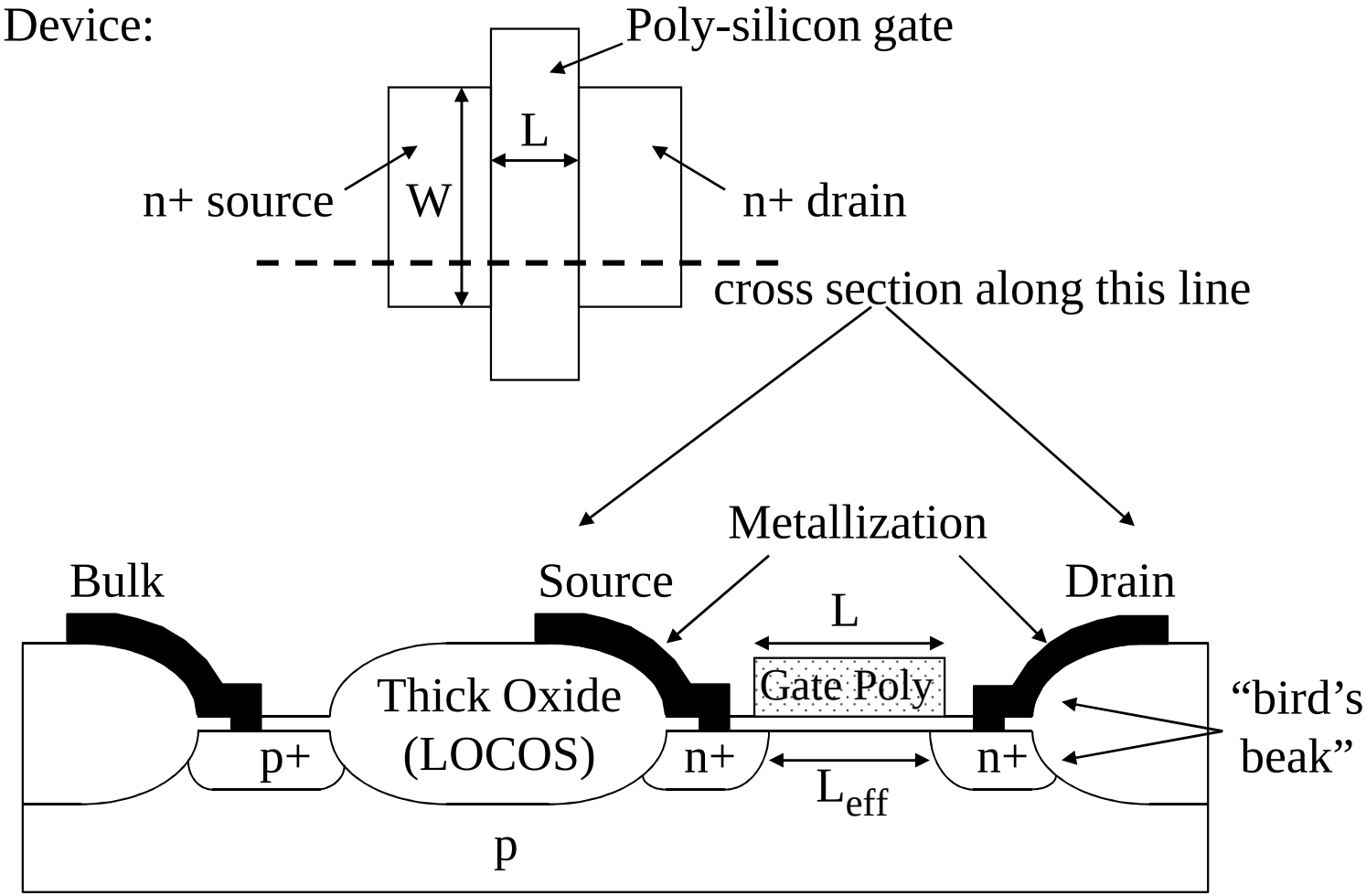
EE4011: RF IC Design
MOSFETs at High Frequency

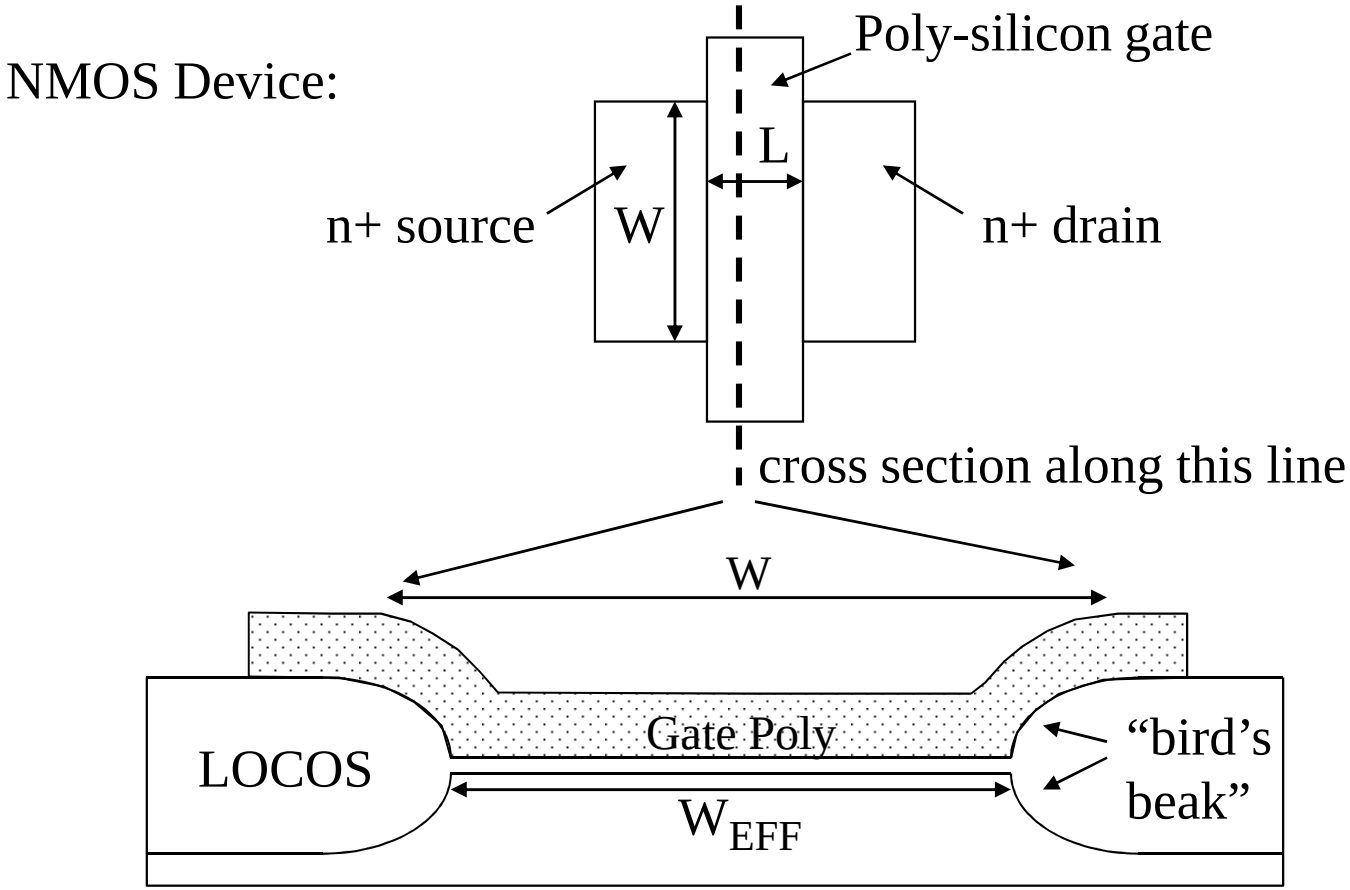
MOSFET Structure - 1

NMOS Device:



MOSFET Structure - 2

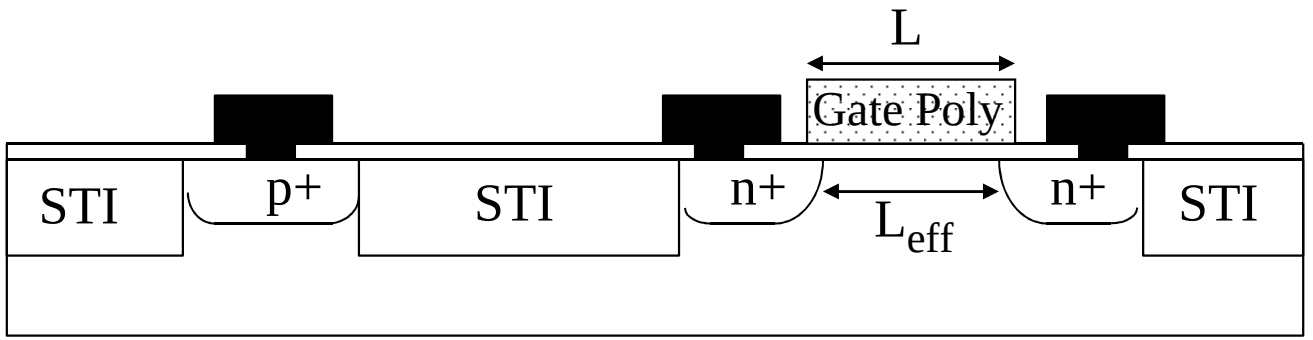
NMOS Device:

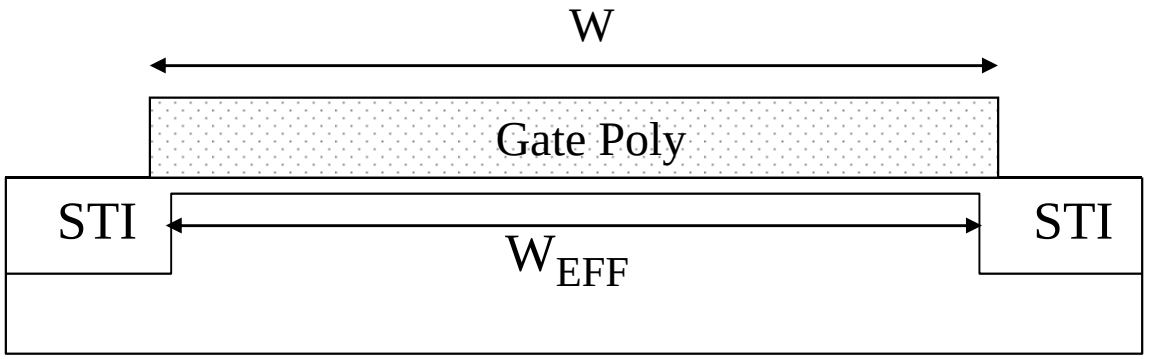


In older processes (L>0.25μm), the thick oxide device isolation regions were mainly formed by a process called local oxidation of silicon (LOCOS = LOCallly Oxidised Silicon). At the boundary between thin and thick oxides this gave a characteristic shape called the bird's beak.

MOSFET Structure - 3

More advanced processes use a shallow trench of oxide (Shallow Trench Isolation) to achieve isolation between devices. This process gives a planar surface topology.





N-type MOSFET I-V Equations

W: MOSFET Width
L: MOSFET Length
T_{OX}: Gate Oxide Thickness
ε_{OX}: Gate Oxide Permittivity
C'_{OX}: Gate Oxide Capacitance per unit Area
μ: electron mobility
λ: output conductance parameter

$$C'_{OX} = \frac{\varepsilon_{OX}}{T_{OX}}$$

Off : $V_{GS} \le V_{TH}$

$I_{DS} = 0$ (simplistic)

Linear : $V_{GS} > V_{TH}$, $V_{DS} < V_{GS} - V_{TH}$

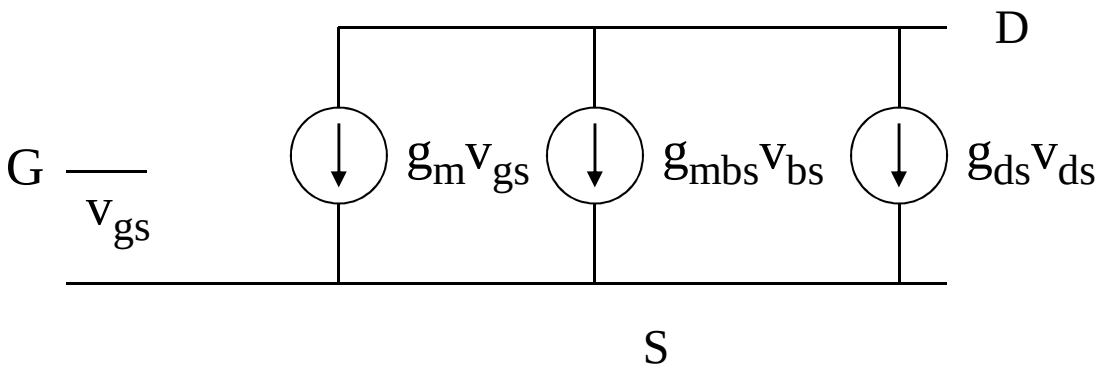
$$I_{DS} = \frac{W}{L}\mu C'_{OX}\left(V_{GS} - V_{TH} - \frac{V_{DS}}{2}\right)V_{DS}$$

Saturation : $V_{GS} > V_{TH}$, $V_{DS} \ge V_{GS} - V_{TH}$

$$I_{DS} = \frac{1}{2}\frac{W}{L}\mu C'_{OX}\left(V_{GS} - V_{TH}\right)^2\left(1 + \lambda V_{DS}\right)$$

Low-Frequency Small-Signal Model - 1

The small-signal operation of a MOSFET at low frequencies can be represented by the following equivalent circuit:



The output conductance may be replaced by an equivalent resistance

For small-signal gain applications MOSFETs are biased in the saturation region. Here:

$$I_{DS} \approx \frac{1}{2}\frac{W}{L}\mu C'_{OX}\left(V_{GS}-V_{TH}\right)^2 \Rightarrow \left(V_{GS}-V_{TH}\right)=\sqrt{2I_{DS}\frac{L}{W}\frac{1}{\mu C'_{OX}}}$$

$$g_m=\frac{dI_{DS}}{dV_{GS}} \approx \frac{W}{L}\mu C'_{OX}\left(V_{GS}-V_{TH}\right)=\sqrt{2\frac{W}{L}\mu C'_{OX}\,I_{DS}}$$

For given device dimensions, the transconductance depends on the square root of I_{DS}. (Recall that transconductance was directly proportional to collector current for bipolars.)

Low-Frequency Small-Signal Model - 2

Output conductance:

$$I_{DS} = \frac{1}{2}\frac{W}{L}\mu C'_{OX}\left(V_{GS} - V_{TH}\right)^2\left(1 + \lambda V_{DS}\right)$$

$$g_{ds} = \frac{dI_{DS}}{dV_{DS}} = \frac{1}{2}\frac{W}{L}\mu C'_{OX}\left(V_{GS} - V_{TH}\right)^2\lambda \approx \lambda I_{DS}$$

Bulk Potential =

$$\varphi_B = \frac{kT}{q}\ln\left(\frac{N_A}{n_i}\right)$$

N_A = doping concentration in the channel region

Output conductance is proportional to current similar to a bipolar transistor.

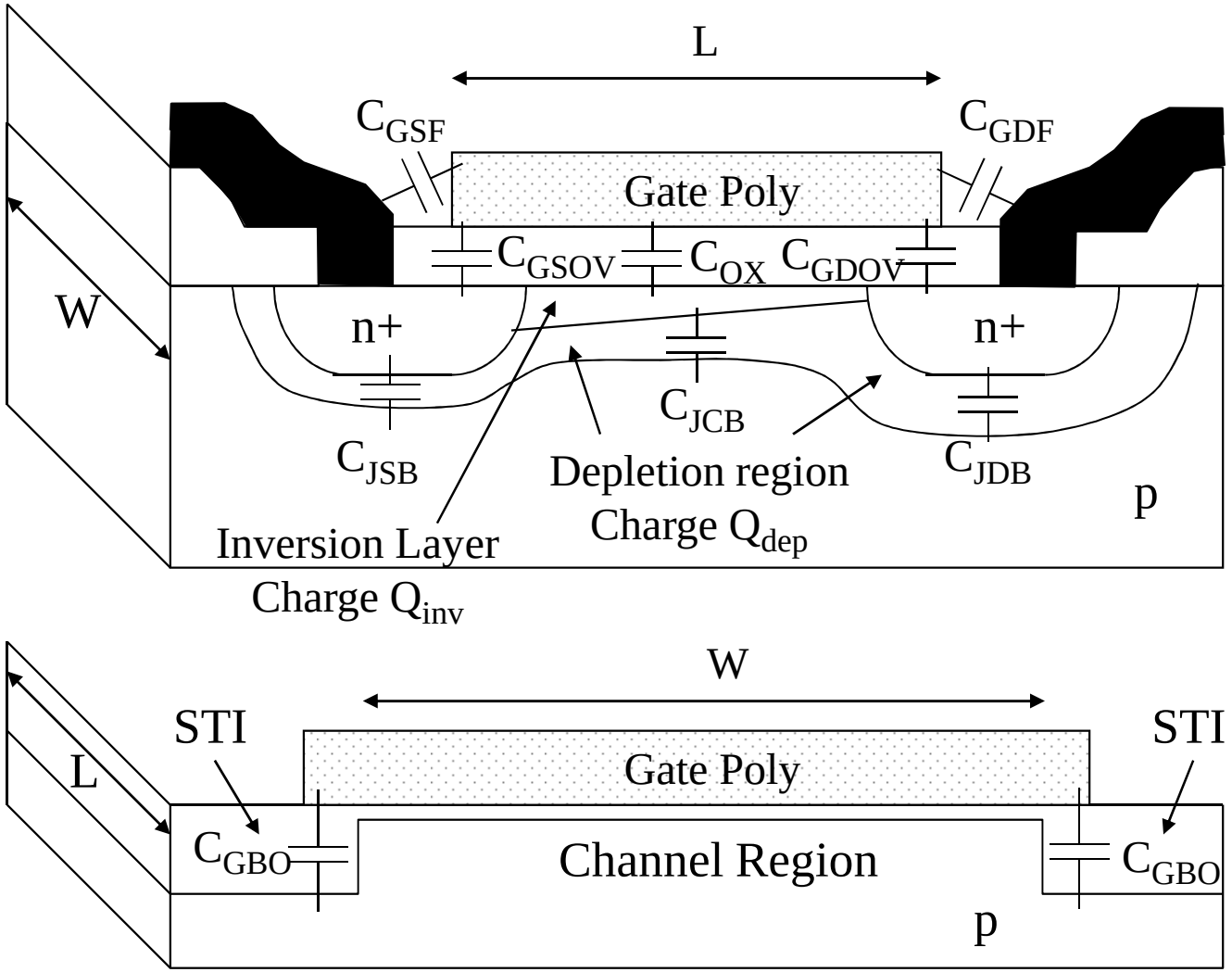
Because V_{BS} influences V_{TH} , it enters the small-signal model through g_{mbs}

$$V_{TH} = V_{TH0} + \gamma\left(\sqrt{2\varphi_B - V_{BS}} - \sqrt{2\varphi_B}\right) \Rightarrow \frac{dV_{TH}}{dV_{BS}} = -\frac{\gamma}{2\sqrt{2\varphi_B - V_{BS}}}$$

$$I_{DS} \approx \frac{1}{2}\frac{W}{L}\mu C'_{OX}\left(V_{GS} - V_{TH}\right)^2 \Rightarrow \frac{dI_{DS}}{dV_{TH}} = -\frac{W}{L}\mu C'_{OX}\left(V_{GS} - V_{TH}\right) = -g_m$$

$$g_{mbs} = \frac{dI_{DS}}{dV_{BS}} = \frac{dI_{DS}}{dV_{TH}}\frac{dV_{TH}}{dV_{BS}} = \left(-g_m\right)\left(-\frac{\gamma}{2\sqrt{2\varphi_B - V_{BS}}}\right) = \frac{\gamma}{2\sqrt{2\varphi_B - V_{BS}}}g_m$$

MOSFET Capacitances - 1



MOSFET Capacitances - 2

The capacitances and charge storage associated with a MOSFET fall into two categories – *intrinsic* and *extrinsic*. The *intrinsic* capacitances occur because of the nature of the MOSFET and have a complicated voltage dependence. The *extrinsic* capacitances arise from parasitic elements of the MOSFET. They are usually constant or have a pn-junction type voltage dependence. Ideally, the process will be designed to minimize these where possible.

Extrinsic Capacitance Elements:

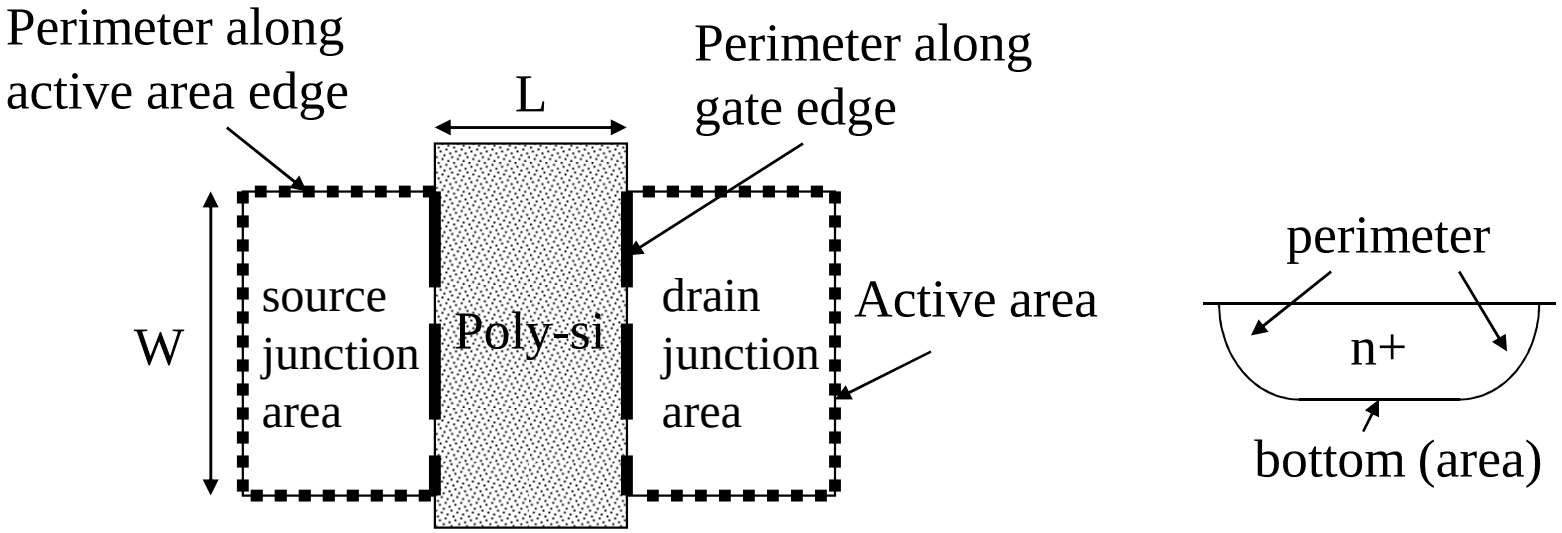
Gate-to-source and gate-to-drain overlap capacitances (C_{GSOV} and C_{GDOV}): These are parallel-plate oxide capacitances which arise due to the overlap of the gate polysilicon with the source and drain regions. These capacitances increase with device width W.

Gate-to-source and gate-to-drain fringing capacitances (C_{GSF} and C_{GDF}): These arise due to the fringing electric fields between the sidewalls of the poly gate and the metallisation which contact the source and drain regions. They are in parallel with the overlap capacitances and are usually "lumped in" with these overlap capacitances in most models but can be separately specified in some models.

Gate-to-bulk overlap capacitance (C_{GBO}): These occur at each side of the device where the gate polysilicon overlaps the thick oxide regions. Even though this occurs at each side of the gate as the capacitances are in parallel they are usually considered as just one overlap capacitor. This overlap capacitance increases with device length L.

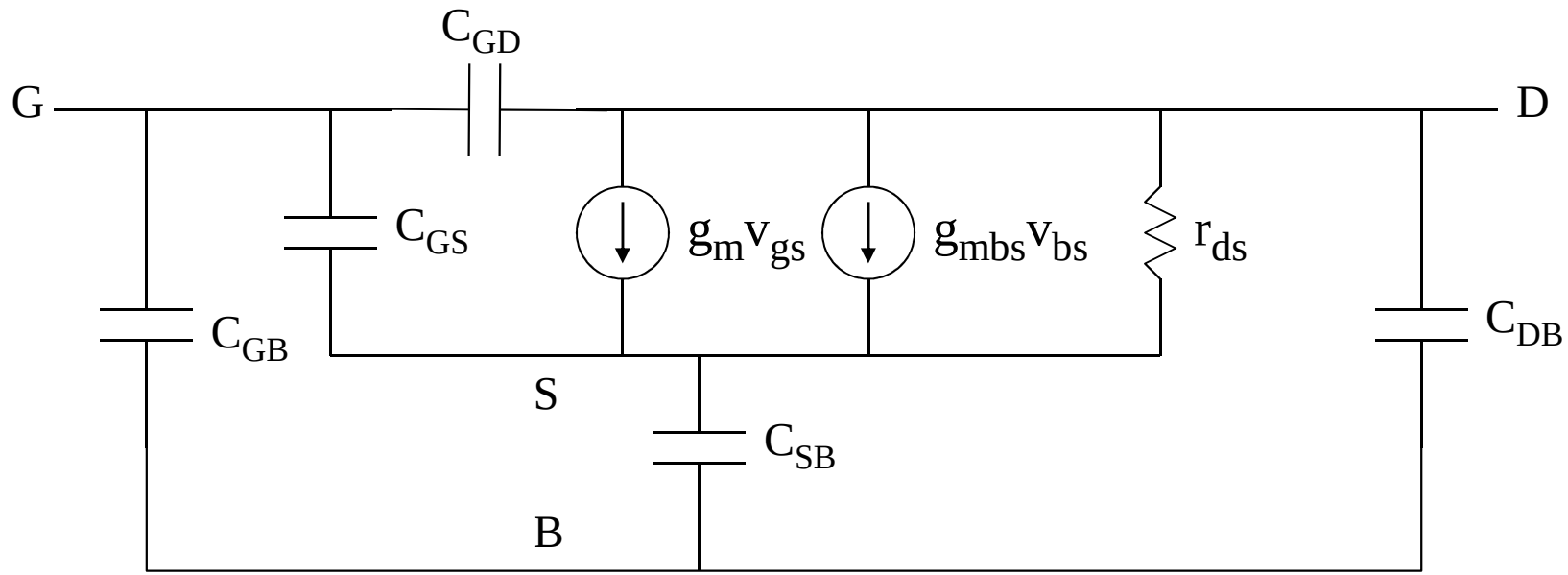
Source-to-bulk and drain-to-bulk depletion capacitances (C_{JSB} and C_{JDB}): These are the capacitances associated with the (normally reverse biased) source/bulk and drain/bulk pn junctions.

Geometrical Definitions for Junction Caps.



Because the bottom plate of the implanted/diffused regions generally has a different doping profile from the sidewalls (perimeters) the voltage dependence of the depletion capacitances is different so they can be modelled separately in SPICE. Some models allow a distinction to be made between the active area perimeter and the gate edge perimeter.

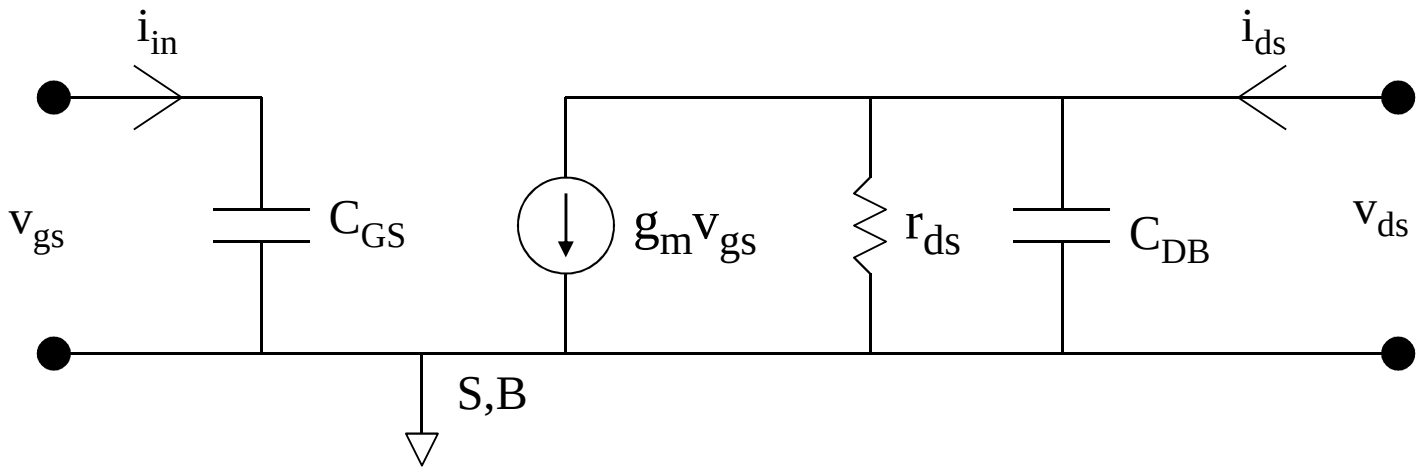
Common-Source Small-Signal Circuit



The above circuit gives a satisfactory representation of MOSFET operation up to "reasonably high" frequencies.

In many circuit topologies the bulk and source are connected together eliminating the g_{mbs} and C_{SB} elements from the equivalent circuit. Because C_{GB} is usually much smaller than the other capacitances in the circuit it is often ignored. Similarly, for a first-order analysis C_{GD} is often neglected.

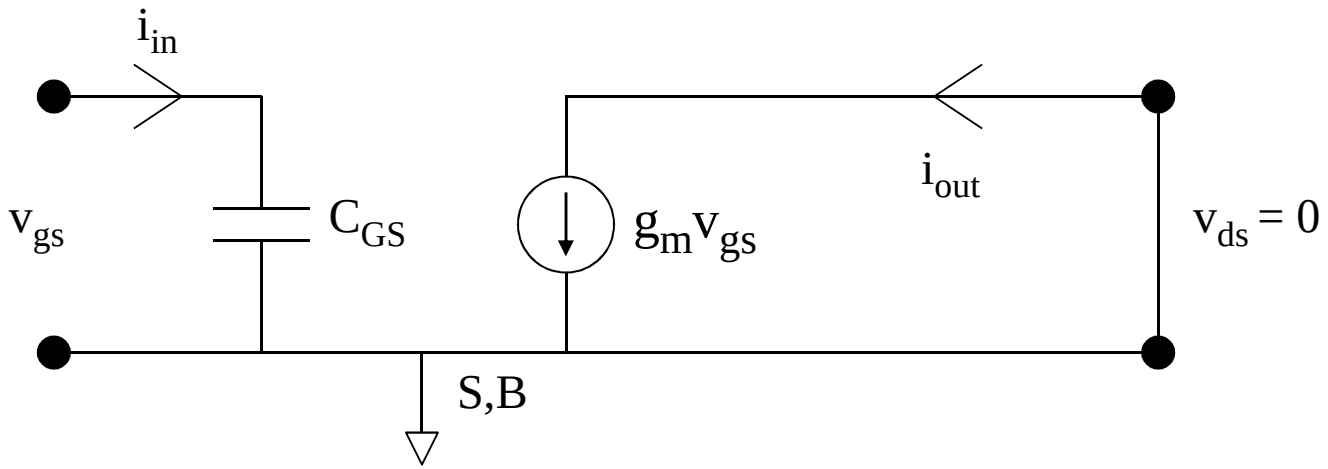
Simplified MOSFET Small-Signal Circuit



This has just 4 elements:

1. Input capacitance
2. Transconductance
3. Output resistance
4. Output capacitance (even this is ignored in a lot of cases or lumped in with the input capacitance of the following stage)

Cut-Off Frequency - 1



Just like the bipolar, the cut-off frequency is defined as the frequency at which the output short-circuit current gain drops to unity (1). Because the output is short-circuited the output resistance and capacitance can be ignored.

$$i_{in} = j\omega C_{GS} v_{gs}$$

$$i_{out} = g_m v_{gs}$$

$$h_{21} = \left.\frac{i_{out}}{i_{in}}\right|_{v_{out}=0} = \frac{g_m v_{gs}}{j\omega C_{GS} v_{gs}} = \frac{g_m}{j\omega C_{GS}}$$

Cut-off Frequency - 2

$$h_{21} = \frac{g_m}{j\omega C_{GS}} = \frac{g_m}{j2\pi f\, C_{GS}} \Rightarrow \left|h_{21}\right| = \frac{g_m}{2\pi f\, C_{GS}}$$

The cut-off frequency f_T is the frequency at which this drops to 1:

$$1 = \frac{g_m}{2\pi f_T\, C_{GS}} \Rightarrow f_T = \frac{g_m}{2\pi C_{GS}}$$

So just like for a BJT, for a MOSFET the cut-off frequency is determined by the ratio of the transconductance to the input capacitance.

It would have been more realistic to include C_{GD} in the derivation giving:

$$f_T = \frac{g_m}{2\pi(C_{GS} + C_{GD})}$$

Cut-off Frequency - 3

The cut-off frequency f_T is given by: $f_T = \frac{g_m}{2\pi C_{GS}}$

g_m in saturation is: $g_m = \frac{W}{L}\mu C'_{OX}\left(V_{GS} - V_{TH}\right)$

A common estimate of C_{GS} in saturation is: $C_{GS} = \frac{2}{3}WLC'_{OX}$

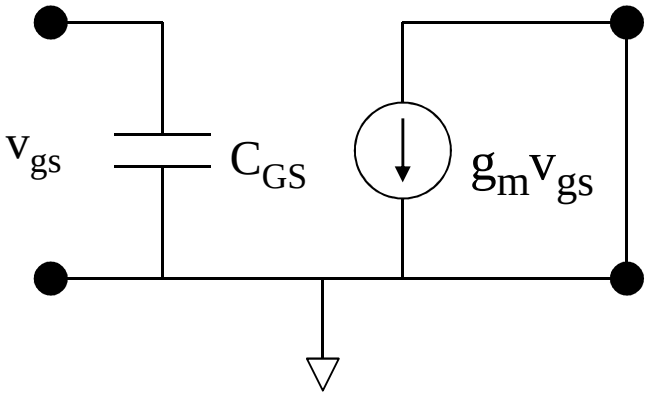
Putting these together:

$$f_T = \frac{g_m}{2\pi C_{GS}} = \frac{\frac{W}{L}\mu C'_{OX}\left(V_{GS} - V_{TH}\right)}{2\pi\frac{2}{3}WLC'_{OX}} = \frac{3\mu\left(V_{GS} - V_{TH}\right)}{4\pi L^2}$$

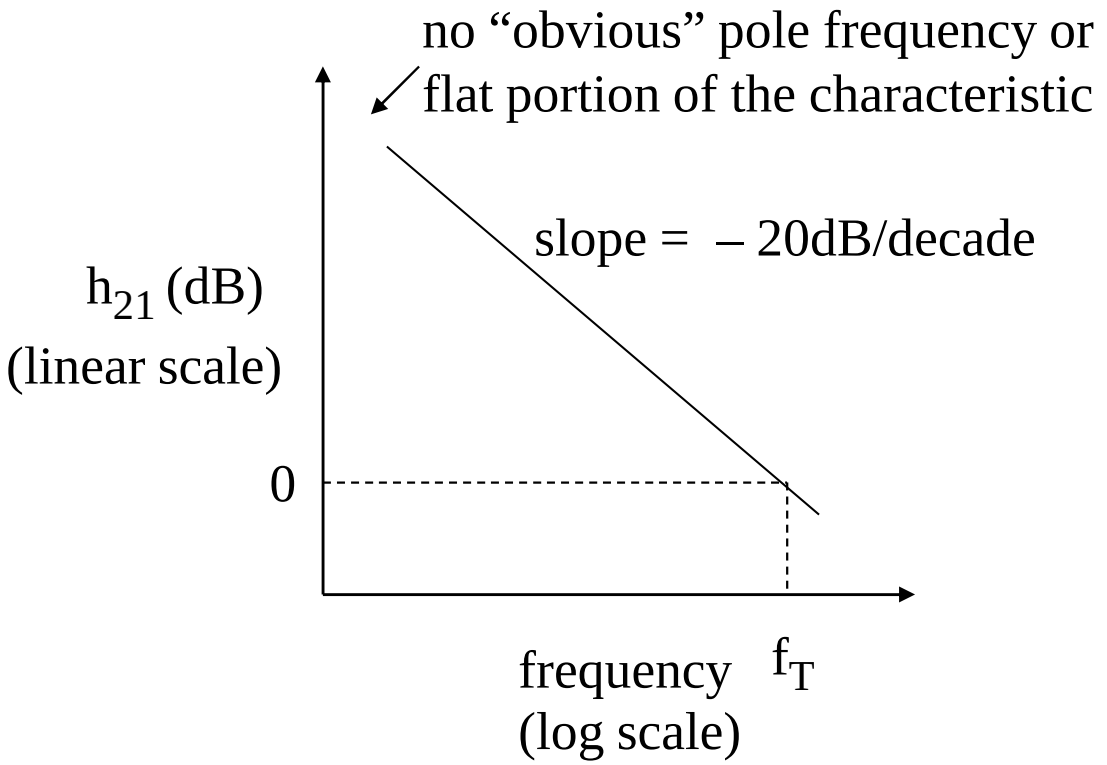
In this simple model: $f_T \propto \frac{1}{L^2}$

This shows that cut-off frequency rises rapidly as L decreases which is a very good motivation for making smaller devices.

h_{21} vs. frequency for the MOSFET

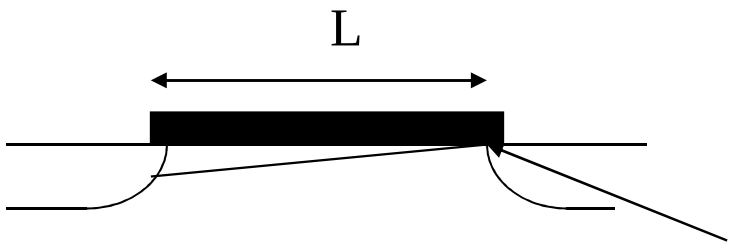


$$|h_{21}| = \frac{g_m}{2\pi f C_{GS}}$$



In the equivalent circuit of the MOSFET the input is purely capacitive. Therefore the input current varies as 1/f for all frequencies and h_{21} varies with 1/f also i.e. there is no low-frequency flat portion of h_{21} as in the BJT because there is no resistance drawing a constant (w.r.t. frequency) small amount of current.

Channel Transit Time



Channel pinch-off
$V_{DS}=V_{GS}-V_{TH}$

At the onset of saturation the drain voltage is $V_{DS}=V_{DSAT}=V_{GS}-V_{TH}$
At this bias the average electric field along the channel and the average carrier velocity are (very approximately):

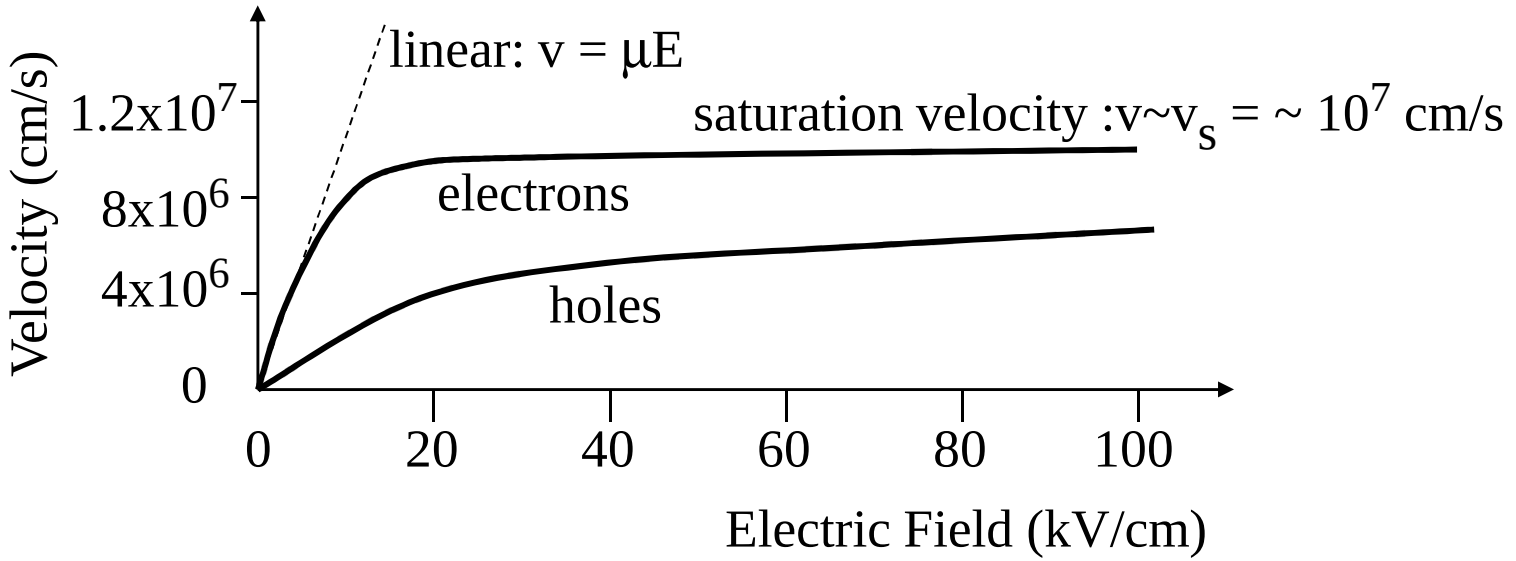
$$E = \frac{V_{DSAT}}{L} \qquad v = \mu E = \mu \frac{V_{DSAT}}{L}$$

The time for a carrier to cross the channel (the transit time) is then approximately:

$$\tau = \frac{L}{v} = \frac{L}{\mu \frac{V_{DSAT}}{L}} = \frac{L^2}{\mu V_{DSAT}}$$

It would seem “reasonable” that the cut-off frequency should be inversely proportional to the transit time – i.e. it should be inversely proportional to L^2.

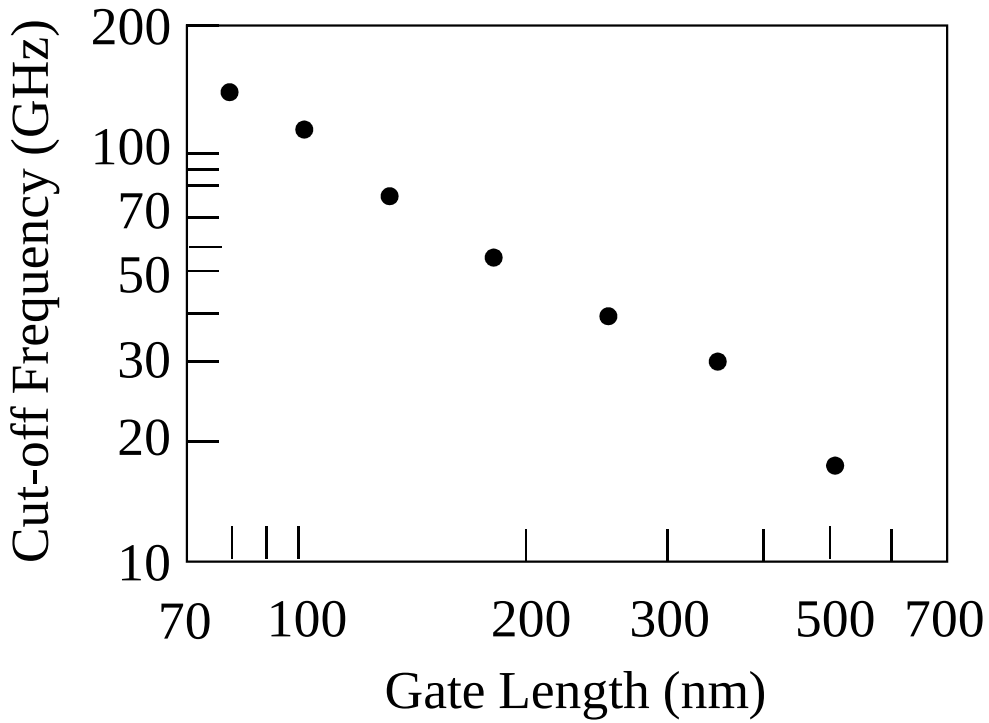
Velocity Saturation



For low fields (< few kV/cm) the carrier velocity increases linearly with field with a constant mobility μ. For high fields (~ 10kV/cm for electrons) the carriers experience more scattering which tends to reduce the effective mobility and the velocity saturates at a maximum value known as the saturation velocity. MOSFETs with short gate lengths operate under velocity saturation conditions in saturation. This changes the f_T vs. L relationship:

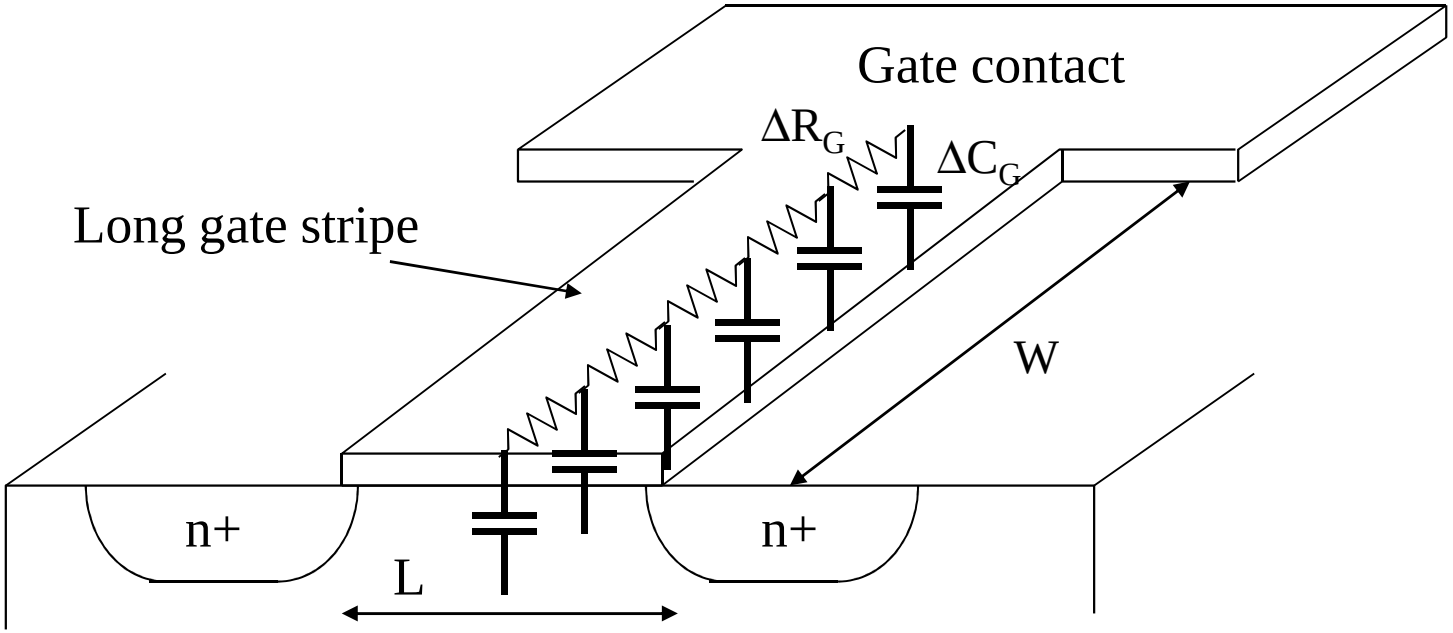
$$v = v_{sat} \Rightarrow \tau = \frac{L}{v} = \frac{L}{v_{sat}} \Rightarrow \tau \propto L \Rightarrow f_T \propto \frac{1}{L}$$

Reported N-MOSFET f_T vs. Channel Length



The points in this graph correspond to results reported by different authors/organizations in papers and conferences.

MOSFET Gate Resistance



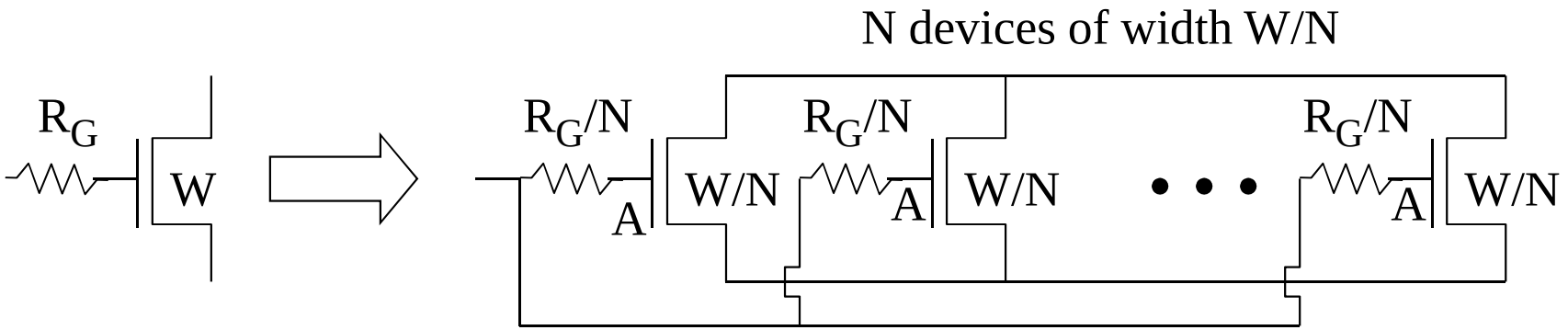
MOSFETs for RF/microwave applications are usually wide so the gate is a long narrow stripe. The resistance of the gate polysilicon in conjunction with the oxide capacitance behaves like a distributed RC network. This can be represented by an equivalent circuit of just one capacitance and one resistance. The capacitance is the total oxide capacitance, whereas the resistance is 1/3 of its "geometrical" value i.e.:

$$R_G = \frac{1}{3}\frac{W}{L}R_{SQ}$$

where R_{SQ} is the resistance per square of the gate material

Multi-fingered Devices

Replacing a very wide device with several narrower devices (with the same gate length as the original) connected in parallel is a common means of reducing the gate resistance in both MOSFETs and MESFETs.

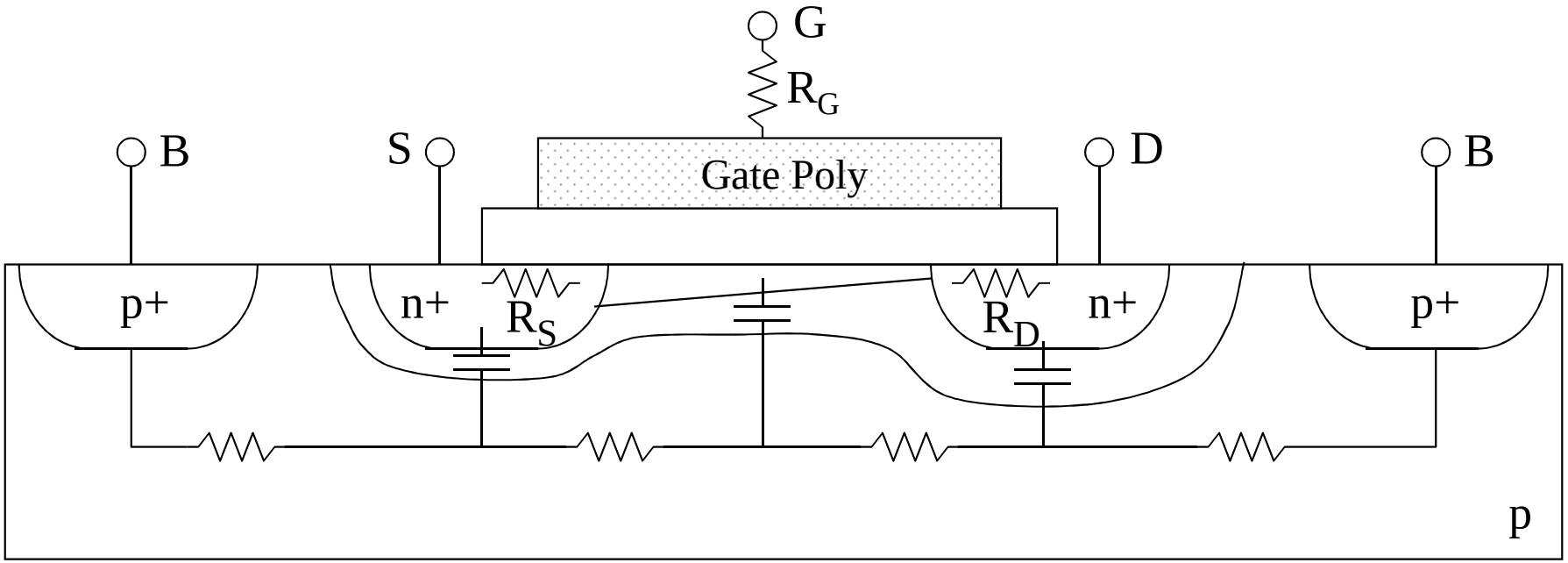


All the smaller MOSFETs (stripes) are all identical so the points labeled “A” will be at the same voltage (equipotential). Therefore the configuration is equivalent to having N resistors of value R_G/N connected in parallel giving an equivalent resistance of:

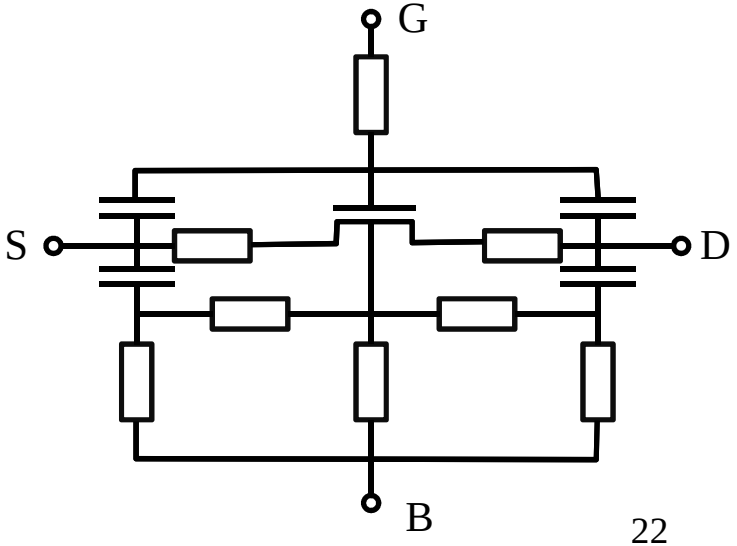
$$R_{Geff} = \frac{R_G / N}{N} = \frac{R_G}{N^2} = \frac{1}{3N^2}\frac{W}{L}R_{SQ}$$

Similarly, if each stripe has a gate contact on both sides of the stripe, then this gives a further reduction in resistance by a factor of 4 compared to the case where the gates are only contacted on one side.

Substrate Resistance Networks



The substrate of the MOSFET forms a distributed resistance system which can be represented by a 3-port network. It has been found that the resistance network can be represented with sufficient accuracy by 3 to 5 resistors.



Sample Exam Question

Q5, EE4005, Summer 2003

(a) Draw appropriate cross-sections of a MOSFET and label the main capacitance elements which contribute to the small-signal equivalent circuit.

(b) Using simplifications as appropriate, determine an expression for the cut-off frequency of a MOSFET used in a common-source configuration and biased in saturation. MOSFET current in saturation may be approximated as

$$I_{DS} = \frac{1}{2}\frac{W}{L}\mu_0 C'_{OX}\left(V_{GS} - V_{TH}\right)^2$$

where the symbols have their usual meaning.

(c) Use the expression derived in part (b) to determine a relationship between the cut-off frequency and the channel length of a MOSFET.

Sample Exam Question

Q2, EE4011, Autumn 2006

(a) Show a small-signal model of a MOS transistor suitable for first-order analysis and from this derive an expression for the cut-off frequency of a MOS transistor. Assume the transistor is biased in saturation and that the current can be approximated by:

$$I_{DS} = \frac{1}{2}\frac{W}{L}\mu C_{OX}(V_{GS} - V_{TH})^2(1 + \lambda V_{DS})$$

where the symbols have their usual meaning. Also assume that the only capacitance to be considered is the gate-source capacitance.

[10 marks]

(b) An NMOS transistor is biased in saturation and configured as a common-source two-port amplifier with the input applied to the gate (port 1) and the output taken from the drain (port 2) with the following bias conditions and device parameters:

W=10μm, L=0.25μm, Tox=4nm, μ=400cm^2/Vs, V_{GS}=2.5V, V_{DS}=2V,

V_{TH}=0.5V, λ=0.1 V^{-1}.

Determine:

(i) The cut-off frequency [2 marks]

(ii) The 4 two-port y-parameters at a frequency of 1GHz. [8 marks]