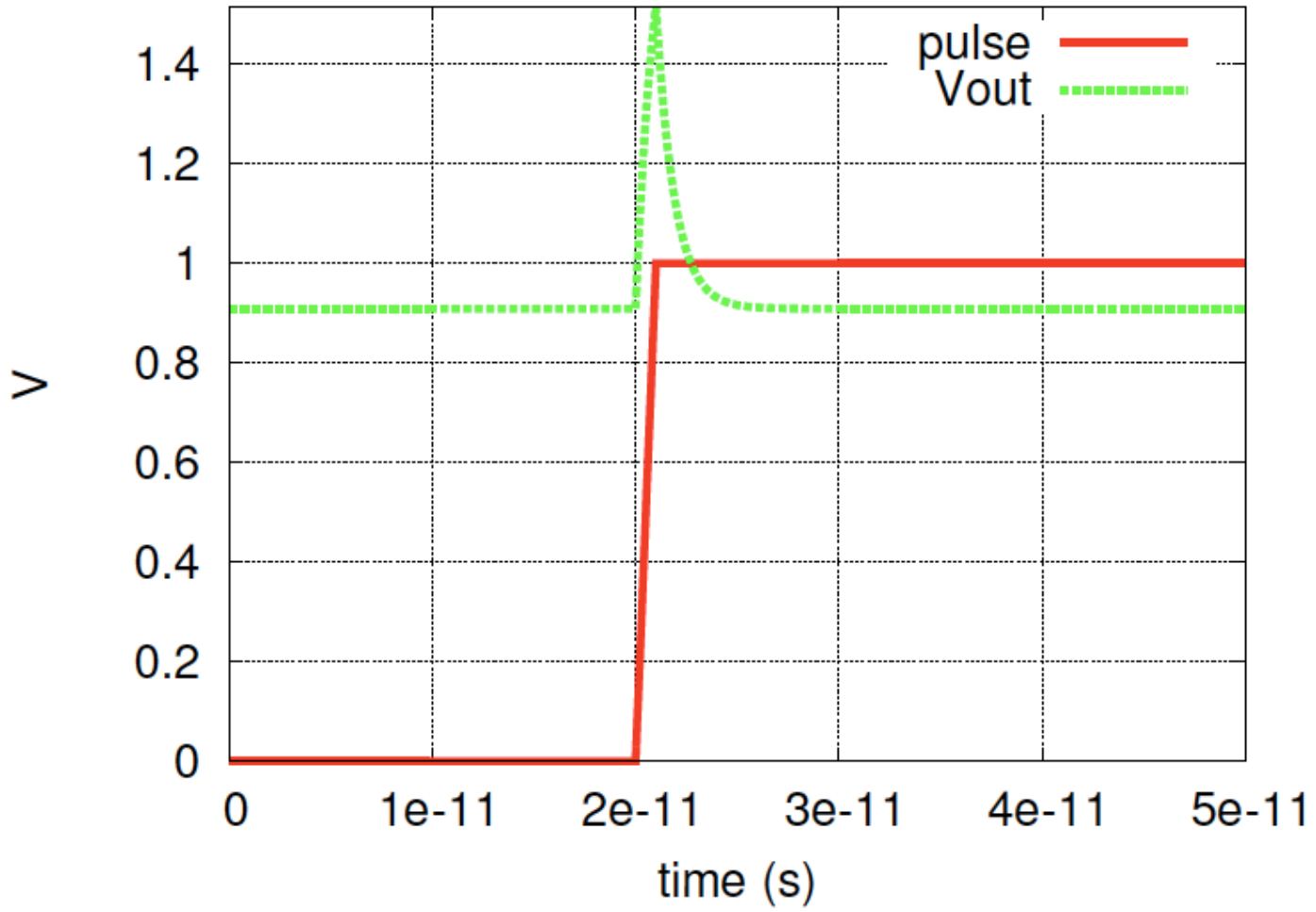
Step Response? (Preclass 3)

- Vin steps from 0 to 1, what does Vout look like as a function of time
  - Initial voltage?
  - Steady state voltage?



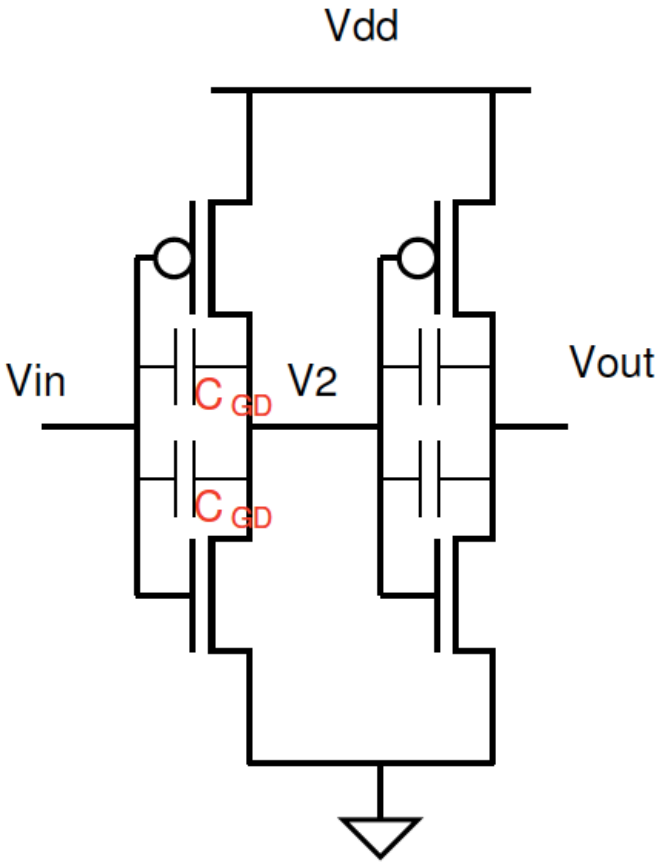
Step Response

Voltage peaking!

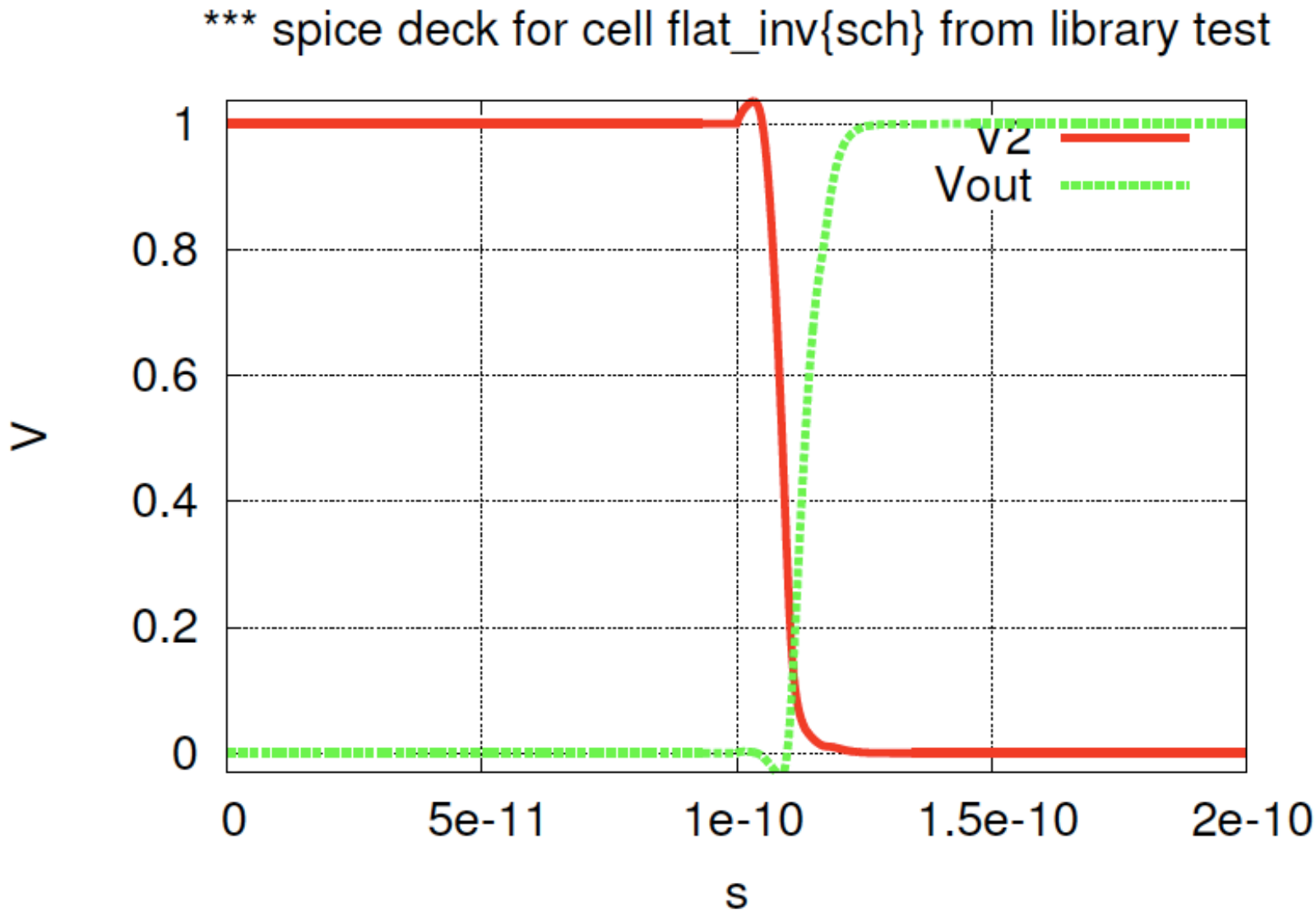


Impact of C_{GD}

- What does C_{GD} do to the switching response here?
  - V_2
  - V_{out}

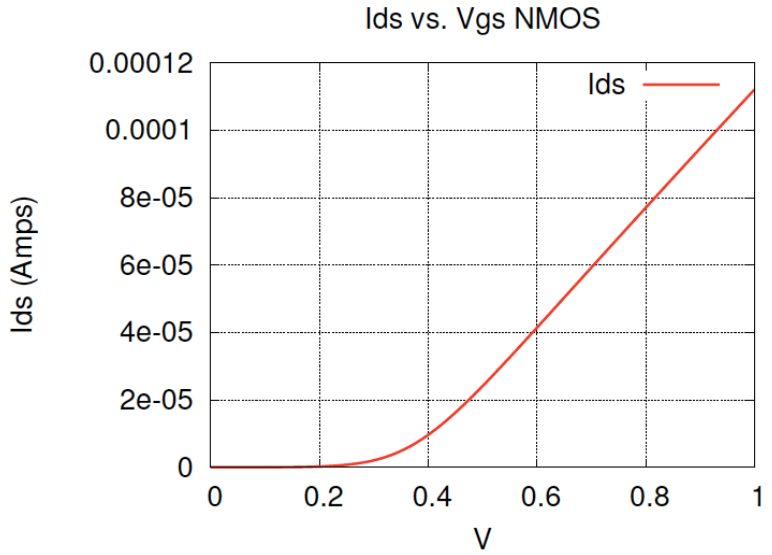


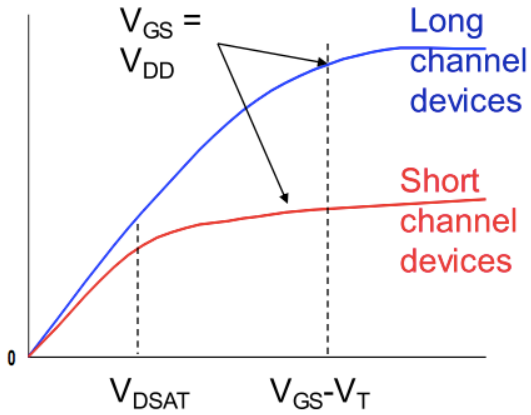
Impact of C_{GD}

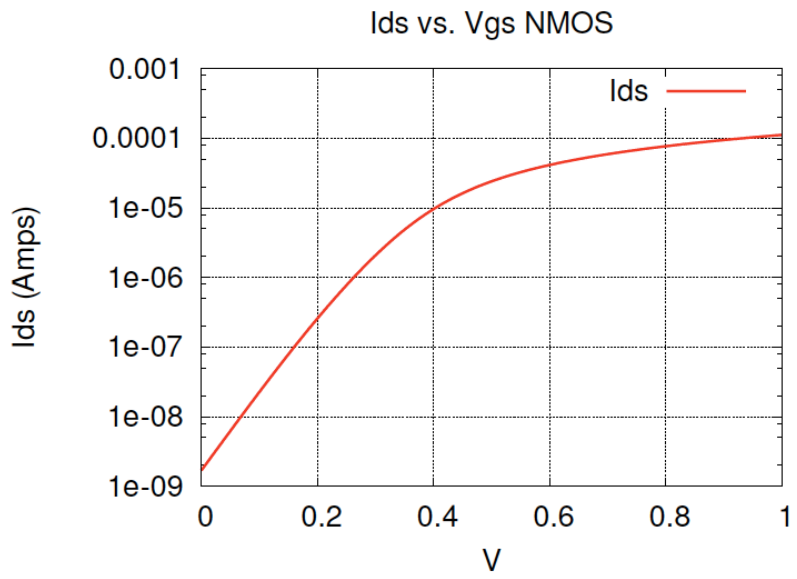


Big Idea

- 3+ Regions of operation for MOSFET
  - Subthreshold
  - Linear
  - Saturation
    - Pinch Off
  - Velocity Saturation, DIBL
    - Short channel

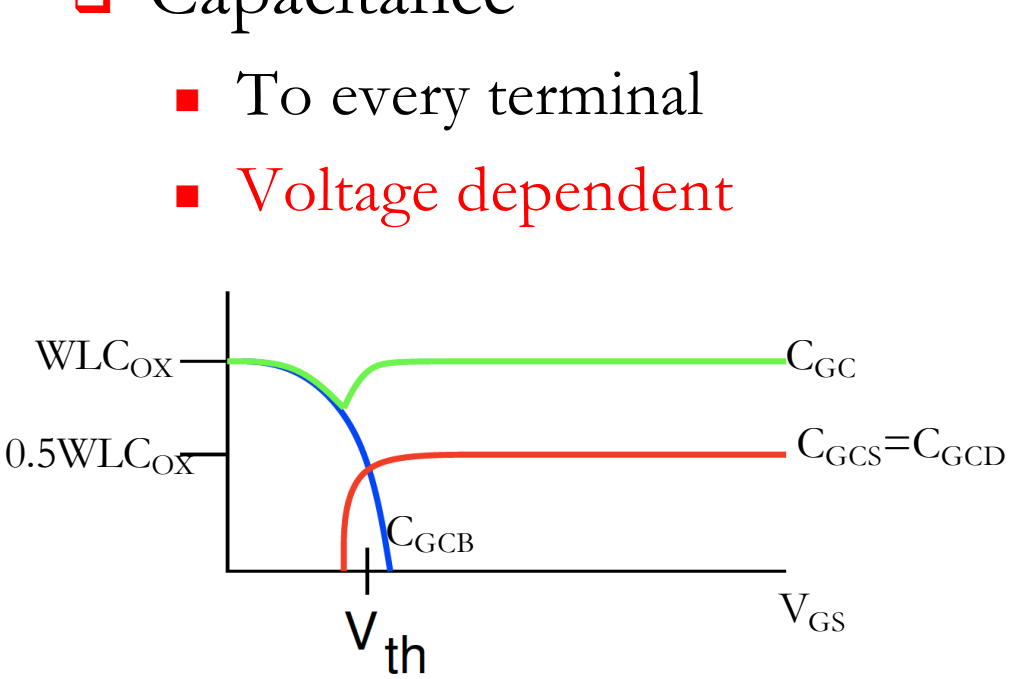


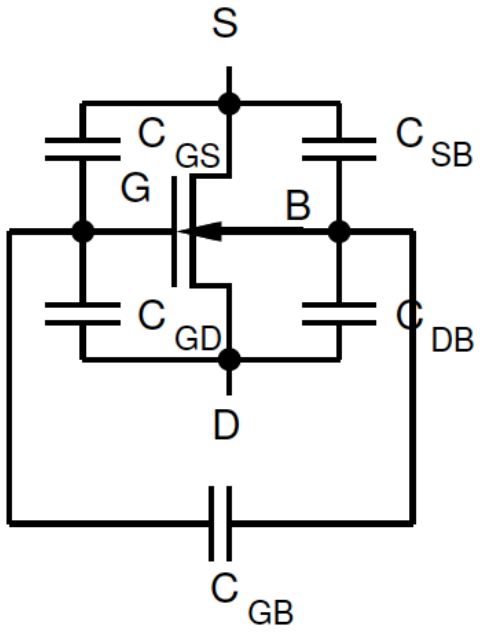


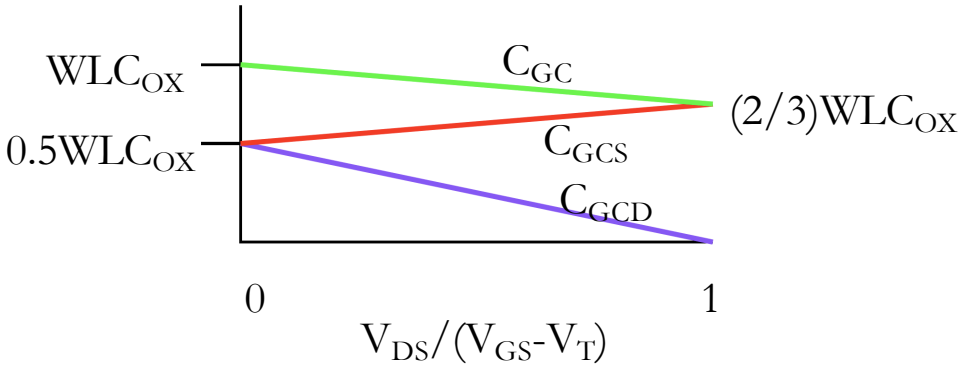


Big Idea

- Capacitance
  - To every terminal
  - Voltage dependent







Admin

- HW3 out now – due 2/16 (Friday)
  - Takes time! Learning curve for how to debug
  - Don't forget the demo/video of SPICE workflow
    - Get TA checkoff or submit video in Canvas
- Midterm 1 Wednesday 2/21 (next week)
  - 1.5 hrs during class in Moore 216
  - Midterm 1 TA review session (TBA)
  - Covers lecture 1-6 (today)
  - Old exams posted on previous years websites
    - Note old exams were for 2hrs

Midterm 1 - Coverage

- Lec 1 - 6
  - Identify CMOS/non-CMOS
  - Identify CMOS function
  - Any logic function → CMOS gate
  - Noise Margins / Restoration
  - Circuit first order switching rise/fall times
    - Output equivalent resistance
    - Load capacitance
  - MOS Model
    - Identify transistor region of operation
    - Analysis with transistor IV models
    - MOS capacitance models

Acknowledgement

- Prof. André DeHon (University of Pennsylvania)
- Prof. Jing Li (University of Pennsylvania)