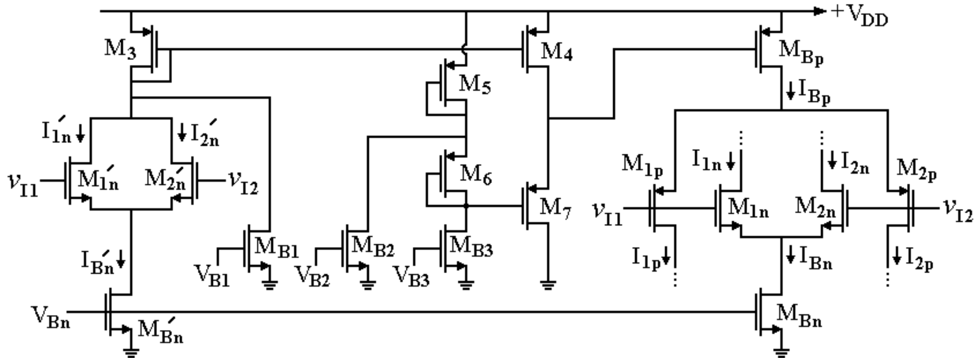
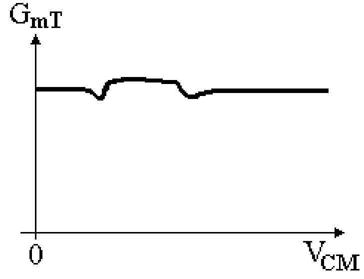
The second circuit using the same method has a different topology but makes use of a similar approach (*A loop along V_{GS} voltages*).



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

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

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



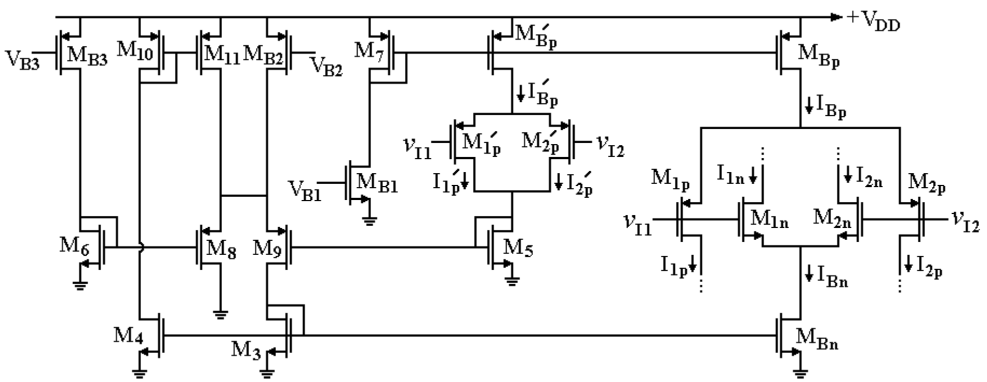
Compared to the first circuit, this circuit seems to be achieving a less $\mathbf{G_{mT}}$ error along the CM input range; probably because a loop along V_{GS} voltages of pMOS transistors (*free of body effect*) are utilized. The relative $\mathbf{G_{mT}}$ error can be kept less than **10%**.

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

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

Try to keep $\sqrt{(\beta_n I_{Bn})}+\sqrt{(\beta_p I_{Bp})}$ fixed ($\beta_n=\beta_p$ ***<u>not</u> required***) :

As an example, the constant-g_m rail-to-rail input stage below will be explained.



The auxilliary differential pair $\mathbf{M_{1p}}$•-$\mathbf{M_{2p}}$• receives a copy of $\mathbf{I_{Bp}}$ and this current is conveyed to $\mathbf{M_5}$. Then, after some processing, a proper $\mathbf{I_{Bn}}$ is obtained to achieve a **constant** $\sqrt{(\beta_n I_{Bn})}+\sqrt{(\beta_p I_{Bp})}$.

Notice that, $\mathbf{V_{GS5}+|V_{GS9}| = V_{GS6}+|V_{GS8}|}$. We can rewrite this as,

$$\sqrt{(2I_5/\beta_5)}+V_{TH5}+\sqrt{(2|I_9|/\beta_9)}+|V_{TH9}| = \sqrt{(2I_6/\beta_6)}+V_{TH6}+\sqrt{(2|I_8|/\beta_8)}+|V_{TH8}|$$

- Since V_{SB5} and V_{SB6} are zero, $\mathbf{V_{TH5}}$ and $\mathbf{V_{TH6}}$ are not affected by body effect; thus they are equal (equal to $\mathbf{V_{THn0}}$).
- Also, since sources of M_8 and M_9 are tied to the same node, $\mathbf{V_{SB8}}$ and $\mathbf{V_{SB9}}$ are equal (***i.e.***, $\mathbf{V_{TH8}}$ and $\mathbf{V_{TH9}}$ are modified equally by body effect). So, - *although it is possible* - there is no requirement for separate wells to obtain $\mathbf{V_{TH8}=V_{TH9}}$ equality.

As a result of $\mathbf{V_{TH}}$ equalities we obtain

$$\sqrt{(2I_5/\beta_5)}+\sqrt{(2|I_9|/\beta_9)} = \sqrt{(2I_6/\beta_6)}+\sqrt{(2|I_8|/\beta_8)}$$

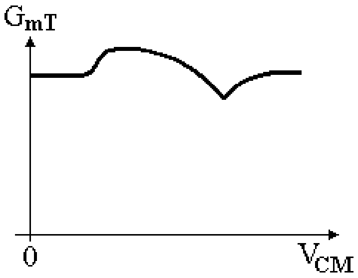
Note also that, $\mathbf{I_5=I_{Bp}}$, $\mathbf{I_6=I_{B3}}$, $\mathbf{|I_9|=I_{Bn}}$ and $\mathbf{|I_8| = I_{Bn}+I_{B2}-|I_9| = I_{B2}}$. So,

$$\sqrt{(2I_{Bp}/\beta_5)}+\sqrt{(2I_{Bn}/\beta_9)} = \sqrt{(2I_{B3}/\beta_6)}+\sqrt{(2I_{B2}/\beta_8)}$$

is obtained. By choosing $\beta_5=\beta_n$, $\beta_9=\beta_p$ and by multiplying both sides by $\beta_n\beta_p/\sqrt{2}$, we reach the desired result:

$$\sqrt{(\beta_n I_{Bn})}+\sqrt{(\beta_p I_{Bp})} = \beta_n\beta_p\,[\,\sqrt{(I_{B3}/\beta_6)}+\sqrt{(I_{B2}/\beta_8)}\,] = \textbf{constant}$$

A typical small-signal $\mathbf{G_{mT}}$ curve along the *rail-to-rail* CM input voltage range is given on the right. Typical max. relative $\mathbf{G_{mT}}$ deviation is **15-20%** (*peak to peak*).



The deviations are mainly due to some imperfections (*deviations from square law, β differences due to body effect on nMOS input transistors, etc.*)

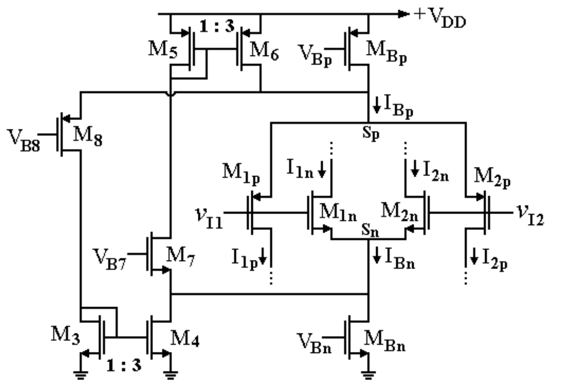
Increase one I_B to $4I_B$ when the other is OFF ($\beta_n=\beta_p$ *required*) :

By utilizing current switching, one differential pair can be forced to overtake the role of the other (when the other dif. pair is OFF). Thus the overall $\mathbf{G_{mT}}$ can be kept constant (*still* $\beta_n=\beta_p$ *required* ☹).

It can be estimated beforehands that, within the proximity of the current switching instants, there will be some "sudden" changes in $\mathbf{G_{mT}}$

Two structures will be supplied for this approach; one is below.

$\mathbf{M_7}$ and $\mathbf{M_8}$ act as switches, biased (not controlled - *at least directly*) by $\mathbf{V_{B7}}$ and $\mathbf{V_{B8}}$, respectively. The *direct* control of these switches are actually via the potential variations at nodes $\mathbf{S_p}$ and $\mathbf{S_n}$.



$\mathbf{V_{B7}}$ and $\mathbf{V_{B8}}$ chosen such that

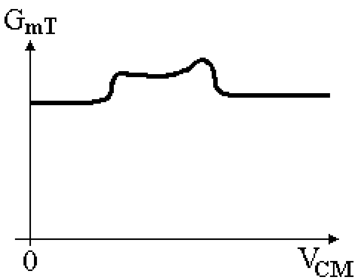
- $\mathbf{M_7}$ is **ON** when $\mathbf{M_{1p}}$-$\mathbf{M_{2p}}$ pair is **OFF** (**i.e.** $\mathbf{V_{Sn}}$ tends to **GND**)
- $\mathbf{M_8}$ is **ON** when $\mathbf{M_{1p}}$-$\mathbf{M_{2p}}$ pair is **OFF** (**i.e.** $\mathbf{V_{Sp}}$ tends to $\mathbf{V_{DD}}$)

☝ $\mathbf{M_{Bn}}$ and $\mathbf{M_{Bp}}$ are both supplying $\mathbf{I_B}$.

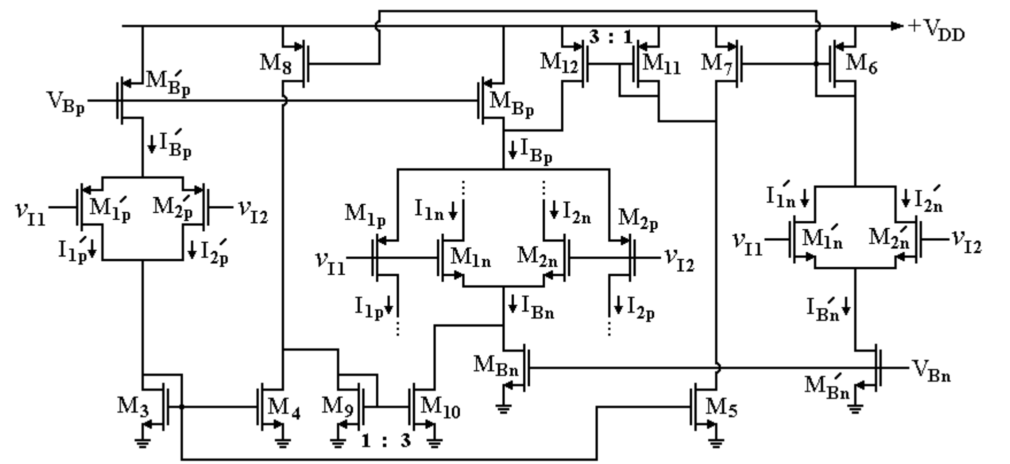
- When both pairs are **ON**: $\mathbf{M_7}$ and $\mathbf{M_8}$ are **OFF** → $\mathbf{I_{D3}}$, $\mathbf{I_{D4}}$, $\mathbf{I_{D5}}$, $\mathbf{I_{D6}}$ = **0A** . Thus, the circuit behaves like a basic *rail-to-rail* input stage.
- nMOS pair **ON**, pMOS pair **OFF**: $\mathbf{M_7}$ is **OFF** → $\mathbf{I_{D5}}$, $\mathbf{I_{D6}}$ = **0A** . On the other hand, $\mathbf{M_8}$ is **ON** → $\mathbf{I_{D3}=I_B}$ (*because*, $\mathbf{I_{Bp}=0}$) → $\mathbf{I_{D4}=3I_{D3}=3I_B}$ → $\mathbf{I_{Bn}=4I_B}$
- pMOS pair **ON**, nMOS pair **OFF**: $\mathbf{M_8}$ is **OFF** → $\mathbf{I_{D3}}$, $\mathbf{I_{D4}}$ = **0A** . On the other hand, $\mathbf{M_7}$ is **ON** → $|\mathbf{I_{D5}}|=\mathbf{I_B}$ (*because*, $\mathbf{I_{Bn}=0}$) → $|\mathbf{I_{D6}}|=3|\mathbf{I_{D5}}|=3\mathbf{I_B}$ → $\mathbf{I_{Bp}=4I_B}$

$$\mathbf{G_{mT}=g_{mn}+g_{mp}}=\sqrt{(\beta_n I_{Bn})}+\sqrt{(\beta_p I_{Bp})}=\begin{cases}\sqrt{(4\beta_n I_B)}=2\sqrt{(\beta I_B)}\ , & n \text{ ON } / \ p \text{ OFF}\\ \sqrt{(\beta_n I_B)}+\sqrt{(\beta_p I_B)}=2\sqrt{(\beta I_B)}\ , & n{+}p \text{ ON}\\ \sqrt{(4\beta_p I_B)}=2\sqrt{(\beta I_B)}\ , & p \text{ ON } / \ n \text{ OFF}\end{cases}$$

As a result, if $\beta_n=\beta_p$ equality can be achieved, an almost constant equivalent transconductance can be obtained for all regions, as shown on the right. This is a typical $\mathbf{G_{mT}}$ curve. The sharp transitions are due to current switching. It is obvious that, it is not easy to obtain a ***fully flat*** curve. **15%** max. relative error is typical.



In an alternative structure (see the figure next page), the current switching is made gradual, such that, the transitions of input pairs from ON to OFF (*and OFF to ON*) regimes are much smoother. Thus, the sharp transitions (*thus, inevitably larger relative errors*) in the $\mathbf{G_{mT}}$ curve are aimed to be eliminated.



The auxilliary pairs $\mathbf{M_{1n}\bullet}$-$\mathbf{M_{2n}\bullet}$ and $\mathbf{M_{1p}\bullet}$-$\mathbf{M_{2p}\bullet}$ helps obtaining precise replicas of drain currents of $\mathbf{M_{Bn}}$ and $\mathbf{M_{Bp}}$. These copiesare compared and depending on which one is larger, $\mathbf{I_{Bn}}$ and $\mathbf{I_{Bp}}$ are assigned appropriate values.

- When $\mathbf{I_{Bn}\bullet > I_{Bp}\bullet}$, $\mathbf{M_9}$ will receive the current difference ($\mathbf{I_{Bn}\bullet - I_{Bp}\bullet}$), thus $\mathbf{I_{Bn} = I_{Bn}\bullet + 3(I_{Bn}\bullet - I_{Bp}\bullet)}$; however, $\mathbf{M_{11}}$ will be **OFF** (*it cannot handle a current inwards, so M_7 will be driven into triode region to "worship" M_5*), thus $\mathbf{M_{12}}$ will be **OFF**, yielding $\mathbf{I_{Bp} = I_{Bp}\bullet}$.
- When $\mathbf{I_{Bn}\bullet < I_{Bp}\bullet}$, $\mathbf{M_{11}}$ will receive the current difference ($\mathbf{I_{Bp}\bullet - I_{Bn}\bullet}$), thus $\mathbf{I_{Bp} = I_{Bp}\bullet + 3(I_{Bp}\bullet - I_{Bn}\bullet)}$; however, $\mathbf{M_9}$ will be **OFF** (*it cannot handle a current outwards, so M_4 will be driven into triode region to "worship" M_8*), thus $\mathbf{M_{10}}$ will be **OFF**, yielding $\mathbf{I_{Bn} = I_{Bn}\bullet}$.

If nominal values (**i.e.** *when $M_{Bn}/M_{Bn}\bullet$ and $M_{Bp}/M_{Bp}\bullet$ are in saturation*) of $\mathbf{I_{Bp}\bullet}$ and $\mathbf{I_{Bn}\bullet}$ are set equal (*equal to I_B*), then,

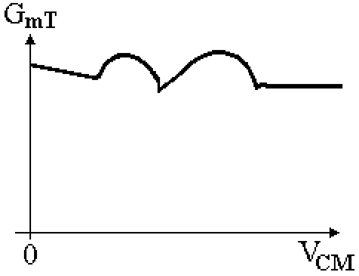
- When both pairs are ON and saturated → $\mathbf{I_{Bn} = I_{Bp} = I_B}$.
- When only the ***n*** pair is ON → $\mathbf{I_{Bn} = 4I_B}$.
- When only the ***p*** pair is ON → $\mathbf{I_{Bp} = 4I_B}$.

… ***BUT*** , in between these 3 regions, the "travel" of each tail current value from $\mathbf{I_B}$ to $\mathbf{4I_B}$ is made *gradual* for this structure.

Setting $\beta_n = \beta_p$ supplies a nominal transconductance of $\mathbf{G_{mT} = 2\sqrt{(\beta I_B)}}$

On the right is a typical $\mathbf{G_{mT}}$ curve.

Typical value for the max. relative $\mathbf{G_{mT}}$ error is **10%**, which is quite sufficient.



Note: *It may seem as if in some G_{mT} curves, the relative error is much larger. Usually the curves are "exaggerated" to supply an idea about the typical nature of the G_{mT}-V_{CM} curve.*

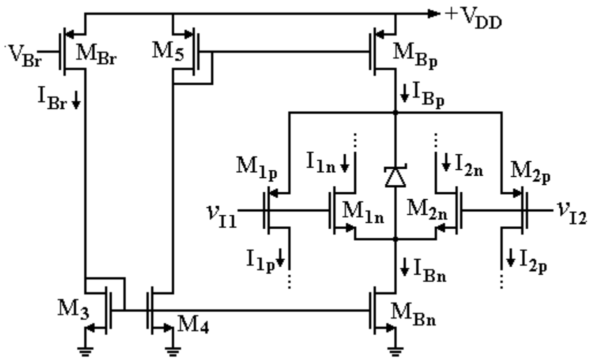
Use a Zener diode (*or equivalent*) to keep $V_{GSn} + V_{GSp}$ constant :

Let's recall the equivalent transconductance equation:

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

This equation says, " If $\beta_n = \beta_p$ and $\mathbf{V_{THn} = V_{THp}}$ can be assumed, then, keeping $\mathbf{G_{mT}}$ constant is reduced to keeping $\mathbf{V_{GSn} + V_{GSp}}$ constant"

Keeping a voltage constant can be regarded as a "DC regulation" issue. Therefore, Zener diodes are good choices; ☹ however they are not found in every process. Nevertheless, let's use the symbol for conceptual explanation purposes.



Compared to the current-based solutions, the most important advantage of this technique is the use of less number of extra current mirrors/sources (*lower power consumption*).

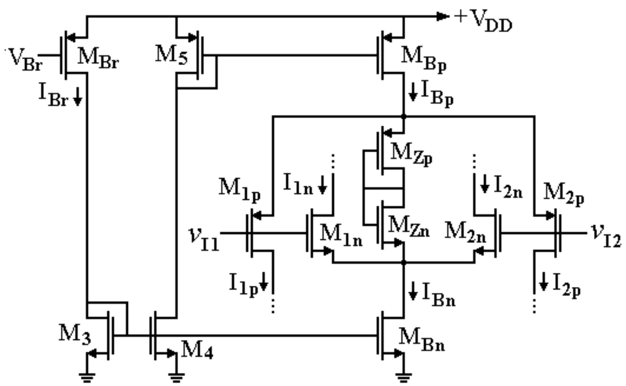
$\mathbf{I_{Bp}}$ and $\mathbf{I_{Bn}}$ must supply enough amount of current to help the Zener diode operate properly, as well the input pairs.

When $\mathbf{V_{CM}}$ is close to **GND** , the ***n*** pair is **OFF** and also Zener is **OFF** (since $\mathbf{V_{Sp}}$ is at a low value). Then the ***p*** pair will receive a tail current of $\mathbf{I_{Br}}$.
When $\mathbf{V_{CM}}$ is close to $\mathbf{V_{DD}}$, the ***p*** pair is **OFF** and also Zener is **OFF** (since $\mathbf{V_{Sn}}$ is at a high value). Then the ***n*** pair will receive a tail current of $\mathbf{I_{Br}}$.

In between, both pairs will be **ON**, each drawing **equal** tail currents (*since $\beta_n = \beta_p$ and $V_{THn} = V_{THp}$ can be assumed*) determined by the Zener voltage. Making this tail current equal to **¼**$\mathbf{I_{Br}}$ is vital to obtain a constant $\mathbf{G_{mT}}$ (then, the Zener receives **¾**$\mathbf{I_{Br}}$).

In the circuit below, a CMOS structure is used in place of the Zener diode, such that, when both pairs are ON, it receives **¾** of $\mathbf{I_{Br}}$.

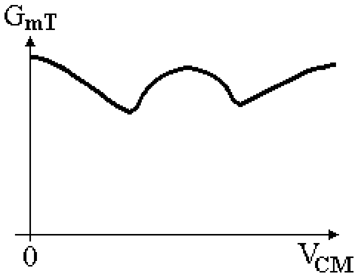
It shouldn't be hard to guess that the design must satisfy $\mathbf{\beta_{Zp} = 6\beta_p}$ and $\mathbf{\beta_{Zn} = 6\beta_n}$ (*in fact, $\beta_{Zp} = \beta_{Zn}$ should also be satisfied since $\beta_p = \beta_n$ is aimed*).



The pseudo-Zener can go OFF not sharply but only gradually (together with *one* of the input pairs). Therefore the $\mathbf{G_{mT}}$ cannot be kept satisfactorily constant.

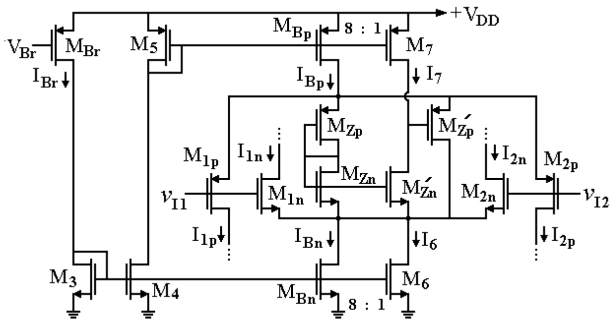
The maximum relative $\mathbf{G_{mT}}$ deviation is worse than that of a current-based method. However, the relatively lower power consumption is still an important advantage.

Typical value of maximum relative $\mathbf{G_{mT}}$ error is **20%**. Sharpening the ON-OFF transition is a must.

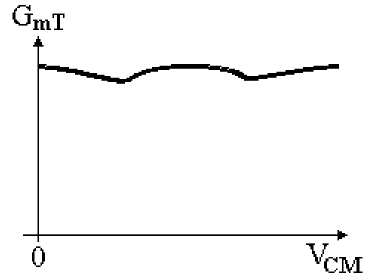


In the circuit on the right, additional devices are used (*but power consumption increases only marginally*) to obtain a better "Zener diode".

A maximum $\mathbf{G_{mT}}$ error less than **10%** is easily achievable.



With help of the added current sources and transistors, the ***V_{CM}-dependent*** "*current sharing* uncertainty" between the "Zener" and the input pairs has been dominantly prevented. In this new version, the "Zener" receives constant current when both input pairs are ON. Then the result is as shown on the right.



Insert voltage level shift between inputs of *p* and ***n* pairs :**

…