**Dynamic Behavior**

The transient behavior of a *pn-junction* was dominated by:

- The movement of **excess minority carrier charge** in the neutral zones.
- The movement of **space charge** in the depletion region.

MOSFETs are *majority carrier* devices.

Their dynamic behavior is determined solely by the time to:

- Charge and discharge the capacitances between the device ports.
- Charge and discharge of the interconnecting lines.

These capacitances originate from three sources:

- The basic MOS structure.
- The channel charge.
- The depletion regions of the reverse-biased pn-junctions of drain and source.

Aside from the MOS structure capacitances, all capacitors are **nonlinear** and **vary with the applied voltage**.

**Dynamic Behavior**

MOS Structure Capacitances:

The gate of a MOS transistor is isolated from the channel by the gate oxide where:

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

For the I-V equations, it is useful to have $C_{ox}$ as **large** as possible, by keeping the oxide very thin.

This capacitance is called gate capacitance and is given by:

$$C_g = C_{ox} WL$$

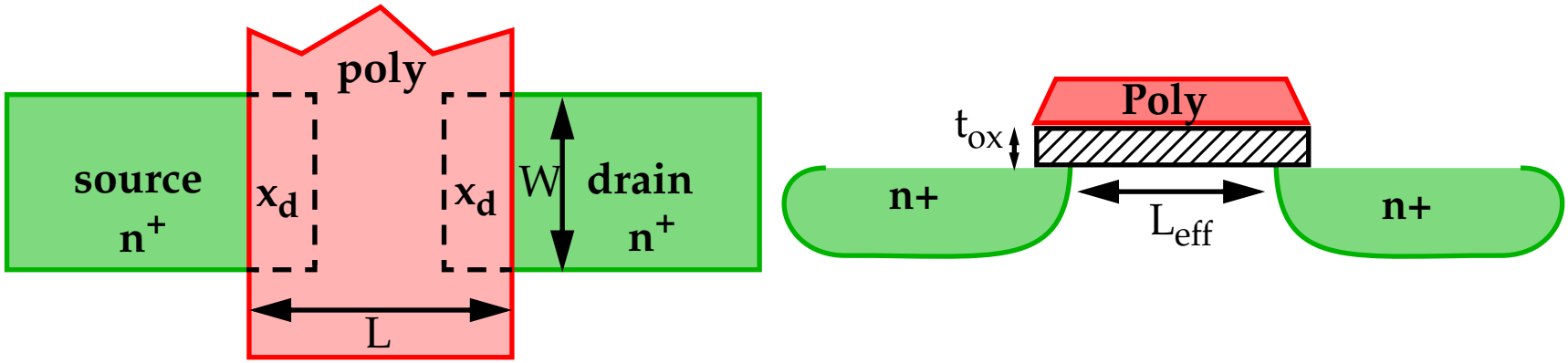
This gate capacitance can be decomposed into several parts:

- One part contributes to the channel charge.
- A second part is due to the topological structure of the transistor.

Let's consider the latter first.

**Dynamic Behavior**

MOS Structure Capacitances, Overlap:




Lateral diffusion: source and drain diffusion extend under the oxide by an amount $x_d$.

The effective channel length ($L_{eff}$) is less than the *drawn length* L by $2*x_d$.

This also gives rise to a linear, fixed capacitance called **overlap capacitance**.

$$C_{gsO} = C_{gdO} = C_{ox}x_d W = C_O W$$

Since $x_d$ is technology dependent, it is usually combined with $C_{ox}$.

**Dynamic Behavior**

Channel Charge:

The gate-to-channel capacitance is composed of three components, $C_{gs}$, $C_{gd}$ and $C_{gb}$.

Each of these is non-linear and dependent on the region of operation.

Estimates or average values are often used:

- Triode: $C_{gb}$ ~=0 since the inversion region shields the bulk electrode from the gate.
- Saturation: $C_{gb}$ and $C_{gd}$ is ~= 0 since the channel is pinched off.

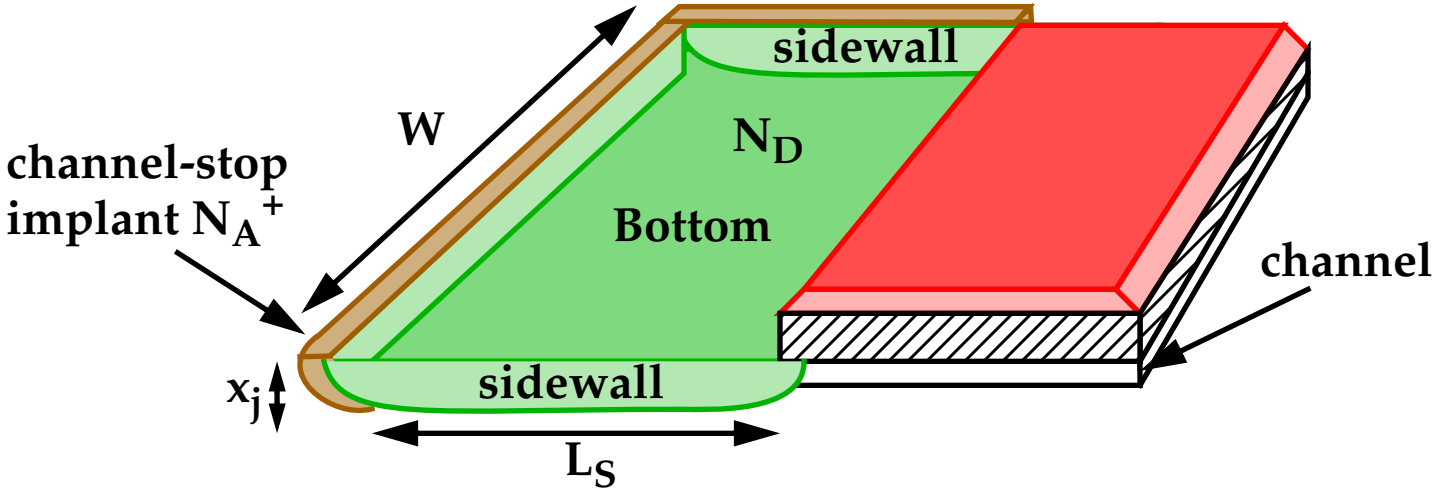
| Operation Region | $C_{gb}$ | $C_{gs}$ | $C_{gd}$ |
|---|---|---|---|
| Cutoff | $C_{ox}WL_{eff}$ | 0 | 0 |
| Triode | 0 | $C_{ox}WL_{eff}/2$ | $C_{ox}WL_{eff}/2$ |
| Saturation | 0 | $(2/3)C_{ox}WL_{eff}$ | 0 |

**Dynamic Behavior**

Junction or Diffusion Capacitances:

This component is caused by the reverse-biased source-bulk and drain-bulk *pn-junctions*.

We determined that this capacitance is **non-linear** and *decreases* as reverse-bias is *increased*.




- Bottom-plate junction:

  Depletion region capacitance is:

  $$C_{bottom} = C_j W L_S$$

  with a grading coefficient of m = 0.5 (for an abrupt junction)

**Dynamic Behavior**

MOS Structure Capacitances, Junction or Diffusion:

- Side-wall junction:

Formed by the source region with doping $N_D$ and the $p^+$channel-stop implant with doping $N_A{}^+$.

Since the channel-stop doping is usually higher than the substrate, this results in a higher unit capacitance:

$$C_{sw} = C'_{jsw} x_j (W + 2 \times L_S)$$

with a grading coefficient of m = 1/3.

Note that the channel side is not included in the calculation.

$x_j$ is usually technology dependent and combined with $C'_{jsw}$ as $C_{jsw}$.

**Total junction** (small-signal) capacitance is:

$$C_{diff} = C_{bottom} + C_{sw} = C_j \times AREA + C_{jsw} \times PERIMETER$$

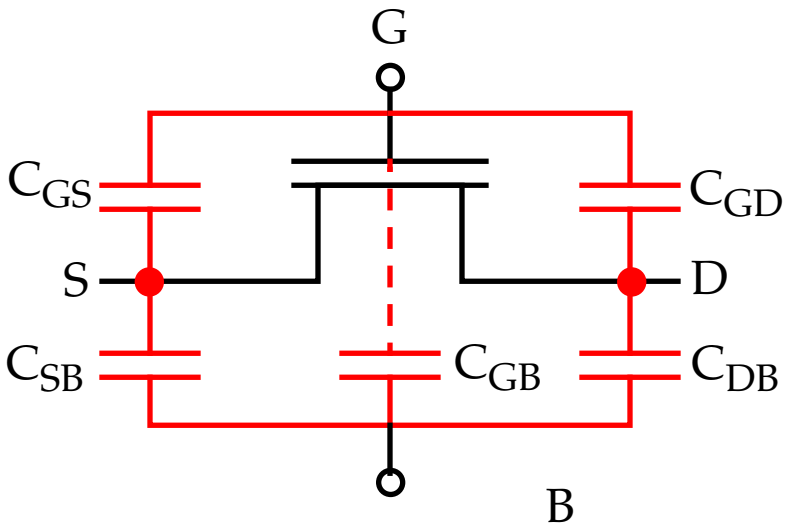
$$C_{diff} = C_j L_S W + C_{jsw}(2L_S + W)$$

As we've done before, we linearize these and use average cap.

**Dynamic Behavior**

Capacitive Device Model:

The previous model can be summarized as:




$C_{GS} = C_{gs} + C_{gsO}$

$C_{GD} = C_{gd} + C_{gdO}$

$C_{GB} = C_{gb}$

$C_{SB} = C_{Sdiff}$

$C_{DB} = C_{Ddiff}$

The dynamic performance of digital circuits is directly proportional to these capacitances.

**Dynamic Behavior**

Example:

Given:

$t_{ox} = 6nm$ $\quad C_O = 3 \times 10^{-10} F/m$

$L = 0.24um$ $\quad C_{j0} = 2 \times 10^{-3} F/m^2$

$W = 0.36um$ $\quad C_{jsw0} = 2.75 \times 10^{-10} F/m$

$L_D = L_S = 0.625um$

Determine the zero-bias value of all relevant capacitances.

Gate capacitance, $C_{ox}$, per unit area is derived as:

$$C_{ox} = \varepsilon_{ox}/t_{ox} = \frac{3.5 \times 10^{-2} fF/um}{6 \times 10^{-3} um} = 5.8 fF/um^2$$

Total gate capacitance $C_g$ is:

$$C_g = WLC_{ox} = 0.36um \times 0.24um \times 5.8 fF/um^2 = 0.5 fF$$

Overlap capacitance is:

$$C_{GSO} = C_{GDO} = WC_O = 0.108 fF$$

Total gate capacitance is:

$$C_{gtot} = C_g + 2 \times C_{GSO} = 0.716 fF$$

**Dynamic Behavior**

Example (cont):

Diffusion capacitance is the sum of bottom:

$$C_{j0} L_D W = 2\,fF/um^2 \times 0.625um \times 0.36um = 0.45\,fF$$

Plus side-wall (under zero-bias):

$$C_{jsw0}(2L_D + W) = 2.75\times10^{-1}(2 \times 0.625um + 0.36um) = 0.44\,fF$$

In this example, diffusion capacitance dominates gate capacitance (0.89 fF vs. 0.716 fF).

Note that this is the worst case condition. Increasing reverse bias reduces diffusion capacitance (by about 50%).

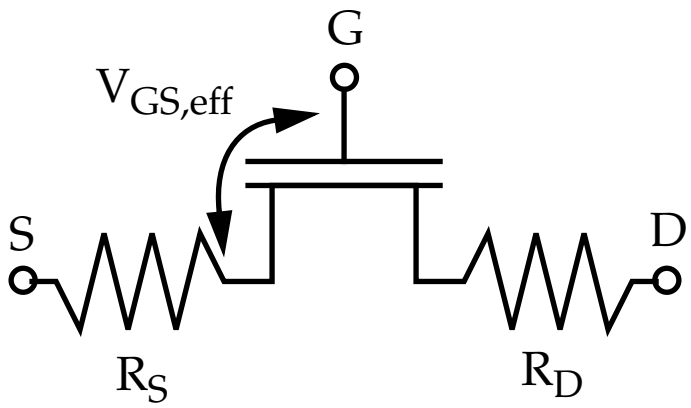
Also note that side-wall dominates diffusion. Advanced processes use $SiO_2$ to isolate devices (trench isolation) instead of $N_A^+$ implant.

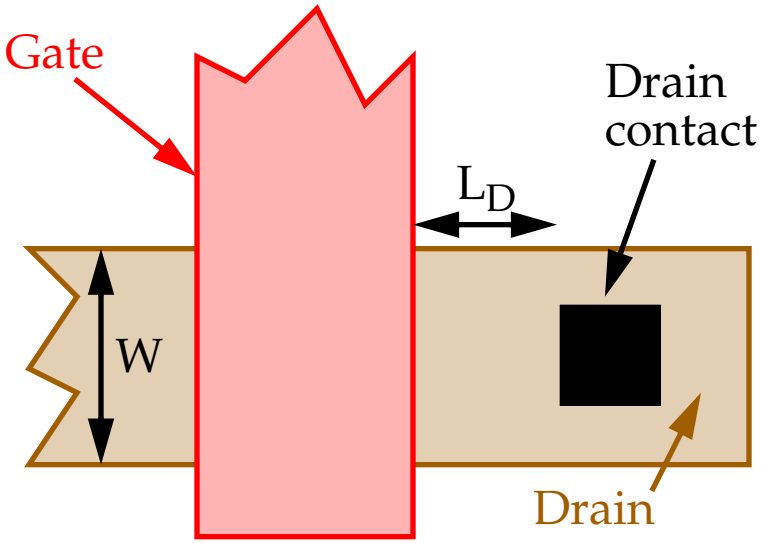
Usually, diffusion is at most equal to gate, very often it is smaller.

**Source-Drain Resistance**

Scaling causes junctions to be *shallower* and contact openings to be *smaller*.

This increases the parasitic resistance in series with the source and drain.




This resistance can be expressed as:

$$R_{S,D} = \frac{L_{S,D}}{W} R_{\square} + R_C$$

$R_C$ = Contact Resistance

$R_{\square}$ = Sheet resistance $(2\Omega - 100\Omega)$

$L_{S,D}$ = length of source/drain region.

The series resistance degrades performance by decreasing drain current.

**Silicidation** used -- low-resistivity material such as *titanium* or *tungsten*.

**Secondary Effects**

**Long-channel devices**:

One-dimensional model discussed thus far.

Assumed:

- All current flows on the surface of the silicon.
- Electric fields are oriented along that plane.

Appropriate for manual analysis.

**Short-channel device**:

Ideal model does not hold well when device dimensions reach sub-micron range.

The length of the channel becomes comparable to other device parameters such as the depth of the drain and source junctions.

Two-dimensional model is needed.

Computer simulation required.

**Secondary Effects**

Threshold variations:

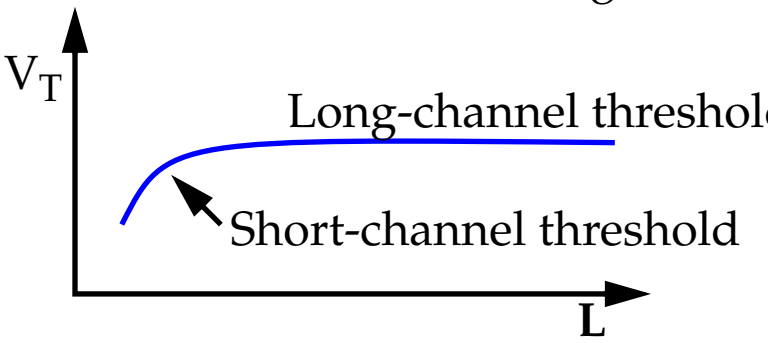
Ideal model assumed *threshold voltage* was only a function of technology parameters and applied body bias, $V_{SB}$.

With smaller dimensions, the $V_{T0}$ becomes a function of L, W and $V_{DS}$.

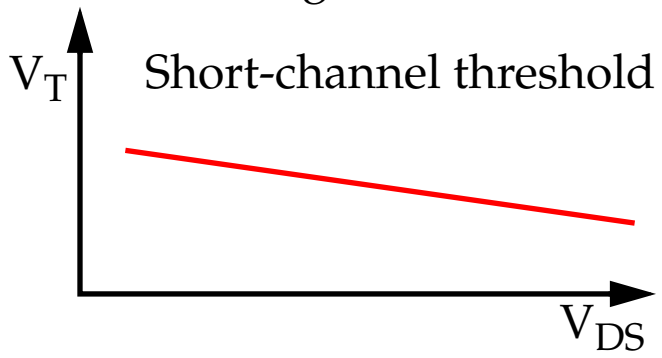
For example, the expression for $V_{T0}$ assumed that all depletion charge beneath the gate originates from the MOS field effects.

We ignored the source and reverse-biased drain depletion regions.

These depletion regions extend under the gate, which in turn **reduces** the threshold voltage necessary to cause strong inversion.




Threshold as a function of length (for low $V_{DS}$)




Drain-induced barrier lowering for small L.

**Secondary Effects**

Threshold variations:

Also, threshold *decreases* with *increasing* $V_{DS}$.

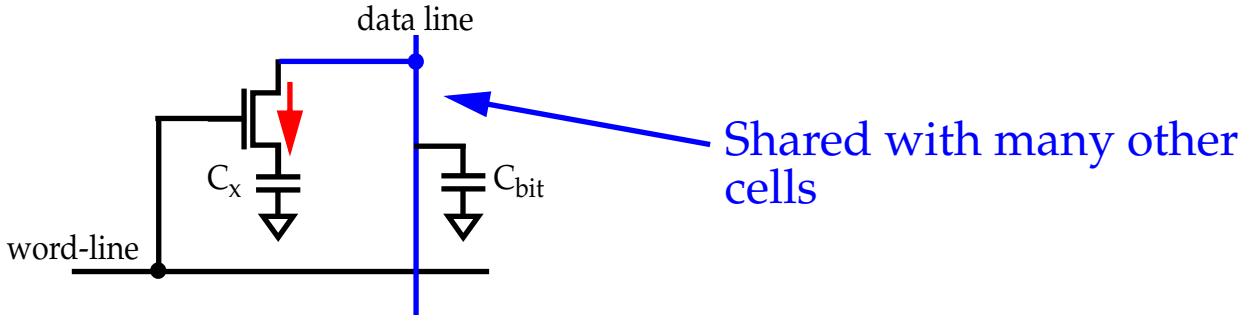
This effect is called *drain-induced barrier lowering* (**DIBL**).

For high values of $V_{DS}$, the source and drain depletion regions can short together (**punch-through**).

**DIBL** is a more serious issue than the variation in $V_{T0}$ as a function of length (since most transistors are minimum length transistors).

Particularly for DRAMs.

Leakage current of a cell (e.g. subthreshold current of the access transistor) is a function of voltage on the data line.



**Secondary Effects**

Threshold variations:

Threshold drift also occurs for short-channel devices over time as a result of **hot-carrier effects**.

In the past, *constant voltage scaling* was used which increased the electric field strength and velocity of the electrons.

The electrons can leave the silicon and tunnel into the gate oxide, given enough energy.

Trapped electrons in the oxide *increase* the threshold of NMOS devices and *decrease* the threshold of PMOS devices.

Field strengths of $10^4$V/cm are easily reached in submicron devices.

This problem causes **long-term reliability problems**.