CHAPTER 2

LITERATURE REVIEW

In this chapter, the basic theories of MOSFET (Metal Oxide Semiconductor Field Effect Transistor) transistors and the characteristics of an ideal operational amplifier will be reviewed. The important relationships of two stages CMOS operational amplifier is reviewed. The conventional method for operational amplifier compensation also discussed.

2.1 MOSFET (Metal Oxide Semiconductor Field Effect Transistor)

The metal-oxide-semiconductor field-effect transistor (MOSFET, MOS-FET, or MOS FET), is by far the most common field-effect transistor in both digital and analog circuits. Figure 2.1 shows the structure of MOSFET.

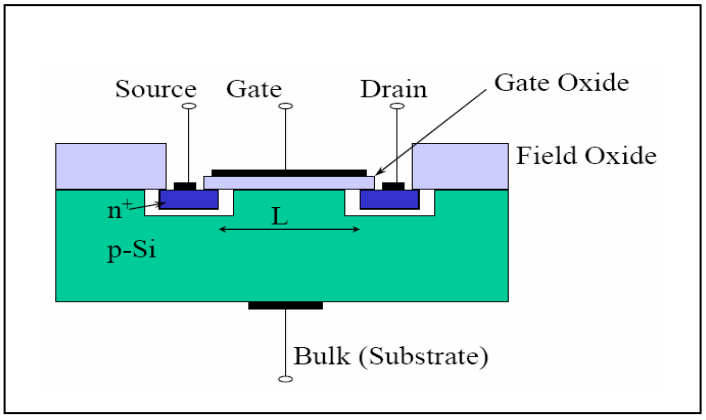


Figure 2.1: MOSFET structure

The MOSFET is composed of a channel of n-type or p-type semiconductor material, and is accordingly called an NMOSFET or a PMOSFET (also commonly nMOSFET, pMOSFET, NMOS FET, PMOS FET, nMOS FET, pMOS FET).

A variety of symbols are used for the MOSFET. The basic design is generally a line for the channel with the source and drain leaving it at right angles and then bending back into the same direction as the channel. Sometimes a broken line is used for enhancement mode and a solid one for depletion mode, but the awkwardness of drawing broken lines means this distinction is often ignored. Another line is drawn parallel to the channel for the gate. Figure 2.2 shows the comparison of enhancement-mode and depletion-mode MOSFET symbol.

The bulk connection, if shown, is shown connected to the back of the channel with an arrow indicating PMOS or NMOS. Arrows always point from P to N, so an NMOS (N-channel) has the arrow pointing in. If the bulk is connected to the source (as is generally the case with discrete devices) it is angled to meet up with the source leaving the transistor. If the bulk is not shown (as is often the case in IC design as they are generally common bulk) an inversion symbol is sometimes used to indicate PMOS.

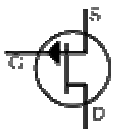
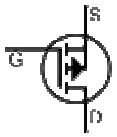
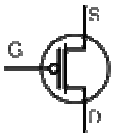
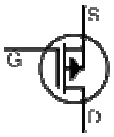
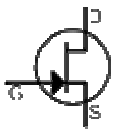
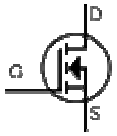
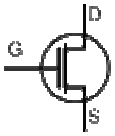
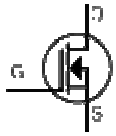
<table>
<tr><td>  </td><td>  </td><td>  </td><td>  </td><td>P-channel</td></tr>
<tr><td>  </td><td>  </td><td>  </td><td>  </td><td>N-channel</td></tr>
<tr><td>JFET</td><td colspan="2">MOSFET enhancement</td><td>MOSFET depletion</td><td></td></tr>
</table>

Figure 2.2: Comparison of enhancement-mode and depletion-mode MOSFET symbol.

2.2 Modes of operation

The operation of a MOSFET can be separated into three different modes, depending on the voltages at the terminals. For an enhancement mode, n-channel MOSFET the modes are:

(1) Cut-off or sub-threshold mode:

When $V_{GS} < V_{th}$ where V_{th} is the threshold voltage of the device and V_{GS} is gate to source voltage. According to the threshold model, the transistor is turned off, and there is no conduction between drain and source. While the current between drain and source should ideally be zero when the transistor is being used as a turned-off switch, there is a weak-inversion current, sometimes called subthreshold leakage.

(2) Triode or linear region:

When $V_{GS} > V_{th}$ and $V_{DS} < V_{GS} - V_{th}$. The transistor is turned on, and a channel has been created which allows current to flow between the drain and source. Figure 2.3 shows cross section of MOSFET operating in linear region The MOSFET operates like a resistor, controlled by the gate voltage.

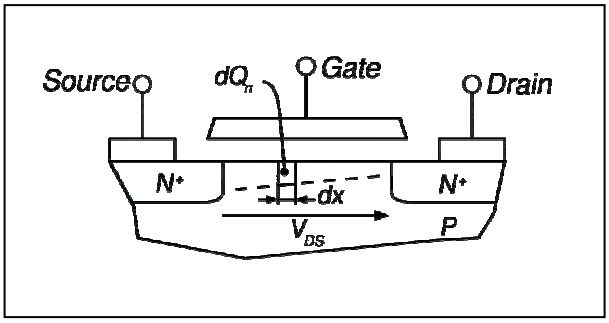


Figure 2.3: Cross section of MOSFET operating in linear region

The current from drain to source is:

$$I_D = \mu_n C_{OX} \frac{W}{L}\left((V_{GS} - V_{th})V_{DS} - \frac{V_{DS}^2}{2}\right) \tag{1.0}$$

where μ_n is the charge-carrier mobility, W is the gate width, L is the gate length and Cox is the gate oxide capacitance per unit area, and V_{DS} is drain to source voltage.

(3) Saturation:

When VGS > Vth and VDS > VGS − Vth . The switch is turned on, and a channel has been created which allows current to flow between the drain and source as shown in Figure 2.4. Since the drain voltage is higher than the gate voltage, a portion of the channel is turned off. The onset of this region is also known as pinch-off.

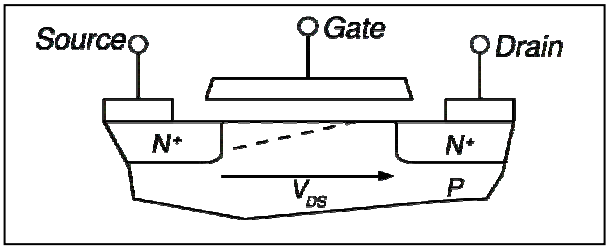


Figure 2.4: Cross section of MOSFET is saturation region

The drain current is now relatively independent of the drain voltage (in a first-order approximation) and the current is only controlled by the gate voltage such that:

$$I_D = \frac{\mu_n C_{OX}}{2}\frac{W}{L}(V_{GS} - V_{th})^2 \tag{1.1}$$

this equation can be multiplied by (1 + λV_{DS}) to take into account the channel length modulation (Early effect), where λ is the channel length.

(4) CMOS (Complementary Metal Oxide Semiconductor)

Complementary Metal Oxide Semiconductor uses complementary and symmetrical pairs of p-type and n-type MOSFETs for logic functions. It is extremely useful for digital circuitry design (act as a simple switches w/o having pull-up resistors). The FETs have their drains and sources connected in parallel, the body of the P-MOS is connected to the high potential (V_{DD}) and the body of the N-MOS is connected to the low potential (Gnd). To turn the switch on the gate of the P-MOS is driven to the low potential and the gate of the N-MOS is driven to the high potential. The CMOS symbols for n-mos and p-mos is shown in Figure 2.5. Main advantage of CMOS is much smaller power dissipation.

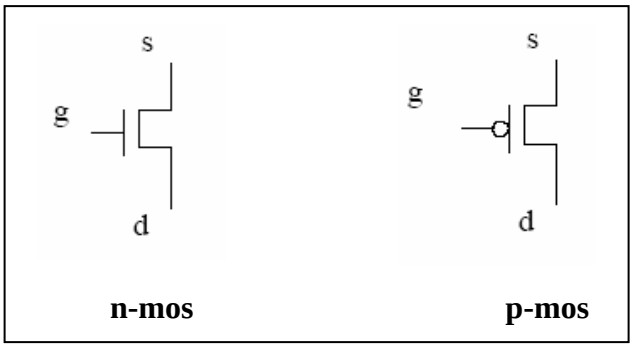


Figure 2.5: CMOS symbol for n-mos and p-mos

As a conclusion, MOSFET transistor operates in three modes. In cutoff region, no current flows between drain and source, transistor is off. In linear region, channel has been created and current flows between drain and source, transistor is turned on. In saturation region, channel has been created and current flows between drain and source but small portion of channel is turned off, which is known as pinch off. In summary, the gate of an MOS transistor controls the flow of current between the source and drain. Simplifying this to extreme allows the MOS transistors to be viewed as simple on/off switches. When the gate of an nMOS transistor is '1,' the transistor is ON and there is a conducting path from source to drain. When gate is low, the nMOS transistor is OFF and almost zero current flows from source to drain. A pMOS transistor is just the opposite, being ON when gate is low and OFF when the gate is high.

2.3 CMOS Operational amplifier

Figure 2.6 shows a block diagram that represents the important aspects of an op amp. CMOS op amps are very similar in architecture to their bipolar counterparts. The differential-transconductance stage forms the input of the op amp and sometimes provides the differential to single-ended conversion. Normally, a good portion of the overall gain is provided by the differential input stage, which improves noise and offset performance.

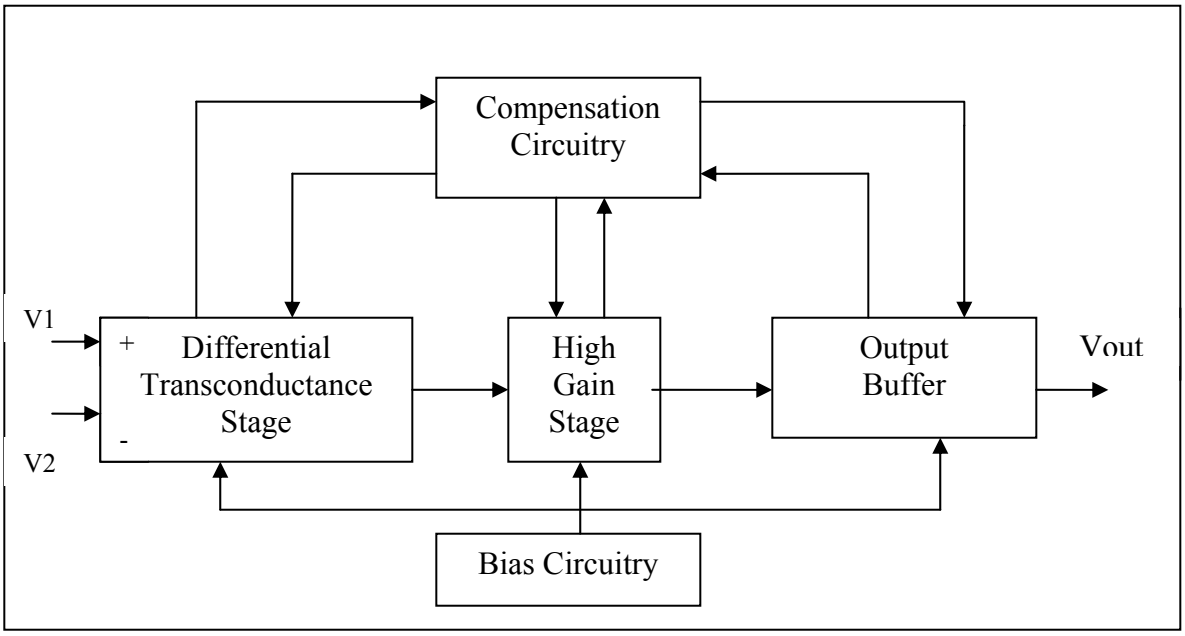


Figure 2.6: Block Diagram of a general two-stage CMOS op amp

The second stage is typically an inverter. If the differential input stage does not perform the differential to single ended conversion, then it is accomplished in the second stage inverter. If the op amp must drive a low resistance load, the second stage must be followed by a buffer stage whose objective is to lower the output resistance and maintain a large signal swing. Bias circuits are provided to establish the proper operating point for each transistor in its quiescent state. Compensation is required to achieve stable closed loop performance.

2.4 Ideal Op Amp

Ideally, an op amp has infinite differential voltage gain, infinite input resistance, and zero output resistance. In reality, an op amp only approaches this values. For most applications where unbuffered CMOS op amps are used, an open loop gain of 2000 or more is usually sufficient. The symbol for an op amp is shown in Figure 2.7.

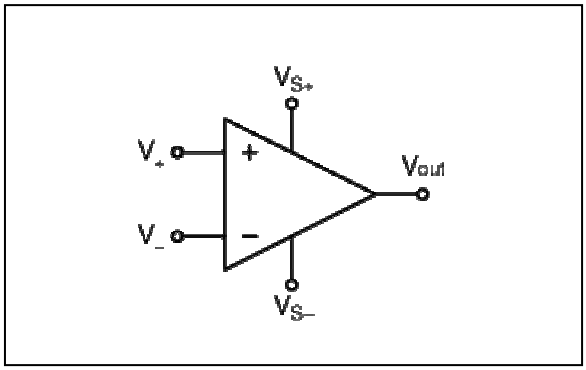


Figure 2.7: Symbol of an op amp

In the nonideal case, the output voltage Vout can be expressed as

$$Vout = Av(v1 - v2) \tag{1.2}$$

Av is used to designate the open loop differential voltage gain. $v1$ and $v2$ are the inputs voltages applied to the noninverting and inverting terminals, respectively. The symbol also shows the power supply connections of V_{DD} and V_{SS}. Generally, these connections are not shown but the designer must remember that they are an integral part of the op amp.

If the gain of the op amp is large enough, the input port of the op amp becomes a null port when negative feedback is applied. A null port (or nullor) is a pair of terminals to a network where the voltage across the terminals is zero and the current flowing into or out of the terminals is zero.

2.5 Modes and parameters

In this section, important operational amplifier modes and several parameters such as open loop gain, input common mode range, output voltage swing, unity gain frequency, slew rate and power dissipation are defined.

Open loop gain:

The open-loop gain of an operational amplifier is the gain obtained when no feedback is used in the circuit. Open loop gain is usually exceedingly high; in fact, an ideal operational amplifier has infinite open-loop gain. Typically an op-amp may have an open-loop gain of around 10^9. Normally, feedback is applied around the op-amp so that the gain of the overall circuit is defined and kept to a figure which is more usable. However the very high gain of the op-amp enables considerable levels of feedback to be applied to achieve required performance. Figure 2.8 shows configuration for the measurement of the open loop gain. The open-loop gain of an operational amplifier falls very rapidly with increasing frequency. Along with slew rate, this is one of the reasons why operational amplifiers have limited bandwidth.

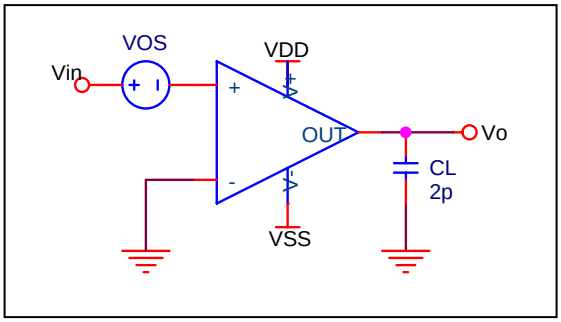


Figure 2.8: Configuration for the measurement of the open loop gain.

Input common mode range:

All operational amplifiers have limitations on the range of voltages over which they will operate. The common mode input voltage range is the range of input voltages which, when applied to both inputs, will not cause clipping or other output distortion. Figure 2.9 shows the configuration for measurement of CMR.

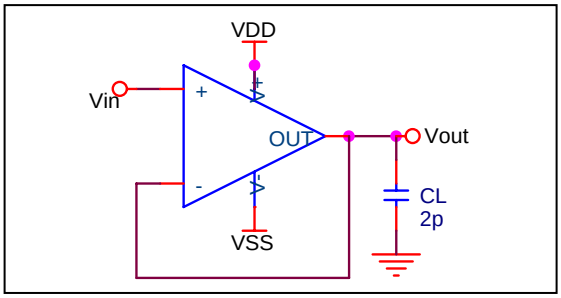


Figure 2.9: The configuration for measurement of CMR.

Output voltage swing:

The output voltage swing is the maximum peak voltage that the output can produce before it starts clipping. This voltage is dependent on the voltage supplied to the op amp - the higher the supply voltage the higher the output voltage swing. It is the swing of the output node without generating a defined amount of harmonic distortion. Figure 2.10 shows the configuration for measurement of output voltage swing.

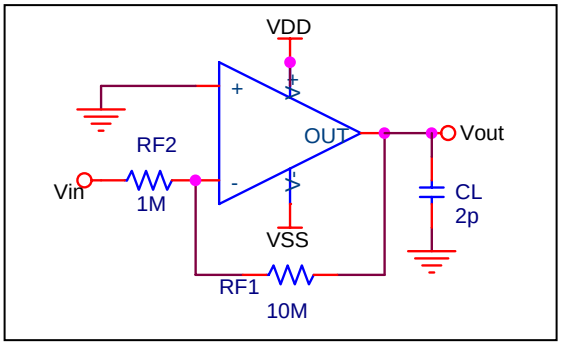


Figure 2.10: The configuration for measurement of output voltage swing

Unity gain frequency:

Figure 2.11 shows the operational amplifier unity gain frequency configuration. Unity gain frequency is the frequency where the voltage gain of an op amp is 1. It indicates the highest usable frequency. It is important because it equals the gain bandwidth product. It is a such frequency of operation for a device where the gain of the component drops to unity. It is the frequency where the open-loop gain is zero. It is also the -3 dB bandwidth in unity-gain closed loop conditions.

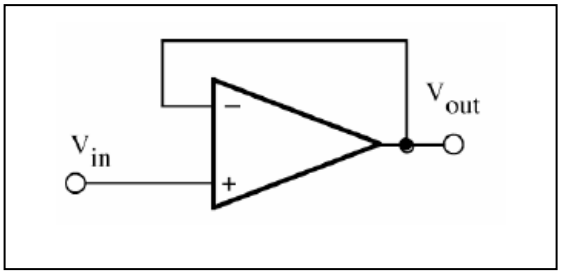


Figure 2.11: Unity gain frequency.

Slew rate:

Slew rate is the maximum slope of the output voltage. Usually it is measured in the buffer configuration as shown in Figure 2.12(a). The slope can be measured from output voltage as shown Figure 2.12(b). The positive slew rate can be different from the negative slew rate. Typically SR =50 ~ 200 V/us, lower values for micropower operation.

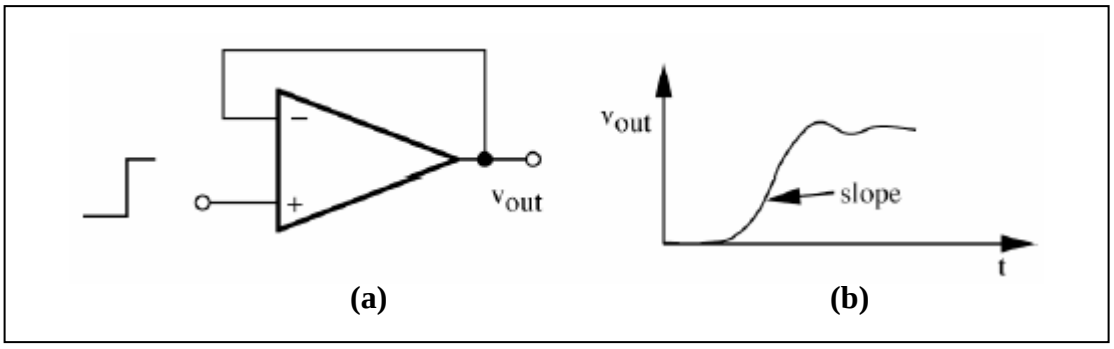


Figure 2.12: Slew rate.

2.6 Two-Stage Operational Amplifier Design

Figure 2.13 shows the basic schematic of two stage CMOS operational amplifier. Basically it is an operational amplifier architecture consisting of two stages. The first part provides of the biasing circuitry for the amplifier. The bias current sets the gate voltage, and this gate voltage is used as a gate bias voltage for the transistor current sources, M5 and M7, which bias the first and second stages of the amplifier.

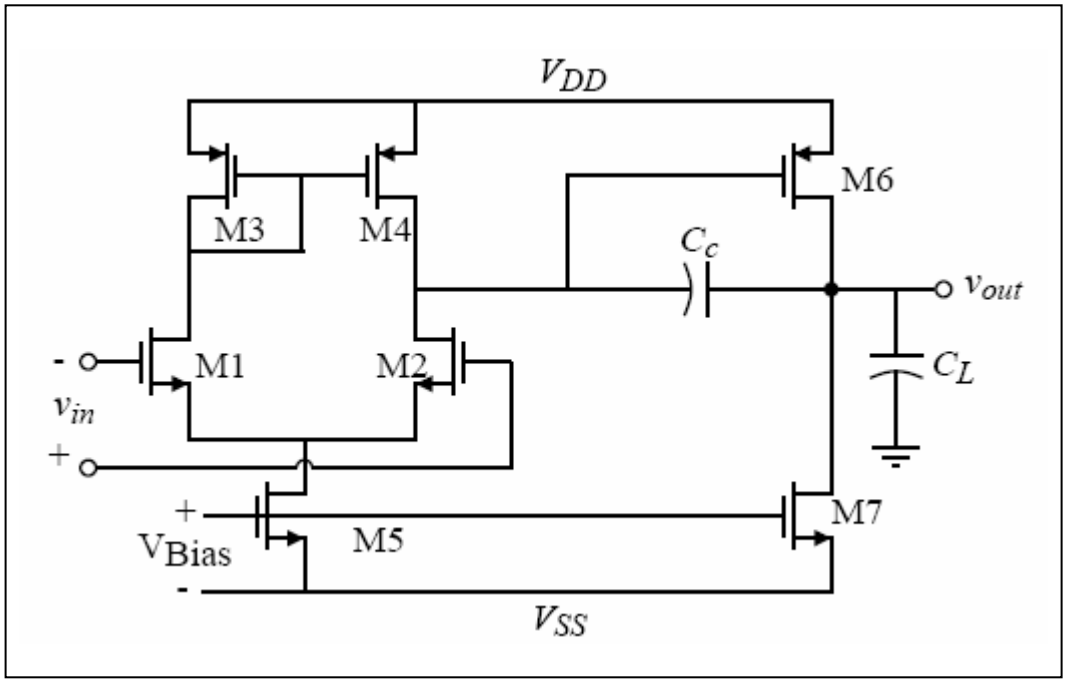


Figure 2.13: Two stage operational amplifier.

The first stage is the differential gain stage consisting transistor M1, M2, M3 and M4. The second gain stage is the common source gain stage consisting transistor M6 and M7. The NMOS transistor M6 is the driver with M7 acting as the load. The important relationship involved in this architecture is discussed next.

Current mirror:

Figure 2.14 shows the current mirror circuit. A current mirror is a circuit designed to copy a current flowing through one active device by controlling the current in another active device of a circuit, keeping the output current constant regardless of loading. Thus 2 NMOS transistors will acts as bias circuit to get same current between I_{ref} and I_{out}. The other NMOS transistor, will acts as a voltage buffer.

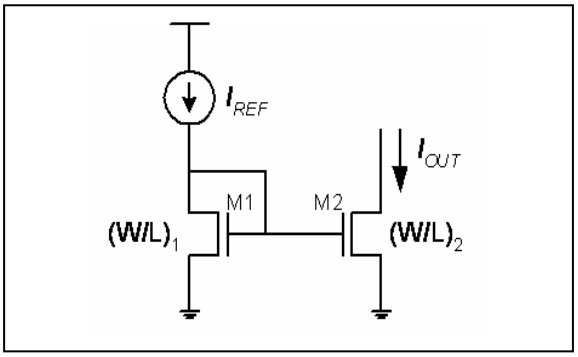


Figure 2.14: Current mirror.

Theoretically,

$I_G = 0$, gate is shorted, therefore

$I_{D1} = I_{ref}$

$I_{D2} = I_{D1}$

$I_{out} = I_{ref}$

By making an assumption,

a) if size of M1 and M2 is same,

So,

$I_{out} = I_{ref}$, it mirror if $(W/L)_{M2} = (W/L)_{M1}$

b) if size of M1 and M2 is different, therefore,

$I_{out} = I_{ref} \ [(W/L)_{M2} / (W/L)_{M1})$

Differential Amplifier:

A differential amplifier is a type of an electronic amplifier that multiplies the difference between two inputs by some constant factor (the differential gain). A differential amplifier is the input stage of operational amplifiers, or op-amps, and emitter coupled logic gates. Given two inputs Vin1 and Vin2, a practical differential amplifier gives an output V_{out}. The general differential amplifier circuit is shown in Figure 2.15 below.

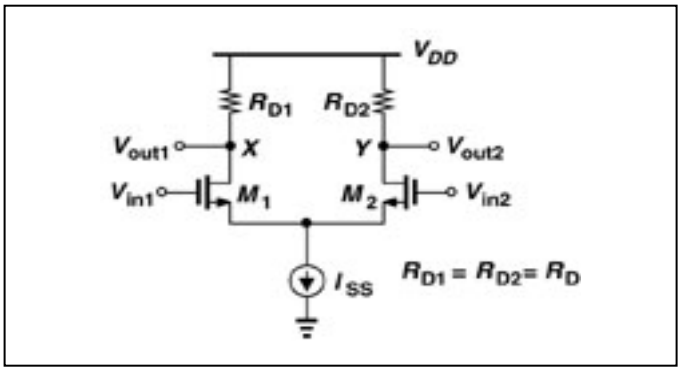


Figure 2.15: Differential amplifier.

The differential amplifier shown in Figure 2.15 has the following relationships.

$I_{D1} = I_{D2} = \frac{I_{SS}}{2}$, when Vin1=Vin2, output CM level $= VDD - RD \times \frac{I_{SS}}{2}$

Two-Stage Operational Amplifier Architecture Relationships

From Figure 2.13 we have, the slew rate is defined as:

Slew rate $SR = \frac{I5}{Cc}$ (1.3)

The differential gain is defined as:

$$\text{First-stage gain } A_{v1} = \frac{gm1}{gds2 + gds4} = \frac{2gm1}{I5(\lambda 2 + \lambda 4)} \tag{1.4}$$

The common source gain is defined as:

$$\text{Second-stage gain } A_{v2} = \frac{gm6}{gds6 + gds7} = \frac{gm6}{I6(\lambda 6 + \lambda 7)} \tag{1.5}$$

Where g_m = transconductance.

g_{ds} = transconductance drain to source.

λ = channel length modulation parameter.

k_p = pmos transconductance parameter (in saturation).

I_D = drain current.

The unity gain bandwidth is given by:

$$\text{Gain-bandwidth } GB = \frac{gm1}{Cc} \tag{1.6}$$

The positive common mode input range is given by:

$$\text{Positive CMR } V_{in(max)} = V_{DD} - \sqrt{\frac{I5}{\beta 3}} - |V_{T03}|(max) + V_{T1}(min)) \tag{1.7}$$

The negative common mode input range is given by:

$$\text{Negative CMR } V_{in(min)} = V_{SS} + \sqrt{\frac{I5}{\beta 1}} + V_{T1}(max) + V_{DS5}(sat) \tag{1.8}$$

Where V_{DSAT5} = drain saturation voltage.

V_{GS} = gate to source voltage.

β = transconductance parameter.

The saturation voltage is given by:

$$VDS(\text{sat}) = \sqrt{\frac{2Ids}{\beta}} \tag{1.9}$$

All transistors are in saturation for the above relationships.

2.7 Compensation

In virtually all op amp applications, feedback will be applied around the amplifier. Figure 2.16 shows the negative feedback system. Therefore, stable performance requires that the amplifier be compensated. Essentially we desire that the loop gain be less than unity when the phase shift around the loop is greater than 135°.

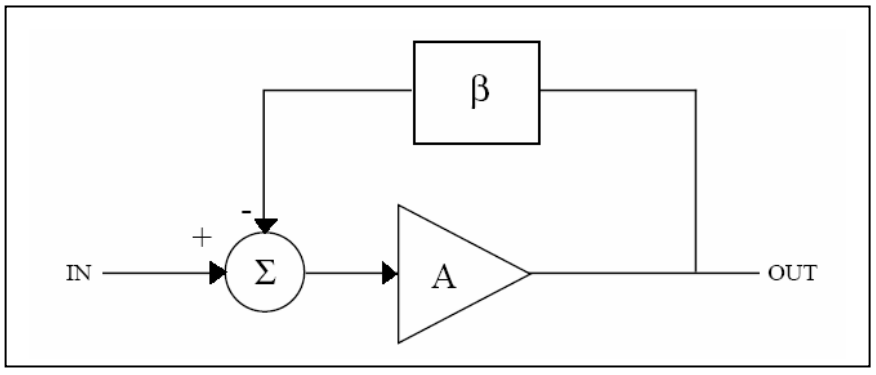


Figure 2.16: Single-loop negative feedback system.

$$\frac{OUT}{IN} = \frac{A}{1 + A\beta} \tag{1.10}$$

A small capacitor across the feedback network will compensate for parasitic capacitance at the (inverting) input of the Op-Amp. Normally 3pF to 10pF will compensate. This is true for Inverting, Non-Inverting, or Voltage Follower configurations. A feed back pole is created when the feedback around any amplifier is resistive. The parallel resistance and capacitance from the input of the device (usually the inverting input) to Ac ground set the frequency of the pole. In many cases the frequency of the pole is much greater than the 3dB frequency of the closed loop gain there is negligible effect on the stability margin. How ever, if the feed back pole is less than approximately six times the expected 3dB frequency, a 'lead' capacitor should be placed from the output to the input of the Op-Amp. The value of the added capacitor should be such that the RC time constant of this capacitor and resistor it parallels is greater than or equal to the original feed back pole time constant.

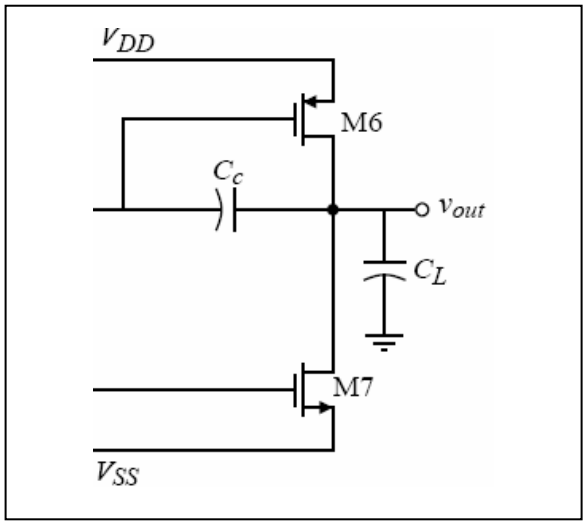


Figure 2.17: Compensation capacitor