Operational Amplifiers Rail to Rail Input Stages Using Complementary Differential Pairs

Submitted to: Dr. Khoman Phang

Submitted by: Ahmed Gharbiya

Date: November 15, 2002

Abstract

A rail to rail input common mode range is an important requirement in operational amplifiers for some applications. The most common method for implementing full range operation is by using a complementary differential pair. It is simply a compound structure that consists of NMOS and PMOS differential pairs connected in parallel. The compound structure achieves rail to rail operation; however, it produces variations in the transconductance over the input common mode range. The variations obstruct the design of an optimal operational amplifier. This paper discusses sources and solutions for the transconductance variations. Several reported approaches to solve the problem will be evaluated. They reduce the variations to 5-28% depending on the technique and method of implementation used. This is a significant improvement over 100% variations we get from complementary differential pair without utilizing any constant g_m technique.

Table of Contents

List of Figures

1. Introduction

Operational amplifiers are the back bone for many analog circuit designs. It is used in numerous applications such as amplifiers and filters. The operational amplifier can be used in two basic configurations: inverting and non-inverting. These configurations place different requirements on the common-mode input range. The required range varies from almost zero to a full rail to rail.

The differential amplifier is used as the input stage for operational amplifiers. The problem is that it will behave as a differential amplifier only over a limited range of common-mode input. Therefore, to make the operational amplifier versatile, its input stage should work for rail to rail common-mode input range. The most common method to achieve this range is to use a complementary differential amplifier at the input stage. This method uses an n-type and a p-type differential pairs simultaneously. Although the method achieves a rail to rail common-mode input operation, it introduces suboptimal operational amplifiers. This is due to the non-constant transconductance (g_m) of the complementary differential amplifier. However, there are methods that keep g_m variations small over the entire input common mode range.

This paper gives an introduction to the concept of designing rail to rail constant gm operational amplifiers using complementary differential pair input stages. Section two starts with explaining the reasons for the need of a rail to rail operation. Section three introduces the complementary differential pair as a solution to achieve the desired range of operation and identifies its drawbacks. Finally, section four examines possible solutions to the suboptimal design problem introduced by the complementary differential pair and discusses implementation methods reported in academic papers.

2. Rail to Rail Operation

The operational amplifier is a circuit building block that can be used in many applications. Op-amps are used in conjunction with feedback networks to achieve several useful functions. However, there are only two main configurations: the inverting and non-inverting configurations. These configurations are shown in Figure 1. Each of the configurations will be analyzed next to determine its input common mode requirements.

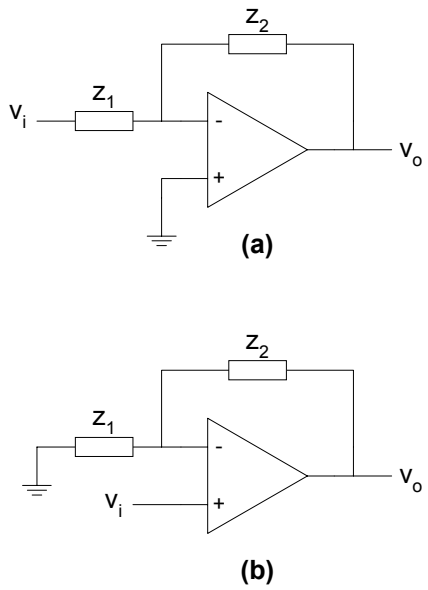

Figure 1: Op Amp in: (a) inverting configuration (b) non-inverting configuration

In the inverting configuration, the op-amp will keep the two input nodes at the same voltage. Since the positive terminal is connected to ground, the non-inverting terminal will have the same voltage. Regardless of the changes of the input voltage vi, both op-amp terminals will be approximately the same and equal ground potential. Therefore, the op-amp needs almost zero common mode input range.

Similarly, in the non-inverting configuration, the op-amp will keep the two input nodes at the same voltage. Since the positive terminal is connected to the input voltage, it will have large variations. The inverting terminal must follow these variations. The worst case allowable variation (v_i swing) is determined by the output voltage and feedback network. Therefore, the common mode input range is given by:

$$V_{CMR} = \frac{Z_1}{Z_1 + Z_2} V_{o,\max} \qquad (1)$$

And if the op-amp output stage in capable of rail to rail operation, $V_{o,max}$ will equal the supply voltage.

A special case of the non-inverting configuration is the voltage follower shown in Figure 2. This configuration will have the most constraint requirement for the common mode input range.

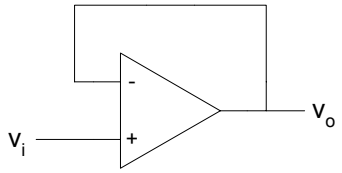

Figure 2: Voltage follower

Similar to the non-inverting configuration, the allowable variation (v_i swing) is determined by the output voltage and feedback network. But, since the feedback is unity, it is only determined by the output voltage swing. If the op-amp output stage in capable of rail to rail operation, then the common mode input range should be rail to rail.

It is clear that to design a versatile op-amp that is useful for any configuration; its input stage should have a rail to rail input common mode range capability. Therefore, we will look into input stages next.

3. The Input Stage

The input stage of every op-amp is a differential amplifier. In CMOS technology the differential amplifier can be realized using a PMOS or NMOS differential pair. There are several tradeoffs that determine which differential pair to use [1]. One criterion that is considered in making the choice is the common mode input range.

To analyze the common mode input range of the NMOS differential input stage, a simplified diagram will be used as shown in Figure 3. Several modifications are made to the simple differential pair in actual implementation such as active loads and cascodes, however this is sufficient for the purpose of illustration. The range extends from the positive supply to $V_{gs,n}+V_{Dsat,b}$ above the negative supply. This minimum voltage is needed to keep the NMOS differential pair and the tail current source in saturation.

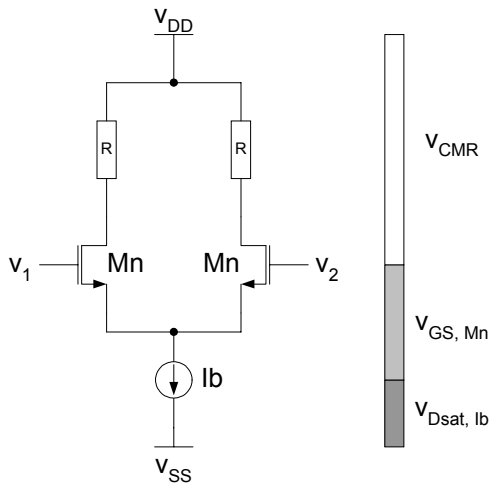

Figure 3: NMOS differential pair common mode input range

A similar analysis can be carried out for the PMOS differential pair shown in Figure 4. The range extends from $V_{gs,n}+V_{Dsat,b}$ below the positive supply to the negative supply. This minimum voltage is needed to keep the PMOS differential pair and the tail current source in saturation.

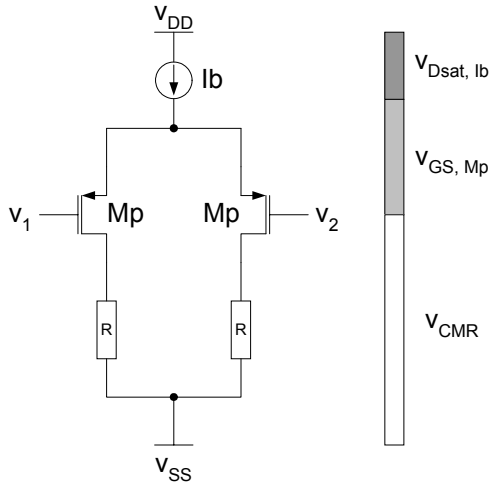

Figure 4: PMOS differential pair common mode input range

The simple differential pair can not meet the rail to rail common mode input requirement. A possible solution to the problem is to use both NMOS and PMOS differential pairs simultaneously. The resulting compound differential pair is called the complementary differential pair and is shown in Figure 5.

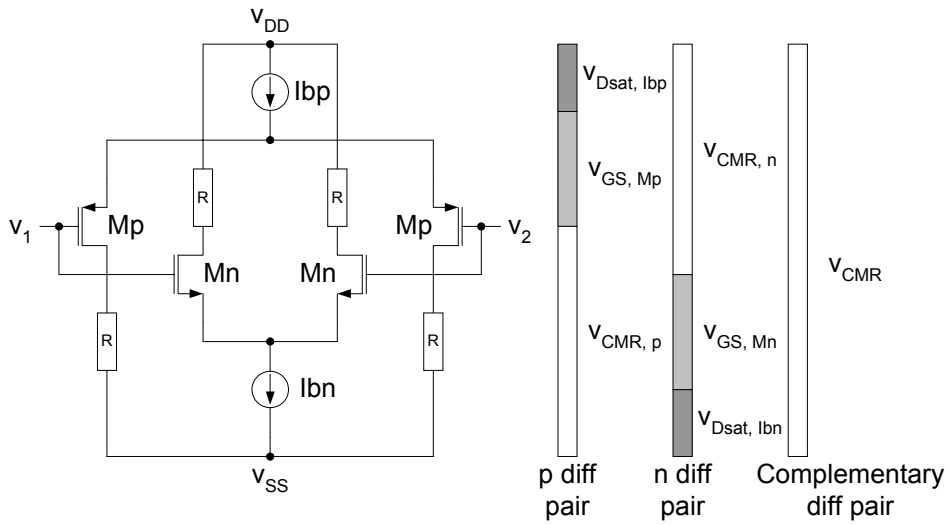

Figure 5: Complementary differential pair common mode input range

For low common mode input, the PMOS differential pair is in saturation and NMOS is off. For high common mode input, the NMOS differential pair is in saturation and PMOS is off. Therefore, the total effect is that the complementary differential pair is always working and the rail to rail common mode input requirement is met. It should be noted that for common mode input in the middle region both pairs are working, this will have a significant effect on the performance of the circuit. To understand the effect, we will investigate how the transconductance of each pair and of the complementary pair changes with common mode input signal. First the transconductance verses input common mode of the NMOS pair is shown in Figure 6.

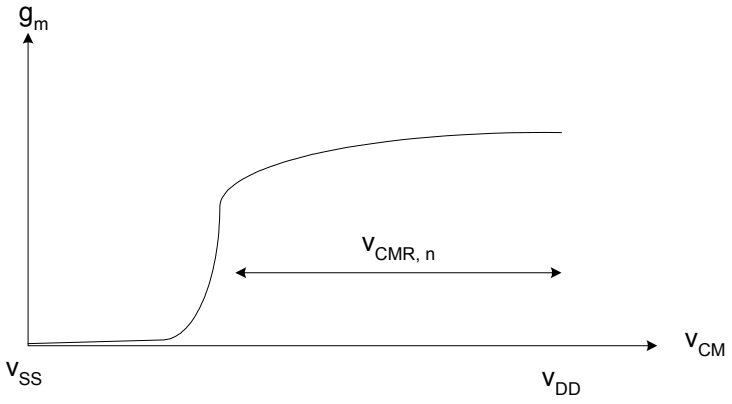

Figure 6: NMOS differential pair transconductance verses input common mode

Similarly, the transconductance verses input common mode of the PMOS pair is shown in Figure 7.

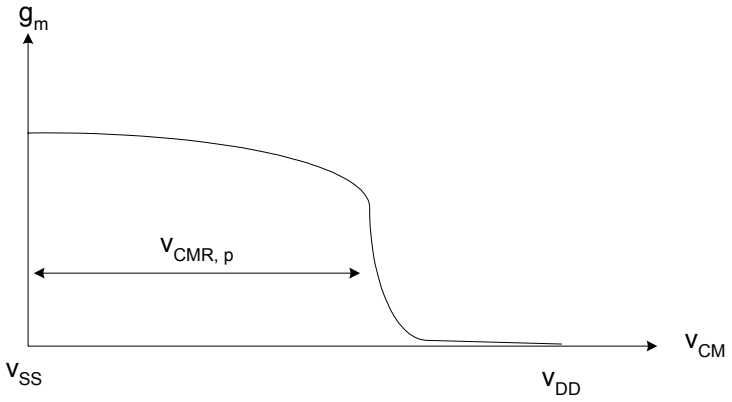

Figure 7: PMOS differential pair transconductance verses input common mode

We see that the transconductance of each pair is almost constant over its common mode range and drops to zero outside this range. Combining these two graphs gives the transconductance verses input common mode of the complementary pair as shown in Figure 8. It is assumed here that both pairs in the complementary structure had been sized appropriately to obtain equal transconductance in their region of operation.

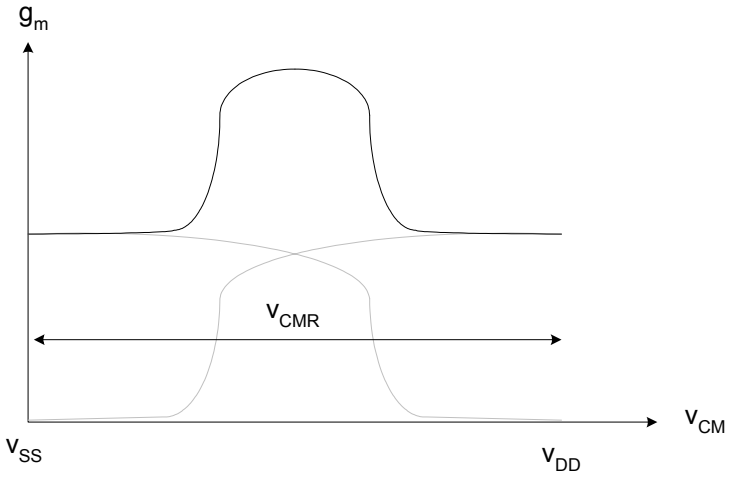

Figure 8: Complementary differential pair transconductance verses input common mode

The transconductance of the complementary differential pair ($g_{m,np}$) is almost constant for high or low common mode input when only one of the pairs is active. In the middle region, both pairs are on and the effective transconductance is twice that of the other regions. The large variations in the transconductance will result in a suboptimal op-amp design with the complementary differential pair used as the input stage. This can be understood by studying the design of a two stage op-amp shown in Figure 9.

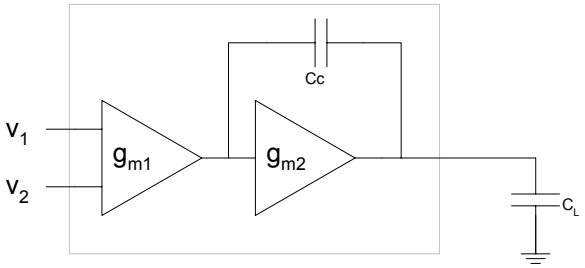

Figure 9: Two stage op-amp

Where C_c is the compensation capacitor and C_L is the load capacitor. The unity-gain bandwidth of the amplifier is approximately given by:

$$\omega_t = \frac{g_{m1}}{C_c} \tag{2}$$

And the second pole is approximately given by:

$$\omega_{p2} = \frac{g_{m2}}{C_L} \tag{3}$$

Also, for stability reasons, we wish to satisfy the following condition:

$$\omega_{p2} > \alpha\, \omega_t = \alpha\, \frac{g_{m1}}{C_c} \tag{4}$$

Where α is a multiplication factor determined by the stability requirement. For example, for a phase margin of 60º, α is about 2.

In order to maintain a minimum unity-gain bandwidth, a lower limit is set on the value of g_{m1} as seen from (2), assuming C_c is constant. Therefore, the worst case condition for ω_t is when g_{m1} is at its minimum. On the other hand, to maintain a minimum distance between first and second poles, an upper limit is set on the value of g_{m1} as seen from (4), assuming C_c and α are constants. Therefore, the worst case condition for ω_{p2} is when g_{m1} is at its maximum. There is a clear conflict of requirements between the two

conditions. In the complementary differential pair, $g_{m,np}$ is twice as large in the middle of the common mode rage as the sides. Therefore, when the minimum required $g_{m,np}$ is evaluated as required by ω_t, it determines $g_{m,np}$ in the side regions. Then, α is set based on the middle $g_{m,np}$ which is the stability's worst case condition. Therefore, α is twice as large as it should be without $g_{m,np}$ variations. From (2), this requires g_{m2} to be twice as large; this translates into increased power consumption.

Another less significant disadvantage of the varying transconductance is harmonic distortion. In op-amps, the distortion will be significantly reduced when the feedback network is added to the amplifier. Therefore, this effect is not usually an issue.

The large variation of the complementary differential pair transconductance is not desirable. Its main disadvantage is the power suboptimal frequency compensation of the amplifier. Methods for reducing the transconductance variation are discussed next.

4. Constant Transconductance Input Stage

We need to formulate $g_{m,np}$ to explore possible methods for achieving a constant transconductance input stage. Assuming all transistors are in the saturation region and using the square law models, the complementary differential pair transconductance is given by:

$$g_{m,np} = g_{m,n} + g_{m,p}$$

$$g_{m,np} = \sqrt{\mu_n C_{ox} \left(\frac{W}{L}\right)_n I_n} + \sqrt{\mu_p C_{ox} \left(\frac{W}{L}\right)_p I_p} \qquad (5)$$

or:

$$g_{m,np} = \mu_n C_{ox} \left(\frac{W}{L}\right)_n V_{eff,n} + \mu_p C_{ox} \left(\frac{W}{L}\right)_p V_{eff,p} \qquad (6)$$

Investigating these formulas gives us suggestions on how to keep $g_{m,np}$ constant. From (5), we see that we can accomplish this by controlling the tail current or the aspect ratio of either NMOS or PMOS differential pair or both. From (6), we see that we can also obtain a constant $g_{m,np}$ by controlling the effective voltage of the transistor.

4.1. Constant $g_{m,np}$ by means of tail current control

We will start with a condition to simplify the relationship given by (5). If we choose:

$$\sqrt{\mu_n C_{ox} \left(\frac{W}{L}\right)_n} = \sqrt{\mu_p C_{ox} \left(\frac{W}{L}\right)_p} = K \tag{7}$$

then, (5) simplifies to:

$$g_{m,np} = K\left(\sqrt{I_n} + \sqrt{I_p}\right) \tag{8}$$

Therefore, to keep $g_{m,np}$ constant, we have to keep $\left(\sqrt{I_n} + \sqrt{I_p}\right)$ constant. This can be done using translinear circuits to obtain a square root biasing scheme. The disadvantage of this design is its relative complexity. Also, it is not very accurate since it uses the square law models of the MOS transistors and for today's deep submicron technology; there is a significant deviation from the ideal relation.

The constant $\left(\sqrt{I_n} + \sqrt{I_p}\right)$ method was used in [4] and 10% $g_{m,np}$ variations was achieved. It was also used in [5] with 5% variations.

Another method that uses tail current to maintain constant $g_{m,np}$ can be understood by putting another condition $\left(\sqrt{I_n} = \sqrt{I_p} = \sqrt{I}\right)$, therefore (8) becomes:

$$g_{m,np} = 2K\sqrt{I} \tag{9}$$

However, $g_{m,np}$ is only half the value given in (9) when only one differential pair is operating because I is half. Therefore, in the regions where one pair is working, an increase in the tail current by a factor of 4 will keep $g_{m,np}$ constant. This can be implemented by diverting the tail current of the non working pair to the working pair after multiplying by a factor of 3. The multiplication by 3 can be accomplished by using a three-times current mirror. The disadvantage of this technique is the limited accuracy due to ideal square law dependence (9). Also, the transition between the three regions is not well controlled and will cause variations in $g_{m,np}$.

The constant $\left(\sqrt{I}\right)$ method was used in [6] and 15% $g_{m,np}$ variations was achieved. It was also used in [7]and [8] with 15% variations in both.

4.2. Constant $g_{m,np}$ by means of voltage control

We can rewrite equation (6) using condition (7) to simplify the expression, we get the following:

$$g_{m,np} = K^2\left(V_{gs,n} + V_{sg,p} - V_{thn} - \left|V_{thp}\right|\right) \qquad (10)$$

Therefore, to keep $g_{m,np}$ constant, we have to keep $\left(V_{gs,n} + V_{sg,p}\right)$ constant. This can be done by connecting a voltage source between the common source of the NMOS and the common source of the PMOS differential pairs. This will keep both pairs working for all common mode input voltage. One method for implementing the voltage source is by using two diode connected MOS transistors and sizing them appropriately. The disadvantage of this technique is that the behavior of the diode connected transistors is a function of the voltage across them. Therefore, $g_{m,np}$ still has some variations over the common mode input range.

The constant $\left(V_{gs,n} + V_{sg,p}\right)$ method was used in [9], 8% and 28% $g_{m,np}$ variations was achieved. The authors in this paper used two different methods to control $g_{m,np}$ by voltage adjustment, this accounts for the two values given here.

4.3. Constant $g_{m,np}$ by means of aspect ratio control

A closer look into (5) or (6), reveals the possibility of controlling $g_{m,np}$ by controlling the aspect ratio of one or both differential pairs. One method for implementing this is by using a second NMOS and a second PMOS differential pairs in standby that are connected in parallel to the complementary differential pair. In the middle region of operation, both pairs in the complementary differential pair are operating and the other two pairs are off. To explain how the circuit operates in the other two regions, let's start with the input common mode voltage in the middle region and going down towards Vss. When the voltage becomes low enough for the PMOS pair to turn off, the other NMOS differential pair will be activated. Effectively, doubling the aspect ratio for low common mode input, therefore, $g_{m,np}$ is constant. The same happens if the voltage increases towards Vdd, the second PMOS differential pair will be activated and the effective aspect ratio is doubled to keep $g_{m,np}$ constant. The disadvantage of this technique is that in the transition between the three regions results in a non-smooth

current transition. This causes relatively large variations in $g_{m,np}$ which results in a suboptimal performance.

The aspect ratio control method was used in [10] and 20% $g_{m,np}$ variations was achieved.

4.4. Constant $g_{m,np}$ by alternative means

This section reports two alternative methods for making $g_{m,np}$ constant. The first method utilizes a maximum selection circuit. The principle of operation is to allow only one pair to operate in the middle region. This can be done by keeping the differential pair that has the larger tail current operating and turn the other differential pair off. The scheme assumes that the differential pair with larger tail current is operating properly while the other pair is going into (or just coming out of) triode region. The principle is illustrated in Figure 10. The disadvantage of this design is its relative complexity. This method was used in [11] and [12] which achieves 5% $g_{m,np}$ variations.

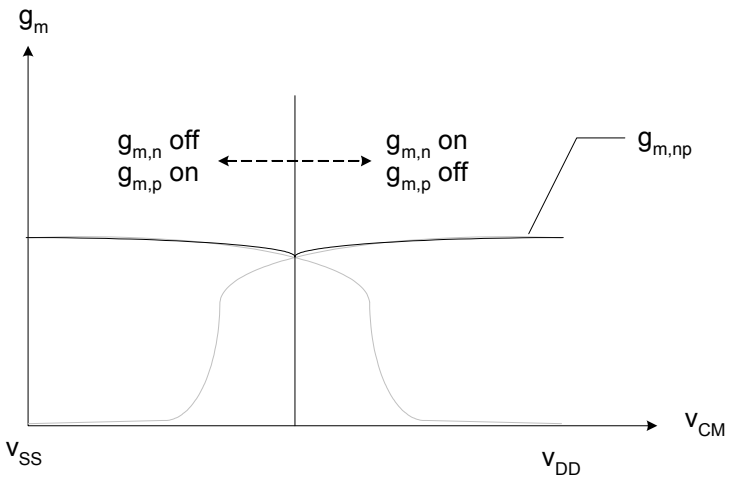

Figure 10: Selecting maximum g_m

Another alternative method uses DC level shifting. The principle of operation is to shift the transconductance characteristics curve of the NMOS differential pair to the left or the transconductance characteristics curve of the PMOS differential pair to the right such that the sum of g_{mn} and g_{mp} is constant. This can be accomplished simply by applying a DC level shift to the NMOS pair (or PMOS pair) to make its turn on voltage higher (or lower). The principle is illustrated in Figure 11 for NMOS shift. The

disadvantage of this method is its need for tuning. Because the characteristics will vary with process, voltage, and temperature, the optimal DC level will change. If the shift level is not tuned, large or small $g_{m,np}$ will appear around the transition region, therefore, instability might occur. This method was used in [13] and achieves 13% $g_{m,np}$ variations before tuning and 5% $g_{m,np}$ variations after tuning.

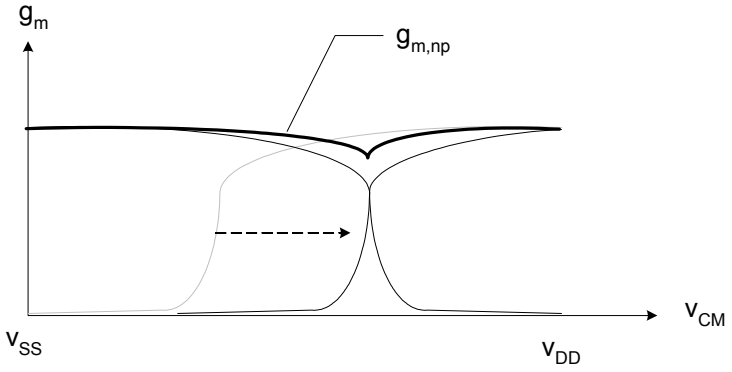

Figure 11: DC level shifting

4.5. Summary of constant $g_{m,np}$ methods

This section provides a summary of all the methods discussed to achieve constant transconductance for the complementary differential input stage.

Principle	$g_{m,np}$ variation
Tail current control I: $\left(\sqrt{I_n}+\sqrt{I_p}\right)=cons\tan t$	10% [4] 5% [5]
Tail current control II: $\left(\sqrt{I}\right)=cons\tan t$	15% [6], [7], and [8]
Voltage control: $\left(V_{gs,n}+V_{sg,p}\right)=cons\tan t$	8% [9] 28% [9]
Aspect ratio control	20% [10]

Maximum selection circuit	5% [11] and [12]
DC level shifting	13% [13] before tuning 5% [13] after tuning

5. Conclusion

Operational amplifiers input stages utilizing a single differential pair have a common mode input range that extends to only one rail. This limits the application where the op-amp can be used. A rail to rail common mode input range is a desirable characteristic for the input stage to make it more versatile. This characteristic can be achieved using a compound differential pair structure called the complementary differential pair. A drawback of this structure is that its transconductance varies by a factor of two over the range of operation. This variation results in a suboptimal frequency compensation of op-amps because it requires much more power. Therefore, regulating the transconductance to keep it almost constant over the entire common mode range is essential.

This paper provides an overview of some techniques used to achieve near constant transconductance for the complementary differential pair operating in strong inversion. Three systematic methods based on mathematical understanding of the transconductance behavior were described. They can be classified as: tail current control, voltage control, and aspect ratio control. Two more alternative methods using maximum selection and DC level shifting were also described. The best performers achieved 5% transconductance variation over the entire common mode range. However, they are also the more complicated techniques.

As the supply voltage shrinks for newer CMOS technologies, some of the techniques presented here will be obsolete. New methods have to be developed to achieve rail to rail common mode operation and capable of operating in future CMOS technologies.

References

[1] David A. Johns and Ken Martin, “Analog Integrated Circuit Design”, New York: Wiley, 1997.

[2] Ron Hogervorst and Johan H. Huijsing, “Design of Low-Voltage, Low-Power Operational Amplifier Cells,” Kluwer Academic publishers, 1996.

[3] Shouli Yan and Edgar Sanchez-Sinencio, “Low Voltage Analog Circuit Design Techniques: A Tutorial,” IEICE Trans. Analog Integrated Circuits and Systems, vol. E00-A, no. 2 February 2000.

[4] S. Sakurai and M. Ismail, “Robust design of rail-to-rail CMOS operational amplifiers for a low power supply voltage,” IEEE Journal of Solid-State Circuits, vol. 31, no. 2, pp. 146-156, February 1996.

[5] J. H. Botma, R. F. Wassenaar, and R. J. Wiegerink, “A low voltage CMOS op amp with a rail-to-rail constant-gm input stage and a class AB rail-to-rail output stage,” EEE Proc. ISCAS 1993, vol. 2, pp. 1314-1317, May 1993

[6] R. Hogervorst, R. J. Wiegerink, P. A. L. de Jong, J. Fonderie, R. F. Wassenaar, and J. H. Huijsing, “CMOS low-voltage operational amplifiers with constant-gm rail-to-rail input stage,” IEEE Proc. ISCAS 1992, pp. 2876-2879.

[7] R. Hogervost, J. P. Tero, R. G. H. Eschauzier and J. H. Huijsing, “A compact power-efficient 3-V CMOS rail-to-rail input/output operational amplifier for VLSI cell libraries,” IEEE Journal of Solid-State Circuits, vol. 29, no. 12, pp. 1505- 1513, December 1994.

[8] J. H. Huijsing, R. Hogervorst, and K.-J. de Langen, “Low-power low-voltage VLSI operational amplifier cells,” IEEE Trans. Circuits and Systems-I, vol. 42. no. 11, pp. 841-852, November 1995.

[9] R. Hogervost, J. P. Tero, and J. H. Huijsing, “Compact CMOS constant-gm rail-to-rail input stage with gm-control by an electronic zener diode,” IEEE Journal of Solid-State Circuits, vol. 31, no. 7, pp. 1035-1040, July 1996.

[10] R. Hogervorst, S. M. Safai, and J. H. Huijsing, “A programmable 3-V CMOS rail-to-rail opamp with gain boosting for driving heavy loads,” IEEE Proc. ISCAS 1995, pp. 1544-1547.

[11] C. Hwang, A. Motamed, and M. Ismail, “LV opamp with programmable rail-to- rail constant-gm,” IEEE Proc. ISCAS 1997, pp. 1988-1959.

[12] C. Hwang, A. Motamed, and M. Ismail, “Universal constant-gm input-stage architecture for low-voltage op amps,” IEEE Trans. Circuits and Systems-I, vol. 42. no. 11, pp. 886-895, November 1995.

[13] M. Wang, T. L. Mayhugh, Jr., S. H. K. Embabi, and E. Sánchez-Sinencio, “Constant-gm rail-to-rail CMOS op-amp input stage with overlapped transition region,” IEEE Journal of Solid-State Circuits, vol. 34, no. 2, pp. 148-156, February 1999.