Transistor Explained

Jiajie Cheng

March 20, 2025

1 Diode Principle

1.1 Fermi-Dirac Distribution

For an electron going around an atom, qunatum mechanics tells us that it will occupy discrete energy levels. If there is a large number of electrons in a system, like a solid, their energy levels will overlap to form **energy bands**. Filled bands and non-filled bands are referred to as valence bands and conduction bands, respectively. The highest energy level that electrons can occupy is the **Fermi level** (E_F). As shown in fig. 1, at $T = 0$ K, all electrons fill the energy levels below E_F, and none occupy the levels above it. When the temperature increases, some electrons gain enough thermal energy to jump to higher energy levels, above E_F. In thermal equilibrium, the distribution of these electrons among the available energy levels is described by the **Fermi-Dirac distribution function**:

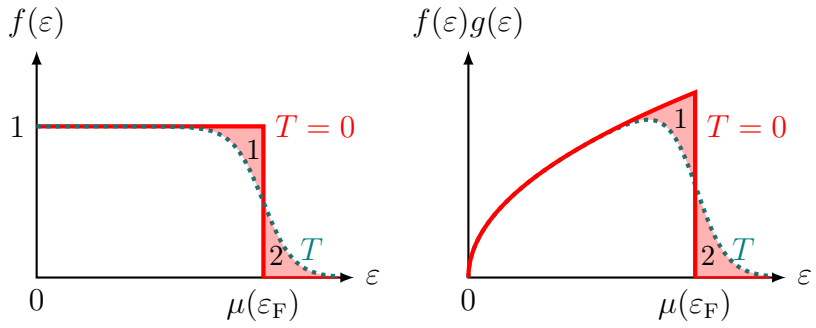


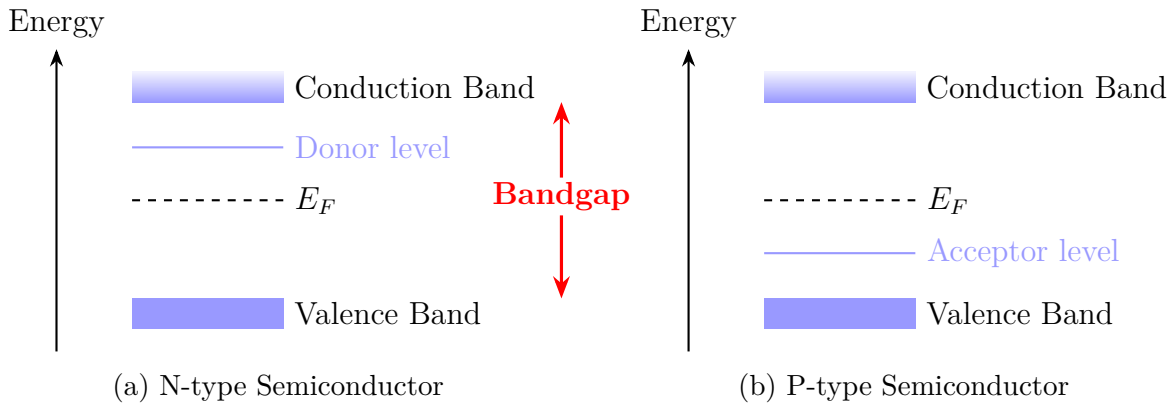
Fig. 1: Distribution function of: particle number and density of state (DOS), if $g(\varepsilon) \propto \varepsilon^{\frac{1}{2}}$

Because valence bands are all filled for insulators, the Fermi level is right in the middle of the bandgap. For metals, the Fermi level is in the conduction band. In insulators, as temperature increases, some electrons may have enough thermal energy, narrowing the bandgap, and that's why some of insulators can turn into **semiconductors** at high temperatures.

Some semiconductors can be widely useful in electronic devices, such as silicon (Si) and germanium (Ge), which have a bandgap of about 1.1 eV and 0.66 eV, respectively. Every silicon atom has 4 protons and 4 electrons in its outer shell, forming 4 covalent bonds with its neighboring atoms. So the silicon is neutrally charged. However, if we dope silicon with

some impurities, we can create **n-type** or **p-type** semiconductors. For example, if we add phosphorus (P) atoms, which have 5 valence electrons, to silicon, the extra electron will be free to move around and contribute to electrical conduction. This creates an n-type semiconductor with excess electrons. On the other hand, if we add boron (B) atoms, which have only 3 valence electrons, it will create "**holes**" in the crystal lattice that can accept electrons, creating a p-type semiconductor with excess holes.

To understand doping, we could use band structure to clarify. In n-type semiconductors, the Fermi level is closer to the conduction band, called the **donor level**, because the extra electrons from the dopant atoms can easily jump into the conduction band. In p-type semiconductors, the Fermi level is closer to the valence band, called the **acceptor level**, because the holes can easily accept electrons from the valence band. The difference in Fermi levels between n-type, p-type semiconductors will creates an electric field when connected.



(a) N-type Semiconductor (b) P-type Semiconductor

Fig. 2: Band structure: conduction band (E_C), valence band (E_V), and Fermi level (E_F).

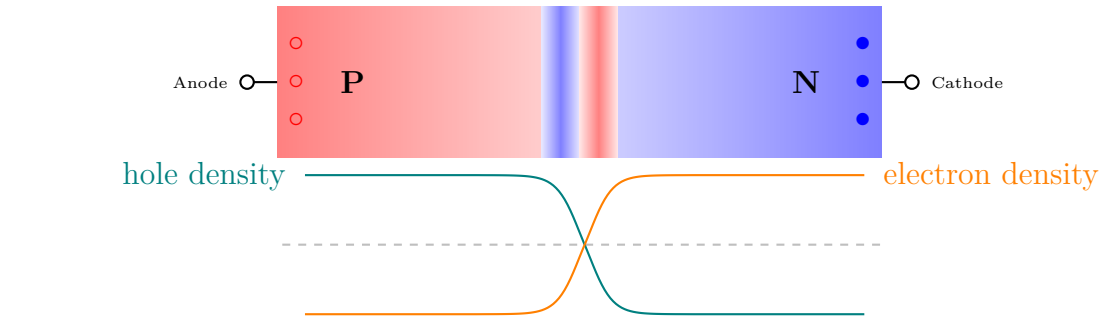
1.2 Depletion Region

If we connect a p-type semiconductor and an n-type semiconductor together, a **pn junction** is formed. At the junction, electrons from the n-type region will diffuse into the p-type region, while holes from the p-type region will diffuse into the n-type region. This diffusion creates a **depletion region** around the junction, where there are no free charge carriers (electrons or holes). The depletion region is electrically neutral but creates an electric field due to the separation of charges.

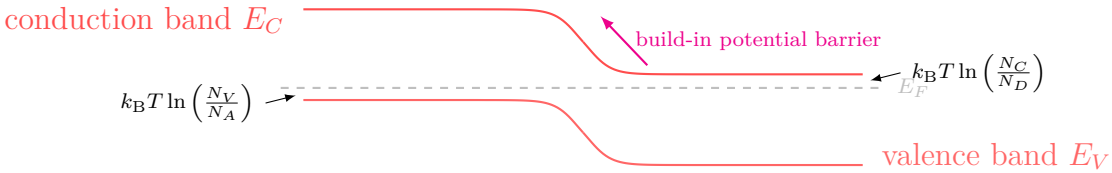
In a PN junction under **forward bias** (whose direction makes current flow readily), the applied voltage reduces the built-in potential barrier, as shown in fig. 3. This allows majority charge carriers—electrons from the n-type region and holes from the p-type region—to move more easily across the junction.

The width of depletion region can be modified by applying an external voltage across the PN junction. That's the principle of **field effect transistor (FET)**, which is similar to a triode, as shown in fig. 5. The depletion region can be expanded or contracted by applying a reverse bias or forward bias, respectively. Different bias voltage may affect the shape of the depletion region, and affect the height of the potential barrier at the junction.

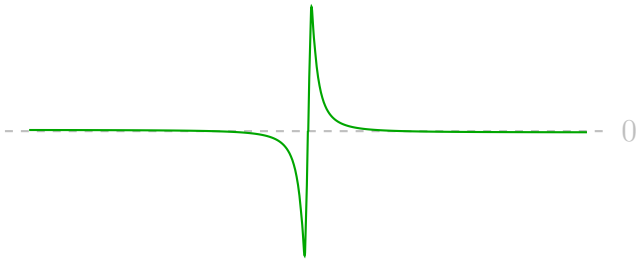
Fig. 3: PN junction diode with depletion region with applied electric field.



(a) charge carrier density sptial distribution



(b) energy band sptial distribution



(c) net charge distribution

- Reverse bias (negative voltage on the p-side, positive on the n-side):
  - Reinforces the built-in electric field, push charge carriers away from the junction, and create an even wider depletion region.
  - example: reduce the gate-source voltage (V_{GS}) in a field-effect transistor (FET) will close the conductive channel between the source and drain, in JFET and MOSFET devices. The I-V curve of the device will fall in the cut-off region (pinch-off region, when $V_{\mathrm{GS}} < V_{\mathrm{TH}}$), where the current is very small or zero.
- Forward bias (positive voltage on the p-side, negative on the n-side):
  - Counteracts the built-in potential barrier, allow more majority charge carriers to cross the junction, and reduce the width of the depletion region.
  - example: increase forward bias voltage will shift the acceptor fermi level up and donor fermi level down, reducing the potential barrier. This allows more charge carriers to flow through the junction, making the diode work. Shockley diode equation can describe the I-V characteristics of a diode ($V_T = \frac{k_B T}{q}$):

$$I = I_0 \left(e^{\frac{V}{n V_T}} - 1 \right) \tag{1.1}$$

2 MOSFET Principle

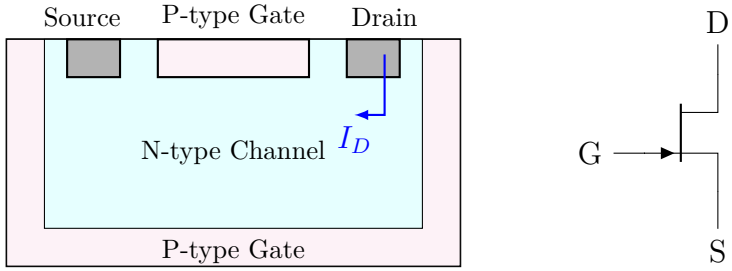


Fig. 5: JFET device (N-Channel) and its circuit symbol

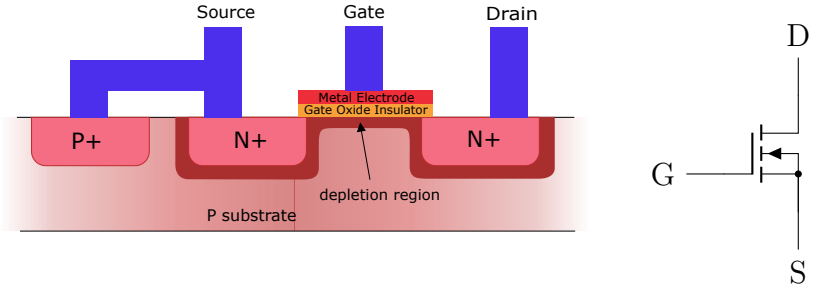


Fig. 6: MOSFET device (N-Channel) and its circuit symbol

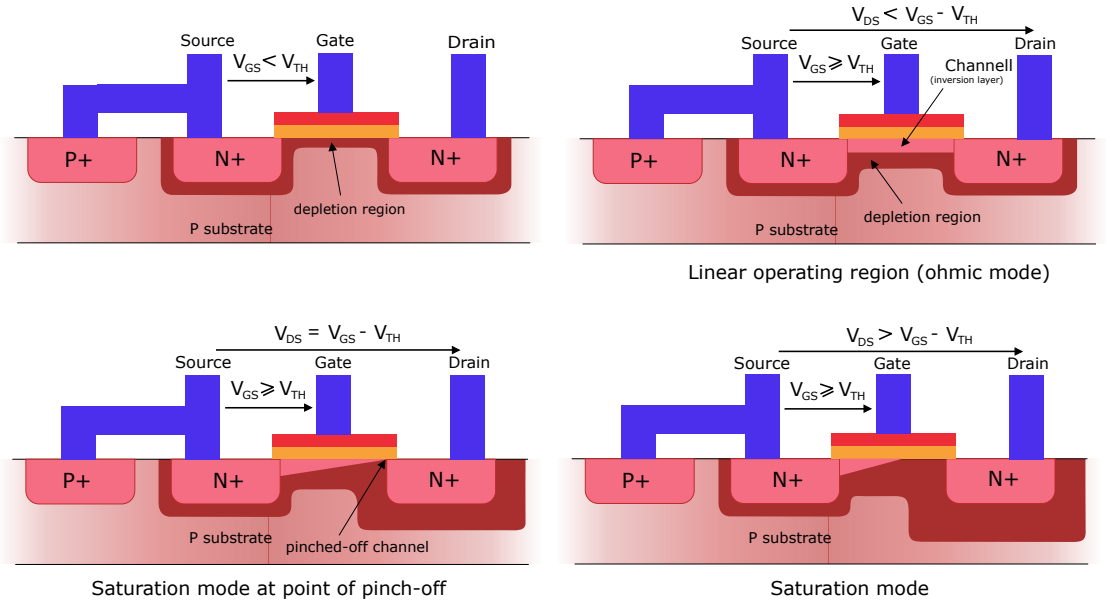


Fig. 7: MOSFET functioning: The operation of an N-type MOSFET (NMOS) in the saturation region. The gate voltage V_{G} is greater than the threshold voltage V_{TH}, creating a conductive channel between the source and drain. The current I_{D} flows from drain to source, controlled by the gate voltage. (Could understand by gate-controlled depletion channel)

2.1 MOSFET bias point

For a N-type Enhancement MOSFET (eMOSFET, NMOS), it consists of three terminals: Gate (G), Drain (D), and Source (S). The operation of the MOSFET is controlled by the voltage V_{GS} applied between the gate and source terminal.

- **Cut-off Region:** When $V_{\text{GS}} < V_{\text{TH}}$, the MOSFET is in the cut-off region, there is no conductive channel formed between the source and drain, and thus no current flows through the device, $I_{\text{DS}} = 0$.
- **Triode Region:** When $V_{\text{GS}} > V_{\text{TH}}$ and $V_{\text{DS}} < V_{\text{GS}} - V_{\text{TH}}$, the MOSFET behaves like a variable resistor, the current I_{DS} is controlled by both V_{GS} and V_{DS}, expressed as:

$$I_{\text{DS}} = k_n \left[2(V_{\text{GS}} - V_{\text{TH}})V_{\text{DS}} - V_{\text{DS}}^2\right] \tag{2.1}$$

  where k is a conduction parameter of the device, could be determined by

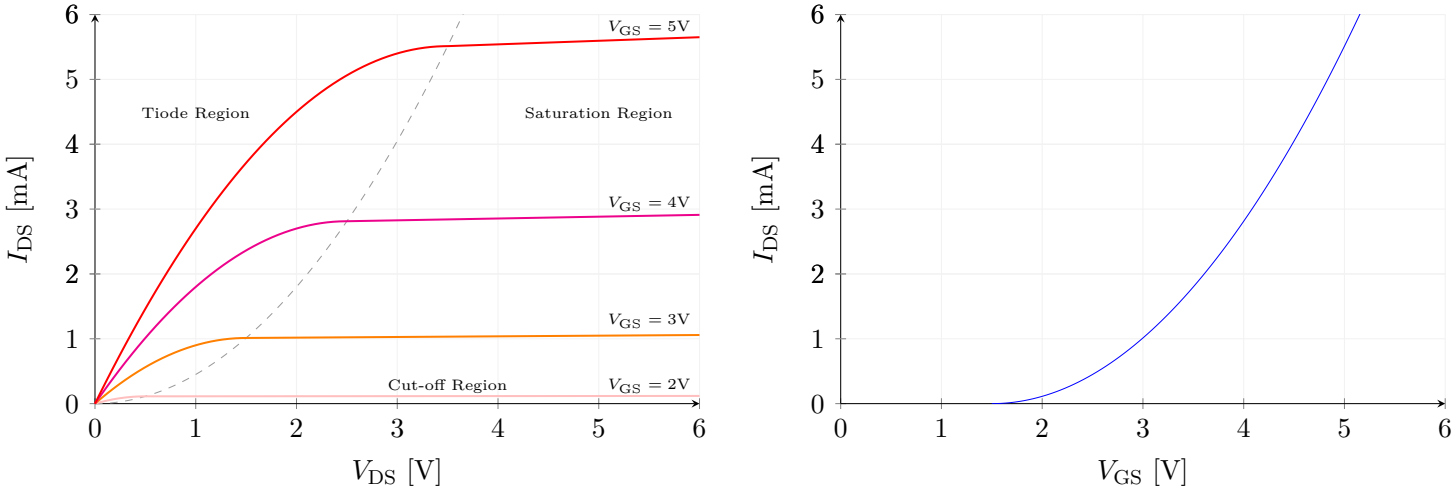
$$k_n = \frac{1}{2}\mu_n C_{\text{ox}} \frac{W}{L}, \tag{2.2}$$

  where μ_n is the mobility of the charge carriers (n-type, electrons), in unit of mA/V^2, C_{ox} is the oxide capacitance per unit area, W/L is the "**aspect ratio**" of the device given by the ratio of width / length of the conductive channel.

- **Saturation Region:** (or active region) When $V_{\text{GS}} > V_{\text{TH}}$ and $V_{\text{DS}} \geqslant V_{\text{GS}} - V_{\text{TH}}$, the MOSFET behaves like a constant current source, and the current flowing through it is primarily controlled by V_{GS}. The current I_{DS} is given by:

$$I_{\text{DS}} = k_n \left(V_{\text{GS}} - V_{\text{TH}}\right)^2 (1 + \lambda V_{\text{DS}}) \tag{2.3}$$

  where λ is the channel-length modulation parameter, which is inversely proportional to the length of the channel L, and it has the unit of V^{-1}. This is called **Early effect**.



(a) $I_{\text{DS}} - V_{\text{DS}}$ curve ($V_{\text{TH}} = 1.5\text{V}$, $k_n = 0.45, \lambda = 0.01$) (b) $I_{\text{DS}} - V_{\text{GS}}$ curve in saturation region ($V_{\text{TH}} = 1.5\text{V}$)

Fig. 8: N-type eMOSFET Characteristics Curves

For a Depletion-type MOSFET (dMOSFET), it can be normally-on (unlike eMOSFET is normally-off), depending on the doping concentration and the applied gate voltage. The I-V curve of a dMOSFET is similar to that of an eMOSFET, but when V_{GS} is negative, it can still conduct current in the saturation region. That's because the threshold voltage $V_{TH} < 0$.

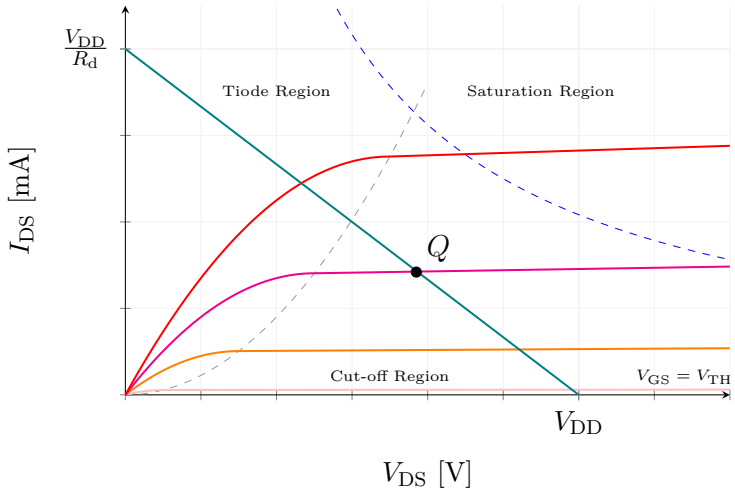


Fig. 9: Graphical solution for bias point Q of MOSFET

The **quiescent point**, or **bias point, DC operating point** is the DC operating conditions (current and voltage) of a device. It defines the steady-state voltages and currents when no AC signal is applied. In order to determine the bias point Q, we could do it simply by **graphical analysis** to find out the intersection point of the MOSFET's I-V curve and the DC load line. After this step, we do AC analysis like in subsection 2.2.

One very important application of the eMOSFET is to act as a **signal amplifier**. In this case, the MOSFET is biased in the **saturation region** to ensure that it can amplify the input signal without being cut off or entering the triode region. But make sure the bias point Q is below the maximum power dissipation curve to avoid thermal overload or device failure (blue hyperbola in fig. 9). The saturation region is the optimal operating zone for MOSFET-based signal amplification due to the following key reasons:

1. High **transconductance** g_m for strong amplification, which is defined as

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = \frac{2k_n(V_{GS} - V_{TH})}{1 + \lambda V_{DS}} \quad \text{(for saturation region)} \tag{2.4}$$

   It actually corresponds to the slope of the I_{DS} vs. V_{GS} curve in fig. 8b. A higher g_m indicates a stronger ability to generate a amplified current.

2. Avoids **non-linear distortion** that can occur in the both triode and cut-off regions. In the triode region, such distortion is a result of nonlinearity of the I-V curve, know as **saturation distortion**, as shown in fig. 10; while in the cut-off region, voltage smaller than threshold voltage is cut off, which is referred to as **cut-off distortion**, as shown in fig. 11. Therefore, working in the saturation region helps to maintain the linearity of waveform during amplification.

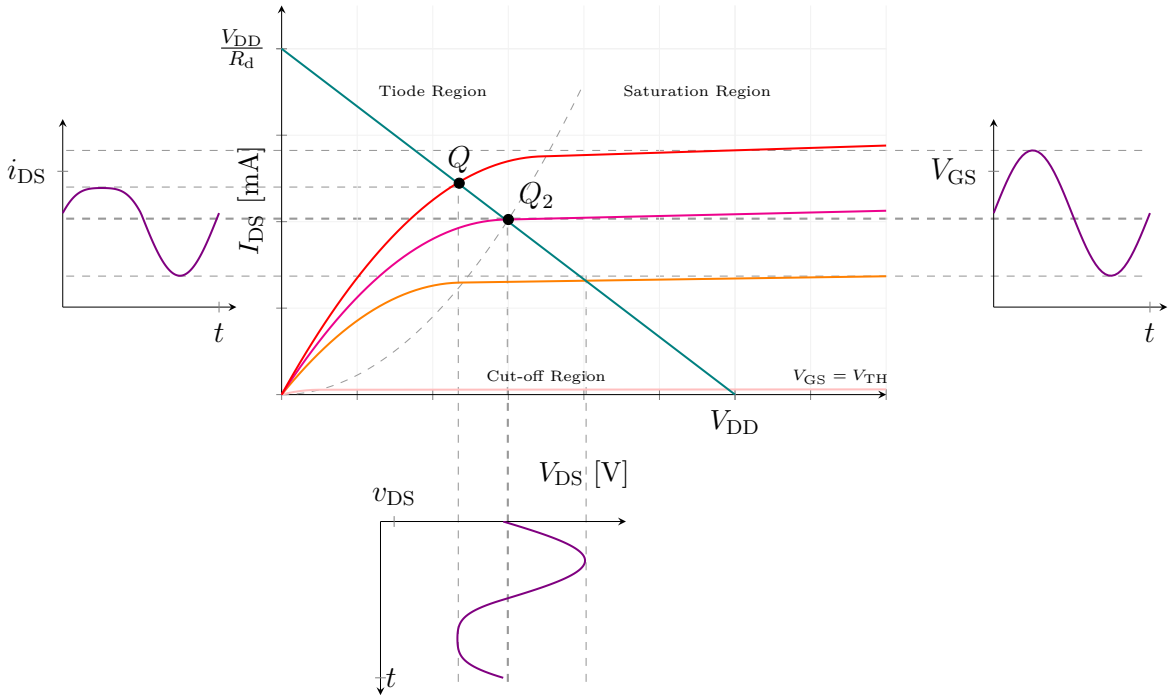


Fig. 10: Graphical analysis: Saturation Distortion

The V_{GS} is controlled by an AC input, varying around Q_2. The output I_{DS} is streched nonlinearly when $V_{DS} < V_{GS} - V_{TH}$, resulting in a distortion in V_{DS}, I_{DS} curve.

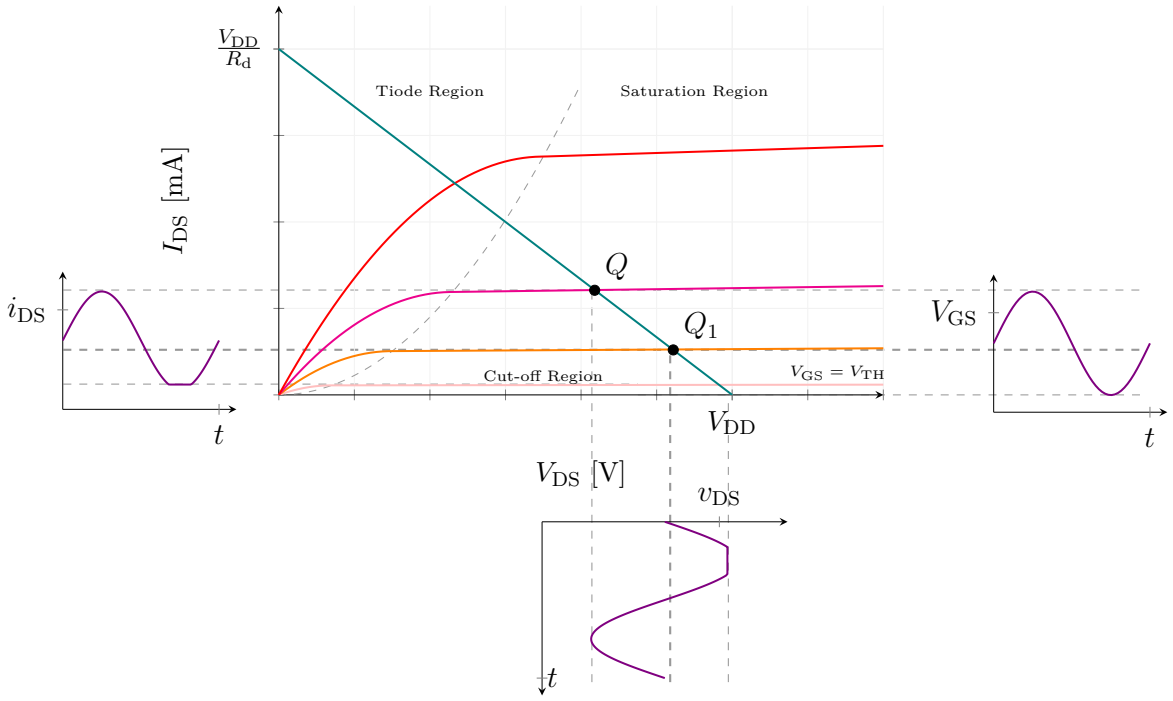


Fig. 11: Graphical analysis: Cut-off Distortion

The V_{GS} is controlled by an AC input, varying around Q_1. The output I_{DS} is cut off when V_{GS} is below the threshold voltage V_{TH}, resulting in a flat line in V_{DS}, I_{DS} curve.

2.2 Common Source Amplifier

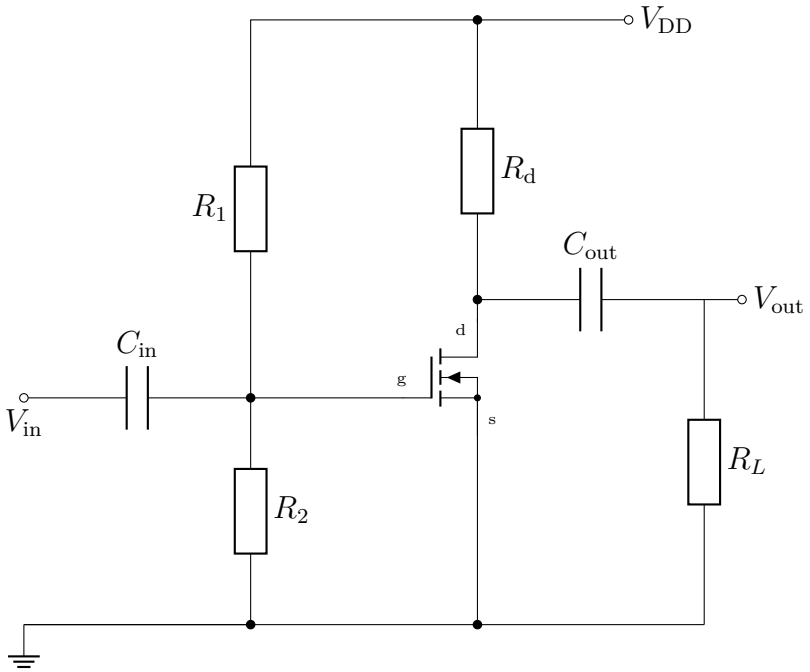


Fig. 12: Common Source Amplifier: Inverting Amplifier

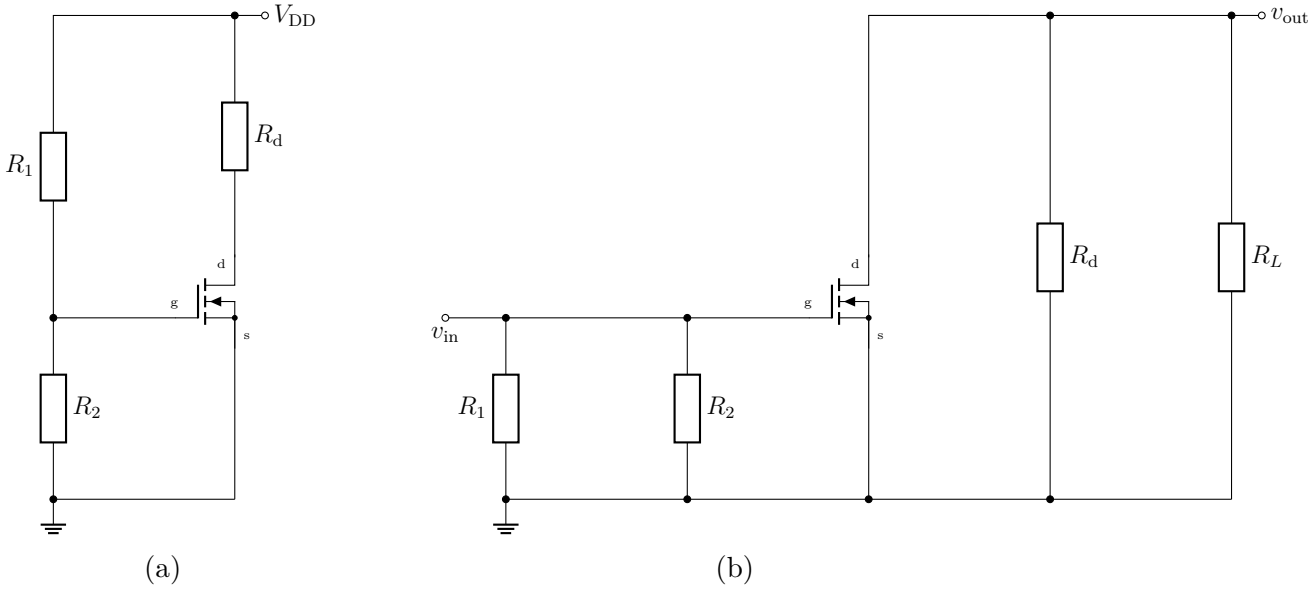


Fig. 13: Common Source Amplifier Equivalent: (a) DC equivalent, (b) AC equivalent.

Let's draw the equivalent of the common source amplifer. For the DC analysis, treat the branches with capacitors as open circuits, since they don't allow DC signals to pass through; For the AC analysis, if the frequency is large enough, We could treat the branches with capacitors as short-circuited. The tricky thing is that there is no AC voltage drop in the DC supply V_{DD}, since an ideal constant volatge source has zero AC impedance, which is referred to as **"virtual ground"**. Therefore, we could view V_{ds} acts on both R_{d} and R_L in parallel.

As an alternative to graphical analysis, we could use **small signal analysis** (must work in saturation region, and small AC signal) to find out the behaviors of the AC equivalent in fig. 13b. The small signal equivalent circuit can be derived by replacing the MOSFET with its small signal model. We could think of a varying tangent in fig. 8b. For an N-channel MOSFET, the small signal model includes the transconductance g_m, the output resistance r_{ds}, and the input capacitance C_{gs}. The equivalent circuit is shown below:

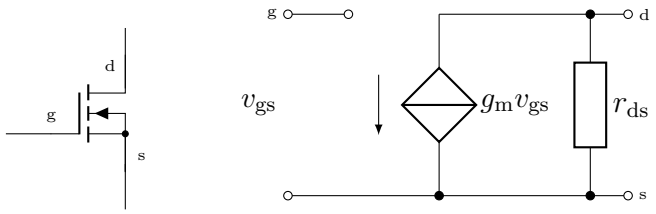


Fig. 14: Small signal equivalent circuit for the AC analysis.

To the right of the circuit, the current i_{ds} going through drain and source is equivalently amplified by:

$$i_{\text{ds}} = g_{\text{m}} v_{\text{gs}} \tag{2.5}$$

Using this equivalent circuit, we could re-draw the AC equivalent circuit for the common source amplifier in fig. 13b.

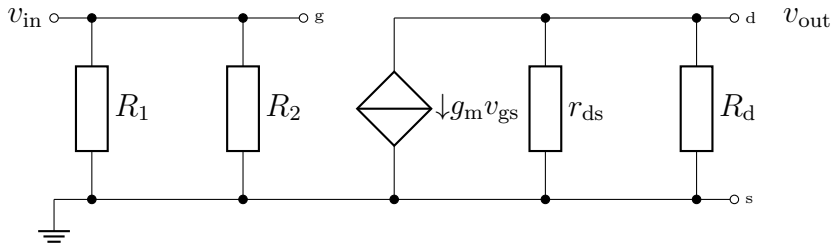


Fig. 15: Small signal equivalent of common source amplifer AC.

Output voltage v_{out} can be derived (neglect r_{ds} if it is much greater than R_{d}, R_L is load resistance which is not included, or think $R_{\text{d}} = R_{\text{d}} \parallel R_L$) as follows:

$$v_{\text{out}} = -i_{\text{ds}} \cdot (r_{\text{ds}} \parallel R_{\text{d}}) \tag{2.6}$$

The voltage gain is then given by:

$$A_v = \frac{v_{\text{out}}}{v_{\text{in}}} = \frac{-i_{\text{ds}} \cdot (r_{\text{ds}} \parallel R_{\text{d}})}{V_{\text{gs}}} = -g_{\text{m}} \cdot (r_{\text{ds}} \parallel R_{\text{d}}) \tag{2.7}$$

Input impedance Z_{in} (imagine v_{in} is applied):

$$Z_{\text{in}} = \frac{v_{\text{in}}}{i_{\text{in}}} = R_1 \parallel R_2 = \frac{R_1 R_2}{R_1 + R_2} \tag{2.8}$$

Output impedance Z_{out} (set input voltage $v_{\text{in}} = 0$, imagine v_{out} is applied):

$$Z_{\text{out}} = \frac{v_{\text{out}}}{i_{\text{out}}} = r_{\text{ds}} \parallel R_{\text{d}} \tag{2.9}$$

Let's consider a variant of the common source amplifier, whose source impedance is non-zero by adding a resistor R_s between the source and ground, as shown in fig. 27. We could use the same small signal analysis to derive the voltage gain, input and output impedance.

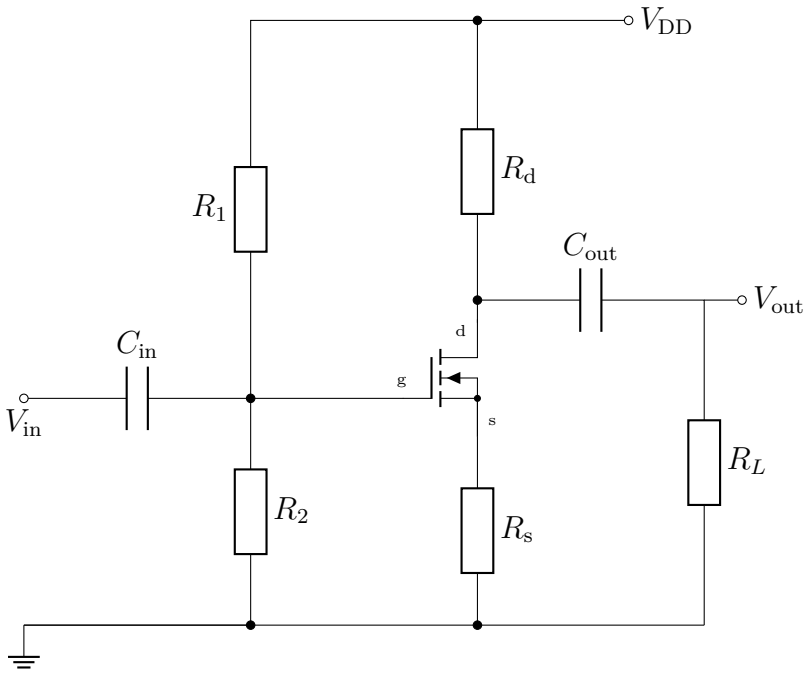


Fig. 16: Common Source Amplifier: Inverting Amplifier (with source impedance)

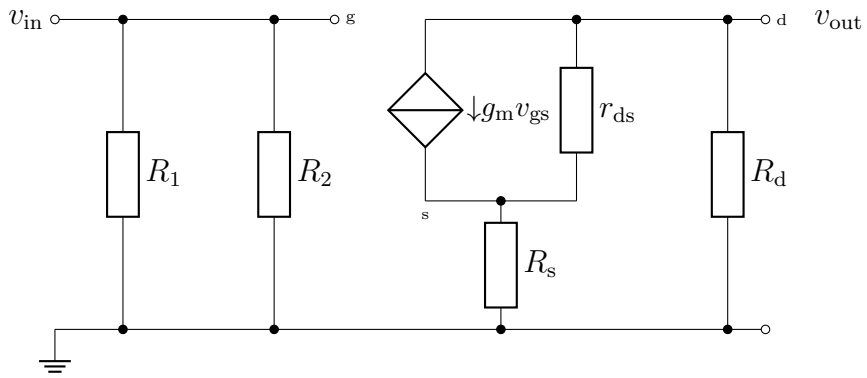


Fig. 17: Small signal equivalent of common source amplifer AC. (with source impedance)

The input voltage becomes:

$$v_{in} = v_{gs} + v_s = v_{gs} + g_m v_{gs} \cdot R_s \tag{2.10}$$

and the output voltage is modified as:

$$v_{out} = -i_{ds} \cdot R_d = -g_m v_{gs} \cdot R_d \tag{2.11}$$

Therefore, the voltage gain is modified as (minus sign indicates phase inversion):

$$A_v = \frac{v_{\text{out}}}{v_{\text{in}}} = \frac{-g_{\text{m}} v_{\text{gs}} \cdot R_{\text{d}}}{v_{\text{gs}} + g_{\text{m}} v_{\text{gs}} \cdot R_{\text{s}}} = -\frac{g_{\text{m}} \cdot R_{\text{d}}}{1 + g_{\text{m}} \cdot R_{\text{s}}} \tag{2.12}$$

Apply the relationship that transconductance g_{m} satisfies from eq. (2.4), $g_{\text{m}} \approx 2k_n(V_{\text{gs}} - V_{\text{TH}})$, we could express the voltage gain as a function of v_{gs}:

$$A_v = -\frac{2k_n(V_{\text{gs}} - V_{\text{TH}}) \cdot R_{\text{d}}}{1 + 2k_n(V_{\text{gs}} - V_{\text{TH}}) \cdot R_{\text{s}}} \tag{2.13}$$

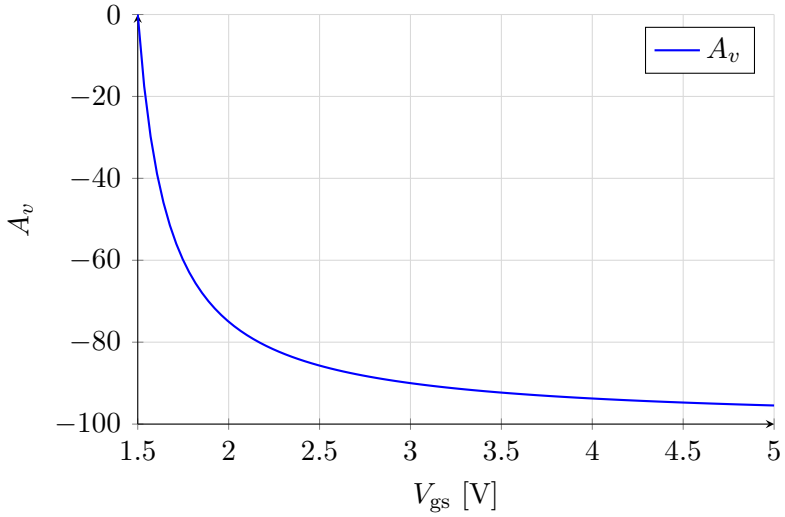


Fig. 18: Voltage gain A_v as a function of V_{gs} for $k_n = 0.3$, $V_{\text{TH}} = 1.5\text{V}$, $R_{\text{d}} = 1\text{k}\Omega$, $R_s = 10\Omega$.

Since the source resistor R_{s} does not affect the input impedance:

$$Z_{\text{in}} = \frac{v_{\text{in}}}{i_{\text{in}}} = R_1 \parallel R_2 = \frac{R_1 R_2}{R_1 + R_2} \tag{2.14}$$

But the output impedance is affected by the source resistor R_{s}. It takes more steps to derive the that. Imagine we apply a voltage v_{out} to the output node, without input voltage $v_{\text{in}} = 0$. Using KCL, the current going through the conductive channel i_{ds} can be expressed by i_{d} as:

$$i_{\text{ds}} = i_{\text{d}} - g_{\text{m}} v_{\text{gs}} \tag{2.15}$$

and KVL in the loop of r_{ds}, R_{s}, and output voltage v_{out} gives us:

$$v_{\text{out}} = i_{\text{ds}} \cdot r_{\text{ds}} + i_{\text{d}} \cdot R_{\text{s}} \tag{2.16}$$

Since the v_{in} is set to zero, this contains another condition:

$$V_{\text{gs}} + i_{\text{d}} \cdot R_{\text{s}} = 0 \tag{2.17}$$

Using the above equations, we can express the impedance to the left of R_{d} as:

$$Z_{\text{d}} = \frac{v_{\text{out}}}{i_{\text{d}}} = \frac{(i_{\text{out}} + g_{\text{m}} \cdot i_{\text{d}} R_{\text{S}}) r_{\text{ds}} + i_{\text{out}} R_{\text{S}}}{i_{\text{out}}} = (1 + g_{\text{m}} R_{\text{s}}) r_{\text{ds}} + R_{\text{s}} \tag{2.18}$$

Thus, the output impedance Z_{out} can be expressed as:

$$Z_{\text{out}} = \frac{v_{\text{out}}}{i_{\text{out}}} = Z_{\text{d}} \parallel R_{\text{d}} = R_{\text{d}} \parallel [(1 + g_{\text{m}} R_{\text{s}}) r_{\text{ds}} + R_{\text{s}}] \tag{2.19}$$

2.3 Common Drain Amplifier

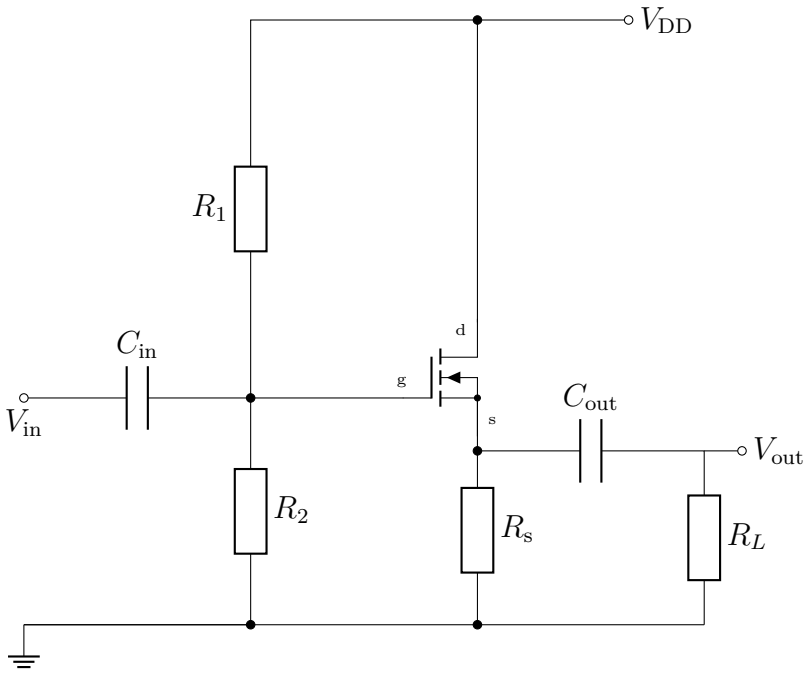


Fig. 19: Common Drain Amplifier: Voltage Source Follower

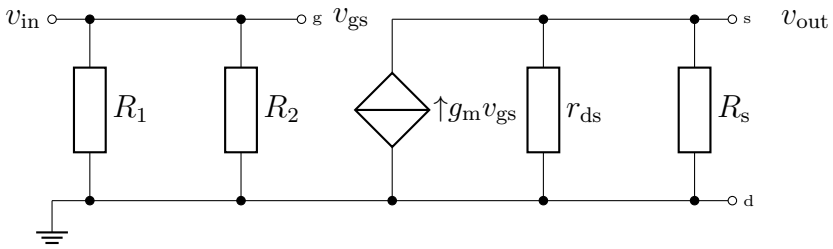


Fig. 20: Small signal equivalent of common drain amplifer AC.

Generally, the load resistance is larger than R_s enough to be neglected. Therefore, the output voltage is approximately equal to

$$v_{out} = g_m v_{gs} \cdot (r_{ds} \parallel R_s) \tag{2.20}$$

The input voltage is the summation of the gate-source voltage v_{gs} and the source-drain voltage $v_{sd} = v_{out}$:

$$v_{in} = v_{gs} + v_{out} = v_{gs} + g_m v_{gs} \cdot (r_{ds} \parallel R_s) \tag{2.21}$$

Thus, we can derive the voltage gain as:

$$A_v = \frac{v_{out}}{v_{in}} = \frac{g_m v_{gs} \cdot (r_{ds} \parallel R_s)}{v_{gs} + g_m v_{gs} \cdot (r_{ds} \parallel R_s)} = \frac{g_m \cdot (r_{ds} \parallel R_s)}{1 + g_m \cdot (r_{ds} \parallel R_s)} \tag{2.22}$$

and divide by g_m in both numerator and denominator, we get:

$$A_v = \frac{(r_\mathrm{ds} \parallel R_\mathrm{s})}{\frac{1}{g_\mathrm{m}} + (r_\mathrm{ds} \parallel R_\mathrm{s})} \tag{2.23}$$

which demonstrates the function of the common drain amplifier as a voltage follower, because the AC gain is close to 1, especially when $1/g_\mathrm{m}$ is small enough compared to $(r_\mathrm{ds} \parallel R_\mathrm{s})$. And the AC part of voltage has the same phase as the input voltage. We may find DC offset on the oscilloscope between the input and output, modulated by the bias point.

The input impedance is still

$$Z_\mathrm{in} = \frac{v_\mathrm{in}}{i_\mathrm{in}} = R_1 \parallel R_2 = \frac{R_1 R_2}{R_1 + R_2} \tag{2.24}$$

But if we add a resistance R_0 between input and before V_gs, the input impedance is then $Z'_\mathrm{in} = R_{12\parallel} = R_1 \parallel R_2 + R_0$. The equivalent input voltage is

$$v'_\mathrm{in} = v_\mathrm{in} \frac{R_{12\parallel}}{R_{12\parallel} + R_0} \tag{2.25}$$

then the gain is modified as:

$$A_{v,s} = \frac{v_\mathrm{out}}{v'_\mathrm{in}} = \frac{g_\mathrm{m} \cdot (r_\mathrm{ds} \parallel R_\mathrm{s})}{\frac{1}{g_\mathrm{m}} + (r_\mathrm{ds} \parallel R_\mathrm{s})} \cdot \frac{R_{12\parallel} + R_0}{R_{12\parallel}} = A_v \cdot \frac{R_{12\parallel} + R_0}{R_{12\parallel}} \tag{2.26}$$

To find the output impedance of the common drain amplifier, we need to modify the small signal model, because $v_\mathrm{in} = 0$ but v_out is applied. So the virtual ground at the source node is no longer valid. We need to consider the current flowing through R_s and r_ds:

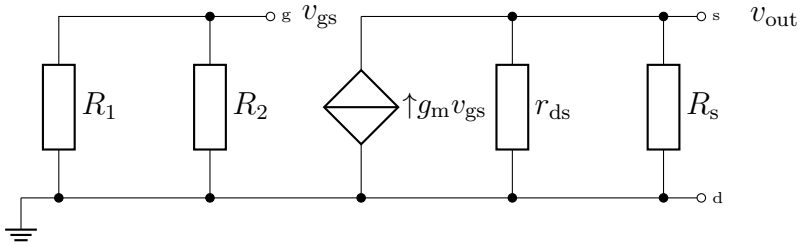


Fig. 21: Small signal equivalent of common drain amplifer AC.

Fig. 22: Small signal equivalent of common drain amplifer AC. (output analysis)

Apply KCL at the source node, we have:

$$i_\mathrm{out} = i_\mathrm{s} + i_\mathrm{ds} - g_\mathrm{m} v_\mathrm{gs} \tag{2.27}$$

the hidden relation is $v_\mathrm{in} = 0$, which means $v_\mathrm{gs} = -v_\mathrm{out}$. So the output impedance is:

$$Z_\mathrm{out} = \frac{v_\mathrm{out}}{i_\mathrm{out}} = \frac{v_\mathrm{out}}{i_\mathrm{s} + i_\mathrm{ds} - g_\mathrm{m} v_\mathrm{gs}} = \frac{1}{\frac{1}{R_\mathrm{s}} + \frac{1}{r_\mathrm{ds}} + g_\mathrm{m}} = R_\mathrm{s} \parallel r_\mathrm{ds} \parallel \frac{1}{g_\mathrm{m}} \tag{2.28}$$

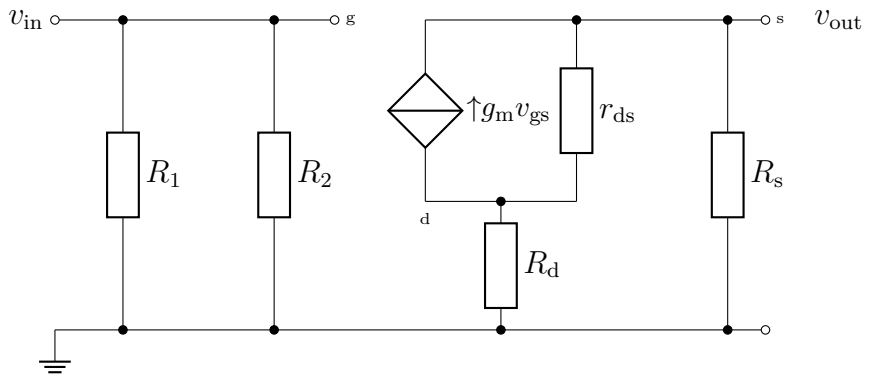


Fig. 23: Small signal equivalent of common drain amplifer AC. (with drain impedance)

Consider the common drain amplifier with a drain resistor R_{d}. Apply KCL at the drain node and KVL in the loop of R_{d}, r_{ds}, and R_{s}:

$$i_{\text{ds}} - i_{\text{out}} = -g_{\text{m}} v_{\text{gs}} \tag{2.29a}$$

$$r_{\text{ds}} i_{\text{ds}} + (R_{\text{d}} + R_{\text{s}}) i_{\text{out}} = 0 \tag{2.29b}$$

That gives us the output current i_{out}:

$$i_{\text{out}} = \frac{g_{\text{m}} v_{\text{gs}} r_{\text{ds}}}{r_{\text{ds}} + R_{\text{s}} + R_{\text{d}}} \tag{2.30}$$

The output voltage becomes:

$$v_{\text{out}} = i_{\text{out}} \cdot R_{\text{s}} = \frac{g_{\text{m}} v_{\text{gs}} r_{\text{ds}}}{r_{\text{ds}} + R_{\text{s}} + R_{\text{d}}} \cdot R_{\text{s}} \tag{2.31}$$

and input voltage is the summation of the gate-source voltage v_{gs} and the output voltage:

$$v_{\text{in}} = v_{\text{gs}} + v_{\text{out}} = v_{\text{gs}} + \frac{g_{\text{m}} v_{\text{gs}} r_{\text{ds}}}{r_{\text{ds}} + R_{\text{s}} + R_{\text{d}}} \cdot R_{\text{s}} \tag{2.32}$$

Thus, we can derive the voltage gain as:

$$A_v = \frac{v_{\text{out}}}{v_{\text{in}}} = \frac{g_{\text{m}} r_{\text{ds}} R_{\text{s}}}{r_{\text{ds}} + R_{\text{s}} + R_{\text{d}} + g_{\text{m}} r_{\text{ds}} R_{\text{s}}} \tag{2.33}$$

And the input impedance is still the same as before:

$$Z_{\text{in}} = \frac{v_{\text{in}}}{i_{\text{in}}} = R_1 \parallel R_2 = \frac{R_1 R_2}{R_1 + R_2} \tag{2.34}$$

To find out the output impedance, set $v_{\text{in}} = 0$ and then $v'_{\text{out}} = -v_{\text{gs}} = i'_{\text{out}} R_{\text{s}}$, and we will have the output impedance:

$$Z_{\text{out}} = R_{\text{s}} \tag{2.35}$$

2.4 Common Gate Amplifier

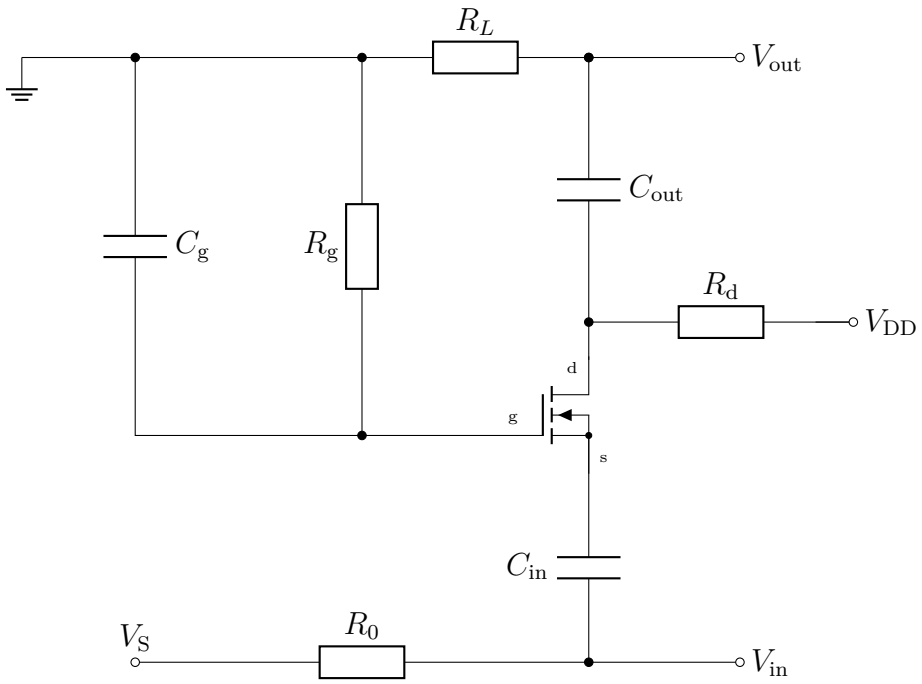


Fig. 24: Common Gate Amplifier: Current Source Follower

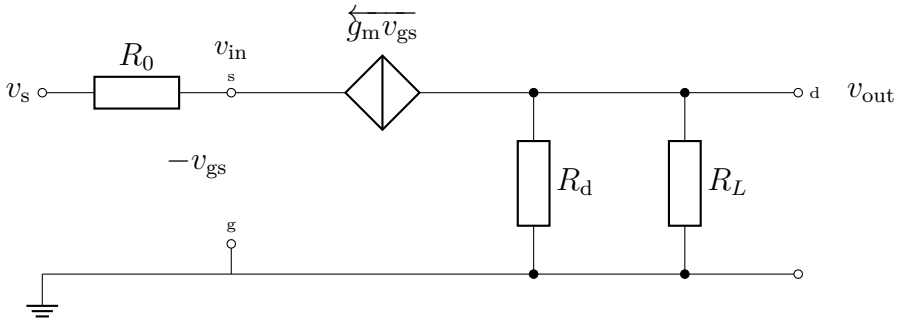


Fig. 25: Small signal equivalent of common drain amplifer AC. (Thévenin equivalent)

The input volatge has following relation with the gate-source voltage (neglect r_{ds} in fig. 25):

$$v_{\text{in}} = -v_{\text{gs}} = v_{\text{s}} + g_{\text{m}} v_{\text{gs}} \cdot R_0 \tag{2.36}$$

The output voltage is given by:

$$v_{\text{out}} = g_{\text{m}} v_{\text{gs}} \cdot (R_{\text{d}} \parallel R_L) \tag{2.37}$$

Therefore, the voltage gain and the voltage gain of source can be derived as:

$$A_v = \frac{v_{\text{out}}}{v_{\text{in}}} = g_{\text{m}} \cdot (R_{\text{d}} \parallel R_L) \tag{2.38a}$$

$$A_{v,s} = \frac{v_{\text{out}}}{v_{\text{s}}} = \frac{g_{\text{m}} \cdot (R_{\text{d}} \parallel R_L)}{g_{\text{m}} R_0 + 1} = \frac{(R_{\text{d}} \parallel R_L)}{\frac{1}{g_{\text{m}}} + R_0} \tag{2.38b}$$

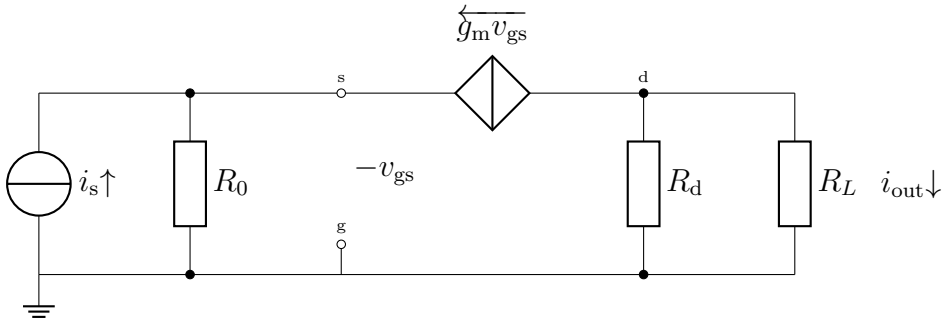


Fig. 26: Small signal equivalent of common drain amplifer AC. (Norton equivalent)

To identify the current gain, we could use the Norton equivalent small signal model in fig. 26. Apply KCL at the source node, we have:

$$i_s + \frac{v_{gs}}{R_0} + g_m v_{gs} = 0 \tag{2.39}$$

and current divider gives us the output current:

$$i_{out} = -\frac{R_d}{R_d + R_L} g_m v_{gs} \tag{2.40}$$

with the input current being:

$$i_{in} = -g_m v_{gs} \tag{2.41}$$

Thus, the current gain and the current gain of source are

$$A_i = \frac{i_{out}}{i_{in}} = \frac{-\frac{R_d}{R_d+R_L} g_m v_{gs}}{-g_m v_{gs}} = \frac{R_d}{R_d + R_L} \tag{2.42a}$$

$$A_{i,s} = \frac{i_{out}}{i_s} = \frac{R_d}{R_d + R_L} \cdot \frac{g_m v_{gs}}{\frac{v_{gs}}{R_0} + g_m v_{gs}} = \frac{R_d}{R_d + R_L} \cdot \frac{g_m R_0}{1 + g_m R_0} \tag{2.42b}$$

when $R_d \gg R_L$, and $g_m R_0 \gg 1$, we can find that the current gain is close to 1, which means common gate amplifier is a current source follower.

The input impedance is given by:

$$Z_{in} = \frac{v_{in}}{i_{in}} = \frac{-v_{gs}}{-g_m v_{gs}} = \frac{1}{g_m} \tag{2.43}$$

where we directly use Thévenin equivalent voltage v_{in} to divide by the Norton equivalent current i_{in}. And similarly, when calculating the output impedance, set the input voltage to zero, and apply KCL at the drain node, and KVL in the left loop of R_d and R_0:

$$i_d = i_{out} - g_m v_{gs} \tag{2.44a}$$

$$i_d = \frac{g_m v_{gs} r_{ds}}{R_d} = \frac{v_{out}}{R_d} \tag{2.44b}$$

Solve this system of equations, we can find the output impedance as:

$$Z_{out} = \frac{v_{out}}{i_{out}} = \frac{g_m v_{gs} r_{ds}}{\frac{g_m v_{gs} r_{ds}}{R_d} + g_m v_{gs}} = \frac{r_{ds} R_d}{R_d + r_{ds}} = R_d \parallel r_{ds} \tag{2.45}$$

2.5 Function of Capacitors

The frequency response of capacitors depends on their capacitance values and frequency.

1. Interal parasitic capacitances: they are capacitances between the terminals of the MOSFET, which are intrisic physical properties of the device. For example, C_{gs}, C_{gd}, C_{sb}, C_{db}, and C_{gd} contributes to the Miller effect **in the high-frequency response**. They are responsible for limited bandwidth, phase shift and instability in amplifiers.

2. Decoupling capacitors: they are used to suppress high frequency AC signals from the power supply voltage, which helps to maintain a steady DC voltage at the power supply. For example, C_{DD} in the common source amplifier, which decouples the AC signals from the power supply V_{DD}. They have critical role **in high-frequency response**.

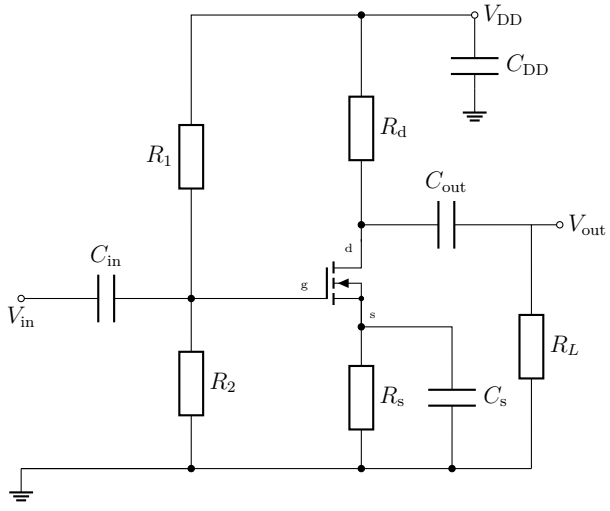
3. Coupling capacitors: act like high-pass filters, pass the AC signals while block DC, **in low-frequency response**. For example, C_{in}, C_{out} in the common source amplifier.

4. Bypass capacitors: they are used to bypass the AC signals around a resistor to ground **in low-pass frequency response**, which helps to maintain a steady DC voltage bias point. For example, C_{s} in the common source amplifier, which bypasses R_{s} for AC signals. Also, it can stablize the AC gain, recall we find the gain of the common source amplifier with R_{s} in eq. (2.12) as:

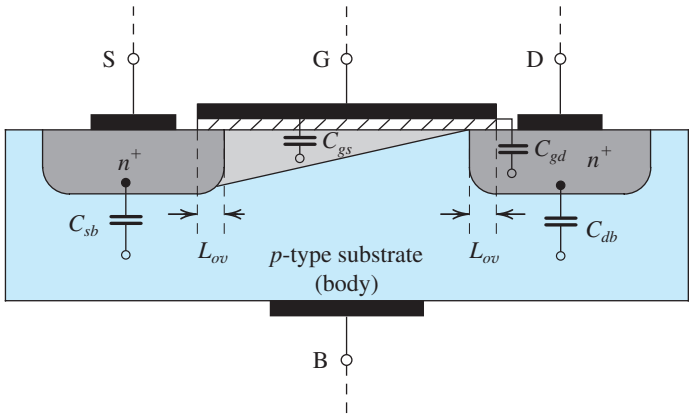
$$A_v = -\frac{g_{\text{m}} R_{\text{d}}}{1 + g_{\text{m}} R_{\text{s}}} \tag{2.46}$$

When C_{s} is present, it bypasses R_{s} for AC signals, and the gain becomes only dependent on R_{d} and g_{m},

$$A_v = -\frac{g_{\text{m}} R_{\text{d}}}{1 + g_{\text{m}} \cdot 0} = -g_{\text{m}} R_{\text{d}} \tag{2.47}$$



(a) Common Source Amplifier Circuit.



(b) Internal parasitic capacitances inside of a n-channel MOSFET (from Sedra & Smith).

Fig. 27: All capacitances in the MOSFET amplifer

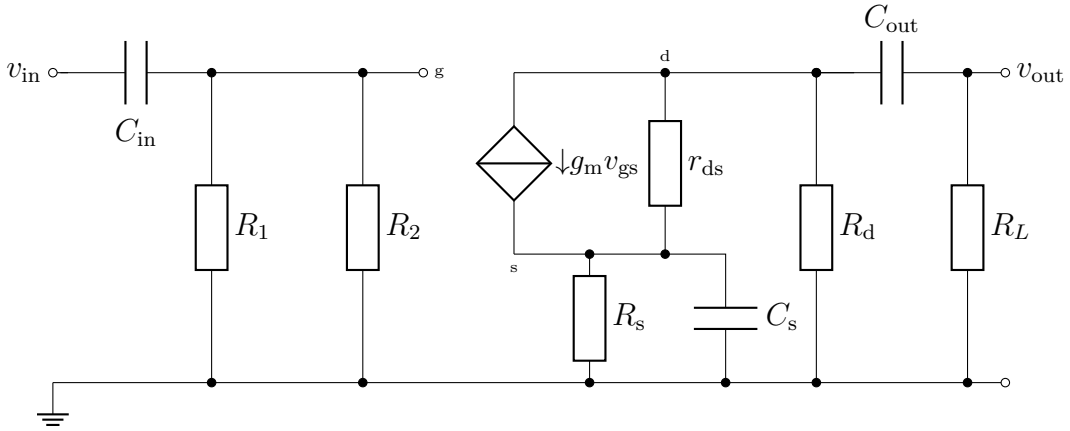


Fig. 28: Small signal equivalent of common source amplifer AC. (low-frequency case)

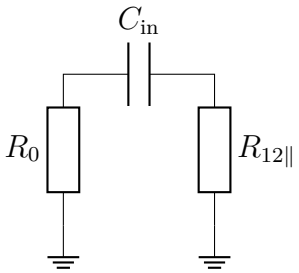
To find out volatge gain of this complicated circuit, we could consider the transfer function with multiple cutoff frequency points (**poles** means the op-amp's gain starts to roll off, while **zeros** are the frequencies where the gain stops rolling off or even rises. We could understand when the denominator/numerator of the transfer function is 0, then poles/zeros.), the general form would be:

$$H(s) = H_{\max} \frac{(1+\omega_{Z_1}/s)\,(1+\omega_{Z_2}/s)\ldots}{(1+\omega_{P_1}/s)\,(1+\omega_{P_2}/s)\ldots} \tag{2.48}$$

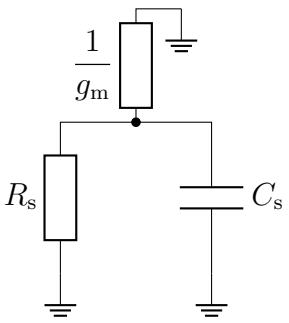
to find the lower cutoff frequency, we need to solve $|H(j\omega_L)|^2 = \frac{1}{2}|H_{\max}|^2$, which gives

$$\omega_L \simeq \sqrt{\sum \omega_{P_i}^2 - 2\sum \omega_{Z_i}^2} \tag{2.49}$$

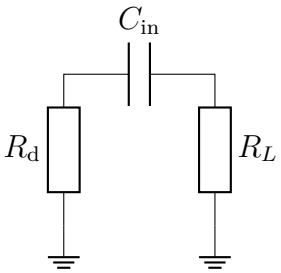
where ω_{P_i} are the poles from the denominator, and ω_{Z_i} are the zeros from the numerator. Let's analyze the small signal equivalent circuit in fig. 28 to find out the poles and zeros. We could treat the circuit as 3 stages:



(a) Stage 1:
$\omega_{P_1} = \dfrac{1}{(R_0 + R_{12\|})C_{\text{in}}}$



(b) Stage 2: $\omega_Z = \dfrac{1}{R_{\text{s}}C_{\text{s}}}$
$\omega_{P_2} = \dfrac{1}{(R_{\text{s}} \parallel \frac{1}{g_{\text{m}}})C_{\text{s}}}$



(c) Stage 3:
$\omega_{P_3} = \dfrac{1}{(R_{\text{d}} + R_L)C_{\text{out}}}$

Fig. 29: Circuits for determining the time constant of each of the three capacitors

Therefore, the normalized transfer function becomes (the voltage gain $A_v' = A_v \cdot H(s)$) :

$$H(s) = \frac{(1+\omega_Z/s)}{(1+\omega_{P_1}/s)\,(1+\omega_{P_2}/s)\,(1+\omega_{P_3}/s)} \tag{2.50}$$

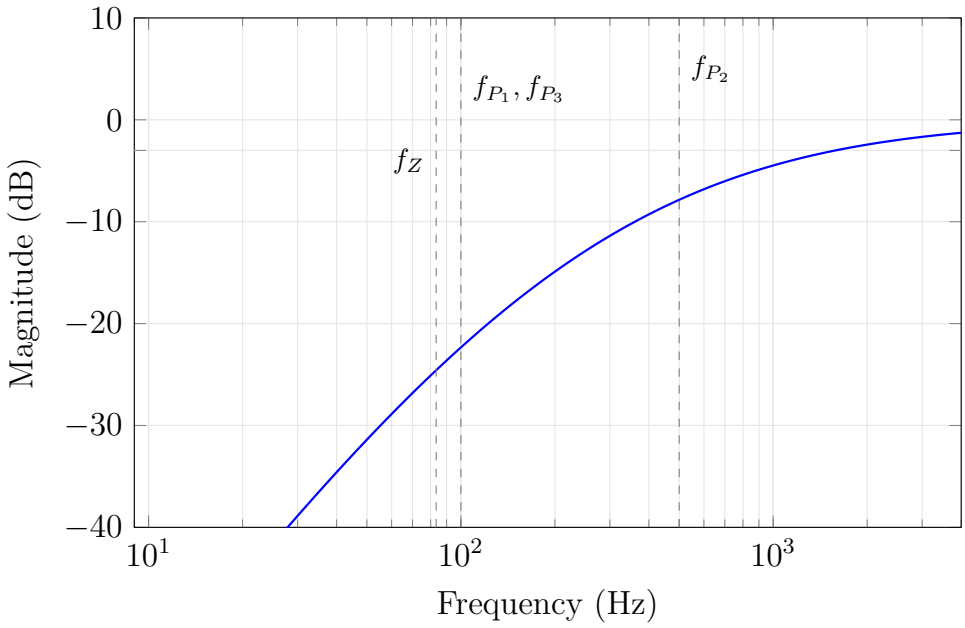


Fig. 30: Bode plot of MOSFET common source amplifer with pole frequencies set very separated to show their distinct effects ($f_Z = 83$ Hz, $f_{P_1} = f_{P_3} = 100$ Hz, $f_{P_2} = 500$ Hz).

Then we could determine the capacitors we should choose in our experiments: if $R_{12\|} = 25\text{k}\Omega$, $R_0 \ll R_{12\|}$, $R_\text{s} = 100\text{k}\Omega$ and $g_\text{m} = 0.05$ A/V, $R_\text{d} = 1\text{k}\Omega$, and $R_L = 0\text{k}\Omega$. If we let $f_{P_1} = f_{P_3} = 100$ Hz, $f_{P_2} = 500$ Hz, we could find the capacitances as:

$$C_\text{s} = \frac{1}{2\pi f_{P_2}(R_\text{s} \parallel \frac{1}{g_\text{m}})} = \frac{1}{2\pi \cdot 500 \cdot (100 \parallel \frac{1}{5\times 10^{-2}})} = 19.1\,\mu\text{F} \tag{2.51a}$$

$$C_\text{in} = \frac{1}{2\pi f_{P_1} R_{12\|}} = \frac{1}{2\pi \cdot 100 \cdot 25 \times 10^3} = 63.6\,\text{nF} \tag{2.51b}$$

$$C_\text{out} = \frac{1}{2\pi f_{P_3}(R_\text{d} + R_L)} = \frac{1}{2\pi \cdot 100 \cdot 1000} = 1.59\,\mu\text{F} \tag{2.51c}$$

and we could find the zero frequency as:

$$f_Z = \frac{1}{2\pi R_\text{s} C_\text{s}} = \frac{1}{2\pi \cdot 100 \cdot 19.1 \times 10^{-6}} = 83.34\,\text{Hz} \tag{2.52}$$

To properly select the bypass capacitor C_s, and coupling capacitors C_in, C_out while setting the lower 3-dB cutoff frequency f_L properly, follow these steps:

- **Choose bypass capacitor C_s first**: The resistance seen by C_s, $\left(\frac{1}{g_m} \parallel R_s\right)$, is typically the smallest among resistance of 3 stages; Set C_s such that its pole frequency $f_{P_2} = f_L$. This minimizes total capacitance while setting the dominant pole.
- **Select coupling capacitors C_out and C_in**: Place their pole frequencies **5∼10 times lower** than f_{P_2}. Ensures they don't significantly affect f_L. Avoid excessively low pole frequencies, which would require impractically large capacitors.

After done with low-frequency analysis, let's analyze the range of the high-frequency response of the common source amplifier. Now the internal parasitic capacitors and decoupling capacitors will step in to play a significant role. The small signal equivalent circuit could be drawn below:

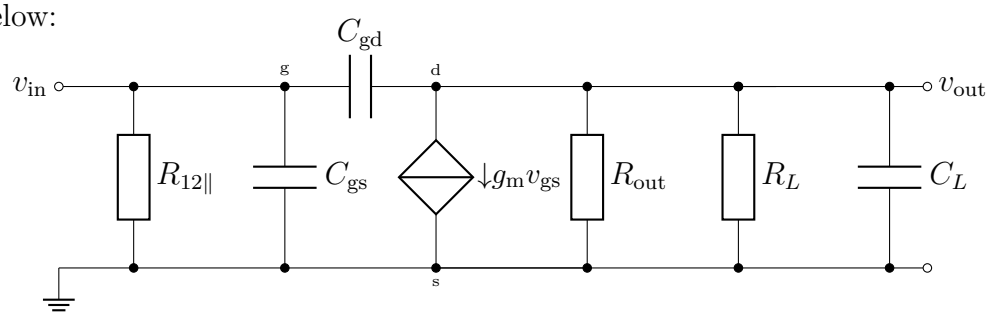


Fig. 31: Small signal equivalent of common source amplifer AC. (high-frequency case)

Using KCL at the drain node,

$$i_{\text{gd}} = g_{\text{m}} v_{\text{gs}} + \frac{v_{\text{out}}}{R_{\text{out}} \parallel R_L} + sC_L v_{\text{out}} \tag{2.53}$$

The volatge gain is given by, with $R'_{\text{out}} = R_{\text{out}} \parallel R_L$ to simplify the notation,

$$A_v = -g_{\text{m}} R'_{\text{out}} \cdot \frac{1 - s\left(C_{\text{gd}}/g_{\text{m}}\right)}{1 + s\left(C_L + C_{\text{gd}}\right) R'_{\text{out}}} \tag{2.54}$$

By setting the numerator to zero, we can find zero frequency

$$f_Z = \frac{1}{2\pi\left(C_{\text{gd}}/g_{\text{m}}\right)} = \frac{g_{\text{m}}}{2\pi C_{\text{gd}}} \tag{2.55}$$

and the high-frequency cutoff frequency can be found by setting the denominator to zero, which gives us the pole frequency, or upper cutoff frequency, f_H:

$$f_H = \frac{1}{2\pi\left(C_L + C_{\text{gd}}\right) R'_{\text{out}}} \tag{2.56}$$

where we can find that, the transmission zero frequency is much higher than the upper cutoff frequency, their ratio can be expressed as:

$$\frac{f_Z}{f_H} = g_{\text{m}} R'_{\text{out}} \cdot \left(1 + \frac{C_L}{C_{\text{gd}}}\right) \tag{2.57}$$

Thus, f_Z does not play a significant role in the vicinity of f_H. Also, the gain decreases from its low-frequency value of $(g_m R'_{\text{out}})$ at a uniform rate of -6 dB /octave (-20 dB/ decade), reaching unity (0 dB) at a frequency $f_{\max}$, which is equal to the **gain-bandwidth product**,

$$f_{\max} = |A_{\max}|\, f_H = g_m R'_{\text{out}} \frac{1}{2\pi\left(C_L + C_{\text{gd}}\right) R'_{\text{out}}} \tag{2.58}$$

The gain–bandwidth product simplifies amplifier design by quantifying the gain vs. speed trade-off. It's critical for: selecting op-amps, predicting bandwidth and avoiding instability. For complicated high frequency transfer functions, similar to eq. (2.49), we can also find f_H:

$$\frac{1}{\omega_H} \simeq \sqrt{\sum \frac{1}{\omega_{P_i}^2} - \sum \frac{2}{\omega_{Z_i}^2}} \tag{2.59}$$

Supplementary Materials

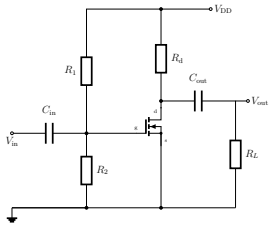
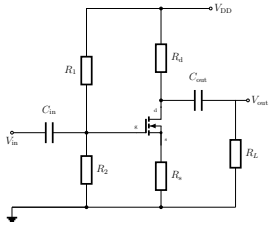
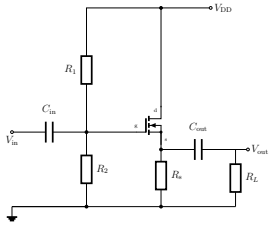
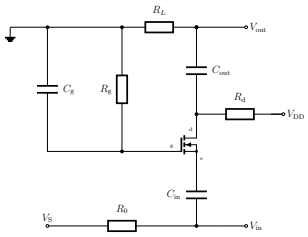
Circuit Type	Voltage Gain A_v	Z_{in}	Z_{out}	Feature
Common Source 	$-g_{\text{m}}\left(R_{\text{d}} \parallel r_{\text{ds}}\right)$	$R_1 \parallel R_2$	$R_{\text{d}} \parallel r_{\text{ds}}$	Inverting amplifier
Common Source with R_{s} 	$-\dfrac{g_{\text{m}}R_{\text{d}}}{1+g_{\text{m}}R_{\text{s}}}$	$R_1 \parallel R_2$	$R_{\text{d}} \parallel \left[\left(1+g_{\text{m}}R_{\text{s}}\right)r_{\text{ds}}+R_{\text{s}}\right]$	Inverting amplifier
Common Drain 	$\dfrac{g_{\text{m}}\left(R_{\text{s}} \parallel r_{\text{ds}}\right)}{1+g_{\text{m}}\left(R_{\text{s}} \parallel r_{\text{ds}}\right)}$	$R_1 \parallel R_2$	$R_{\text{s}} \Vert r_{\text{ds}} \Vert \dfrac{1}{g_{\text{m}}}$	Voltage Source Follower
Common Gate 	$A_i \approx 1$	$\dfrac{1}{g_{\text{m}}}$	$R_{\text{d}} \parallel r_{\text{ds}}$	Current Source Follower

Table 1: Summary of the three basic MOSFET amplifier configurations.