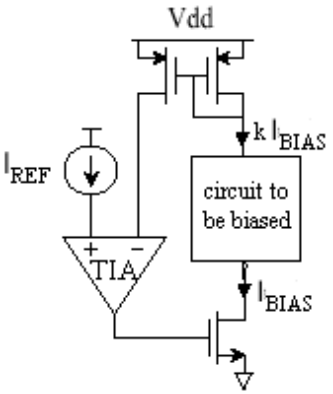
So as to take the advantage, V_{shift} must be positive; therefore $|V_{GS3}|>|V_{GS4}|$ must be satisfied (Choose $I_{B2}<½I_{B2}$ and/or $\beta_3<\beta_4$).

There exists a type of ***very-low-V_{OUTmin}*** *and* ***high-output-resistance*** current source structure that deserves mentioning. The conceptual structure is shown on the right.



Although the employed technique is not widely applicable, it can be very helpful for biasing of several widely-used structures, including differential pairs and regulated-cascode transconductors.

The technique works in circuits where a copy (kI_{BIAS}: **e.g. ½I_{BIAS}**) of the supplied bias current $\mathbf{I_{BIAS}}$ can be supplied for a prompt evaluation in a feedback loop. The *differential **T**rans**I**mpedance **A**mplifier* (**TIA**) helps keeping $\mathbf{I_{BIAS}=I_{REF}}$ (*or* $I_{BIAS}=I_{REF}/k$).

! ! ! *The stability of the utilized loop is another important issue.*

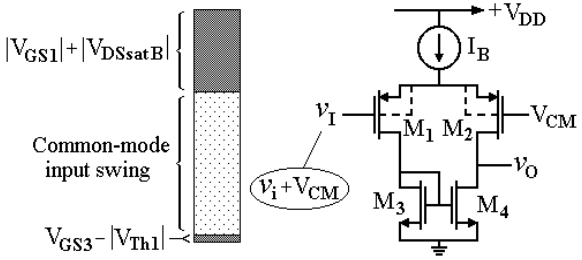
H.W.#2 : Find and explain <u>one</u> transistor-level example for this circuit.

Constant-g_m rail-to-rail input stages

The input stage of an operational amplifier is very typically a *differential pair* (often called **"long tailed pair"**).

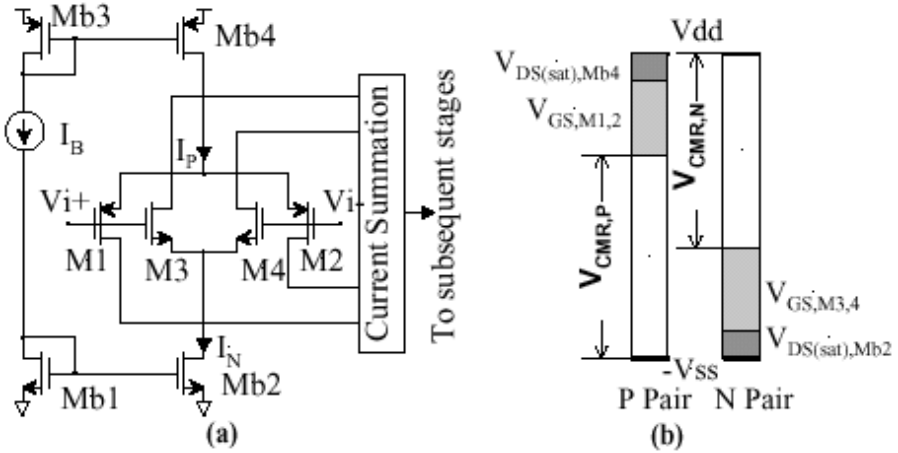
For linear applications, opamps are usually used in local closed loop configurations, therefore the differential input voltage is quite small. Therefore, the main voltage swing limitation comes from the common-mode (CM) input voltage swing.

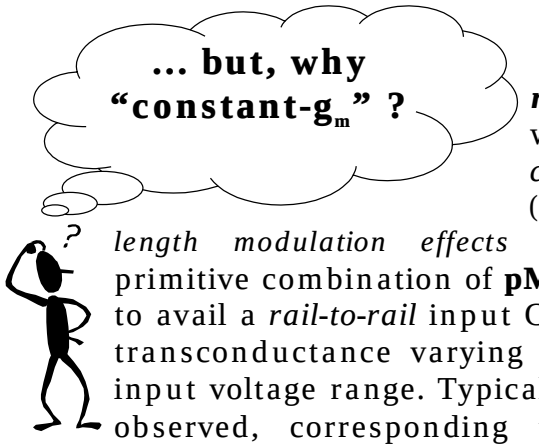
CM *input voltage swing* of the conventional differential pair is very good in one direction, but restricted in the other direction.



For LV operation, the CM input range must be widened as much as possible. A **rail-to-rail** CM input range would be appreciated a lot.

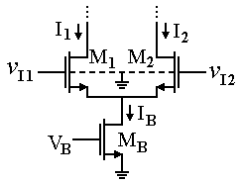
Combining one **pMOS-input** and one **nMOS-input** differential pair helps obtaining an almost *rail-to-rail* input CM voltage swing.



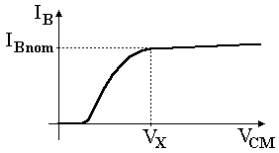


An ***only-pMOS*** or ***only-nMOS*** differential pair would supply an *almost constant* transconductance (*if body effect and channel length modulation effects are ignored*). However, a primitive combination of **pMOS** and **nMOS** input pairs, to avail a *rail-to-rail* input CM range, yields an effective transconductance varying significantly along the CM input voltage range. Typically, **3** different $\mathbf{G_m}$ levels are observed, corresponding to **3** different regions of operation summarized below:

$\mathbf{V_{CM}}$ close to **GND** : **pMOS** pair operates, **nMOS** pair **OFF**
$\mathbf{V_{CM}}$ close to $\mathbf{V_{DD}}$: **nMOS** pair operates, **pMOS** pair **OFF**
$\mathbf{V_{CM}}$ close to ½$\mathbf{V_{DD}}$: **both** pairs operate



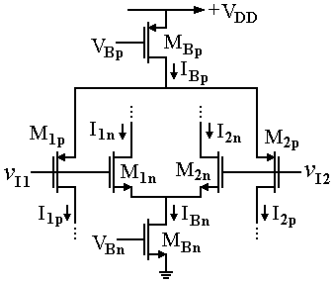
Down to the CM input voltage limit ($\mathbf{V_{CM}=V_X}$), a conventional differential pair is biased by an almost constant tail current $\mathbf{I_B=I_{Bnom}}$; thus the transconductance of the differential pair is almost constant: $\mathbf{G_m}= \partial(I_1-I_2)/\partial(v_{I1}-v_{I2})=2g_{m1}v_{GS1}/(2v_{GS1})= \mathbf{g_{m1,2}}$



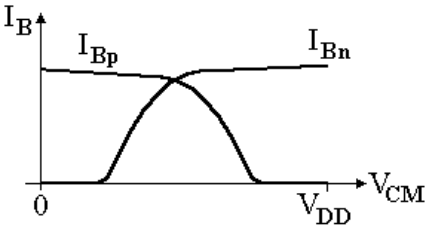
As shown above (*body effect omitted for brevity*), the transconductance of the differential pair is equal to the transconductance of the input transistors.

For the pMOS-nMOS combination case, a basic summation of the currents of the pairs will supply the replacement of $\mathbf{I_1}$ with $\mathbf{I_{1n}-I_{1p}}= I_{1n}+|I_{1p}|$ and $\mathbf{I_2}$ with $\mathbf{I_{2n}-I_{2p}}= I_{2n}+|I_{2p}|$.

Thus, the *new* $\mathbf{I_1}$ and $\mathbf{I_2}$ currents will allways be **non-zero** (*though not constant*), as $\mathbf{V_{CM}}$ is sweeped from one rail to the other.



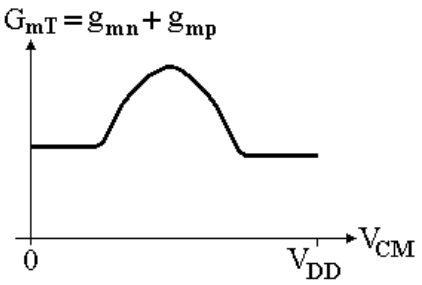
On the right, the tail currents are supplied for such a rail-to-rail input stage. There seems to exist 3 different regions (as mentioned before).



The equivalent transconductance ($\mathbf{G_{mT}}$) will simply be the the **sum** of the transconductances of the **nMOS** and **pMOS** input transistors.

$$\mathbf{G_{mT}} = \partial[(I_{1n}+|I_{1p}|)-(I_{2n}+|I_{2p}|)]/\partial(v_{I1}-v_{I2})$$
$$= \mathbf{g_{mn}}+\mathbf{g_{mp}} = \sqrt{(\beta_n I_{Bn})} + \sqrt{(\beta_p I_{Bp})}$$

It is apparent that a non-zero $\mathbf{G_{mT}}$ is available for the whole input CM range; thus amplification will allways be possible.



... BUT, what is the reason forcing us to keep the transconductance (G_{mT}) of the input stage <u>constant</u>?

For low frequencies, the open-loop small-signal differential voltage gain for a typical *2-stage* CMOS opamp can be generalized as

$$\mathbf{A_v = G_m R_{out1} A_{v2}}$$

where $\mathbf{G_m}$ and $\mathbf{R_{out1}}$ are respectively the transconductance and the output resistance of the first stage, and $\mathbf{A_{v2}}$ is the gain of the second stage (*the signs of the gains are deliberately omitted*). The dominant pole frequency of such an amplifier is created (*forced to be located*) at roughly $\mathbf{f_{p,d}=1/(2\pi R_{out1}A_{v2}C_c)}$. Here, $\mathbf{C_c}$ is the artificial (*compensation*) capacitance, used to shift the dominant pole away from the second (*non-dominant*) pole, towards low frequencies (Actually, what helps to create a *dominant* and a *non-dominant* pole is nothing but **compensation**).

The value of $\mathbf{C_c}$ is chosen such that, the non-dominant pole is located after the **GBW** frequency (after $|\mathbf{A_v}(f)|$ drops below **1**), to supply a sufficient phase margin. Since bandwidth will be approximately equal to the dominant pole, the **gain-bandwidth product** is,

$$\mathbf{GBW} = A_v f_{p,d} = G_m R_{out1} A_{v2}/(2\pi R_{out1} A_{v2} C_c) = \mathbf{G_m/(2\pi C_c)}$$

This result says, "**The GBW is directly dependent on G_m!**"

So, a ***varying*** $\mathbf{G_m}$ (**i.e.** a varying GBW) will require much more strict precautions to guarantee stability (**e.g.** Choosing a very large $\mathbf{C_c}$ to achieve a large enough phase margin for the *worst-case* GBW value), in turn, slowing down the amplifier severely.

This is a **critical problem** in a basic rail-to-rail opamp where, with its primitive form, $\mathbf{G_{mT}}$ will have 3 different values, depending on the common-mode input level $\mathbf{V_{CM}}$.

"Constant-g_m rail-to-rail input stages" come as a solution for this problem. There are several different techniques trying to "*regulate*" the transconductance of such an input stage along the *rail-to-rail* input common-mode range.

Numerous ***constant-g_m rail-to-rail input stage*** structures (more than **5** different approaches) will be described throughout this course.

Let's recall the transconductance of the rail-to-rail input stage:

$$G_{mT} = g_{mn} + g_{mp}$$
$$= \sqrt{(\beta_n I_{Bn})} + \sqrt{(\beta_p I_{Bp})}$$
$$= \beta_n(V_{GSn}-V_{THn}) + \beta_p(V_{GSp}-V_{THp})$$

Keeping G_{mT} constant along the CM input range can be achieved by using the one (*or simultaneously more the one*) of the following methods:

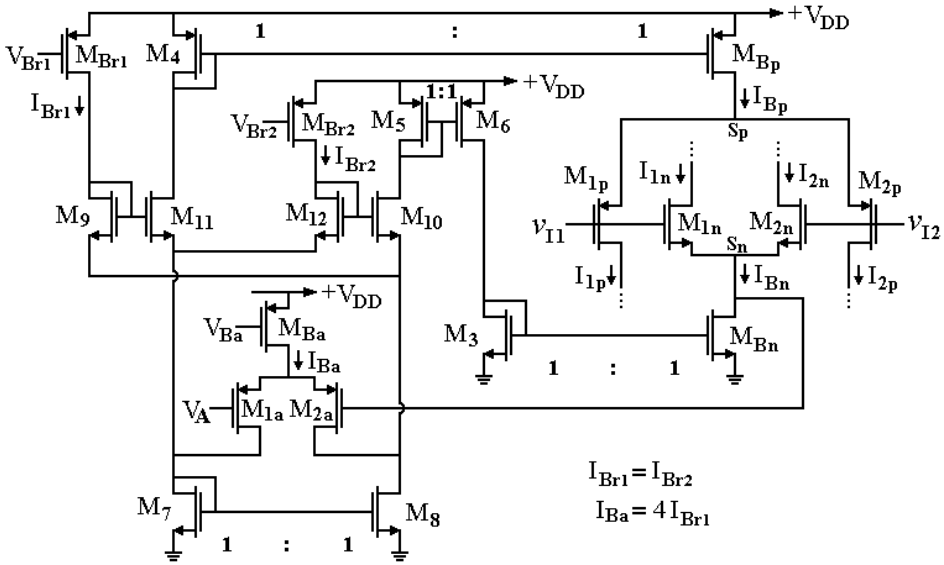
Current-based methods

- Set $\beta_n=\beta_p$ (*not easy*) and try to keep $\sqrt{I_{Bn}} + \sqrt{I_{Bn}}$ constant.
- Try to keep $\sqrt{(\beta_n I_{Bn})} + \sqrt{(\beta_p I_{Bp})}$ fixed ($\beta_n= \beta_p$ *not required*).
- Increase one I_B to $4I_B$ when the other is OFF ($\beta_n=\beta_p$ *required*).

Voltage-based methods

- Use a Zener diode (*or equivalent*) to keep $V_{GSn}+ V_{GSp}$ constant.
- Insert a voltage level shifter between inputs of ***p*** and ***n*** pairs.

Set $\beta_n=\beta_p$ and try to keep $\sqrt{I_{Bn}} + \sqrt{I_{Bn}}$ constant :



2 circuits using this method will be supplied. First one is shown above. In this circuit, $I_8=I_7$ with the help of the current mirror M_7-M_8. From the circuit, we observe $I_7 = I_{11}+ I_{12}+|I_{1a}|$ and $I_8 = I_9+ I_{10}+|I_{2a}|$. Then, since $I_9 = I_{Br1}$ and $I_{12}= I_{Br2}$, we obtain

$$I_{Br1}+ I_{10}+|I_{2a}|= I_{11}+ I_{Br2}+|I_{1a}|$$

Setting $I_{Br1}= I_{Br2}$ yields $I_{10}+|I_{2a}|= I_{11}+|I_{1a}|$. Also, it is apparent from the circuit that, $I_{Bn}= I_{10}$ and $I_{Bp}= I_{11}$. Therefore, we can write

$$I_{Bn}+|I_{2a}|= I_{Bp}+|I_{1a}|$$

Since $|I_{2a}|+|I_{1a}| = I_{Ba}$, i.e. $|I_{2a}|= I_{Ba}-|I_{1a}|$, we obtain

$$I_{Bn}+ I_{Ba}-2|I_{1a}|= I_{Bp}$$

This means, if I_{Bn} starts to decrease, $|I_{1a}|$ will also decrease ($|I_{2a}|$ will increase) to satisfy the equation (*actually, the equation is satisfied with aid of the feedback*). Then, for the $I_8=I_7$ equality to hold, I_{11} must be larger than I_{10} ; so it increases. Thus, I_{Bp} increases together with I_{11} . In this way, the a reduction in g_{mn} is compensated by a *counter action* (increment in g_{mp}) to keep overall G_{mT} constant (***Proven below***).

Using the loop along V_{GS} voltages of M_9 , M_{10} , M_{11} , M_{12} , we obtain

$$V_{GS9}- V_{GS10}+ V_{GS12}- V_{GS11}= 0 \quad \rightarrow \quad V_{GS9} + V_{GS12} = V_{GS10}+ V_{GS11}$$

Assuming all these transistors matched, we get

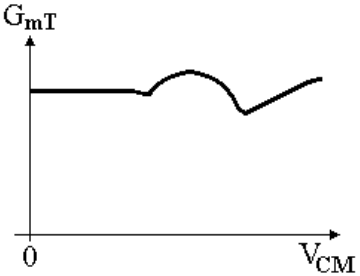
$$\sqrt{I_9} + \sqrt{I_{12}} = \sqrt{I_{10}} + \sqrt{I_{11}}$$

Since I_9 and I_{12} are constant and $I_{Bn}= I_{10}$ and $I_{Bp}= I_{11}$, finally

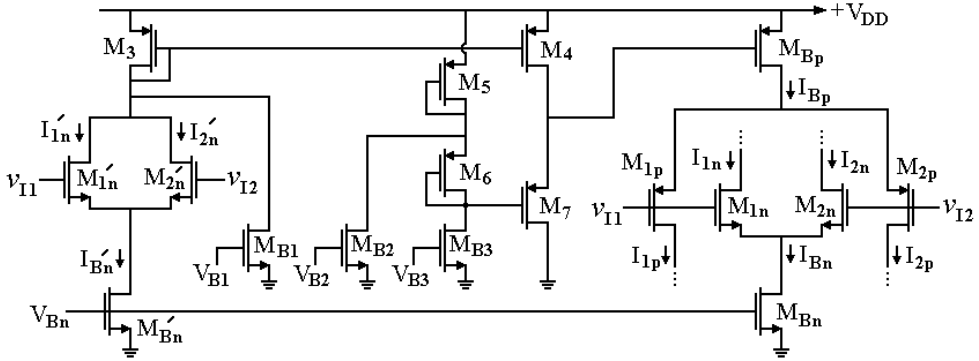
$$\sqrt{I_{Bn}} + \sqrt{I_{Bp}} = \sqrt{2}\, I_{B1} = \textbf{constant}$$

is obtained ($\beta_n=\beta_p$ should be achieved, which is not easy).

Although this result theoretically estimates a constant G_{mT} , especially because of the deviations from the square law and due to body effect, practically G_{mT} will not be absolutely constant along the CM range. A max. relative decline of **15%** (*or a bit more*) from the nominal G_{mT} value is typical.



The second circuit using the same method has a different topology but makes use of a similar approach (*A loop along V_{GS} voltages*).

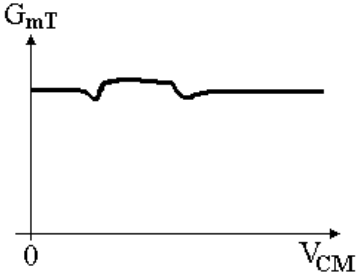


M_{Bn}• supplies an exact copy of $\mathbf{I_{Bn}}$ to M_3, in addition to $\mathbf{I_{B1}}$. Thus, $|I_7|=|I_4|=|I_3|=I_{Bn}+I_{B1}$. Also, $|I_5|=I_{B2}+I_{B3}$ and $|I_6|=I_{B3}$ can be observed. It is obvious that, $|V_{GSBp}|+|V_{GS7}|=|V_{GS5}|+|V_{GS6}|$. Then, if all these four transistors are matched, we obtain

$$\sqrt{\mathbf{I_{Bp}}}+\sqrt{(\mathbf{I_{Bn}}+\mathbf{I_{B1}})}=\sqrt{(\mathbf{I_{B2}}+\mathbf{I_{B3}})}+\sqrt{\mathbf{I_{B3}}}$$

Then, choosing $\mathbf{I_{B1}} << \mathbf{I_{Bn}}$ will be enough for achieving $\sqrt{\mathbf{I_{Bp}}}+\sqrt{\mathbf{I_{Bn}}}=$ **constant** .

Compared to the first circuit, this circuit seems to be achieving a less $\mathbf{G_{mT}}$ error along the CM input range; probably because a loop along V_{GS} voltages of pMOS transistors (*free of body effect*) are utilized. The relative $\mathbf{G_{mT}}$ error can be kept less than **10%**.



***A note by A.Z.**: Actually, using only $\mathbf{I_{B3}}$ would be sufficient; $\mathbf{I_{B1}}$ and $\mathbf{I_{B2}}$ seem useless. In this case, mathematically, $\sqrt{\mathbf{I_{Bp}}}+\sqrt{\mathbf{I_{Bn}}}=$ **constant** is achieved more accurately ($\sqrt{\mathbf{I_{Bp}}}+\sqrt{\mathbf{I_{Bn}}}=2\sqrt{\mathbf{I_{B3}}}$).*

*However, $\mathbf{I_{B1}}$ and $\mathbf{I_{B2}}$ are probably used for **trimming** purposes. **i.e.** to compensate for the errors caused by the deviation from square law, body effect, $\boldsymbol{\beta_n}$ **vs.** $\boldsymbol{\beta_p}$ inequalities, etc.*