

Operational Transconductance Amplifiers

Achim Gratz

<mailto:Stromeko@Stromeko.net>

<http://Synth.Stromeko.net/diy/OTA.pdf>

Permission granted to link to the URL given above.
Do not make publicly available copies (aka mirroring).

This is a collection of various bits and pieces that I have found about OTA so I have to look in only one place. It was also motivated by the fact that the datasheets for commonly available OTA IC contain way too much handwaving and errors, some of them not very easy to spot. Of course I've probably added some errors of my own in this document, corrections and ideas for improvement are always welcome.

1 Preface

The OTA is popular for implementing voltage controlled oscillators (VCO) and filters (VCF) for analog music synthesizers, because it can act as a two-quadrant multiplier as we'll see later. For this application the control input has to have a wide dynamic range of at least 60 dB, while the OTA should behave sensibly when overdriven from the signal input (in particular, it should not lock up or phase reverse). Viewed from a slightly different angle an OTA can be used to implement an electrically tunable resistor that is referenced to ground, with extra circuitry floating resistors are possible as well.

The primary application for an OTA is however to drive low-impedance sinks such as coaxial cable with low distortion at high bandwidth. Hence, "improved" OTA such as the MAX436 or OPA660 have optimized these characteristics, but made it either impossible (MAX436) or considerably harder (OPA660) to use them as two-quadrant multipliers. Four quadrant multipliers on the other hand are hideously expensive, so that "obsolete" OTA like the CA3080 are still in widespread use.

2 Principle of Operation

An OTA is a voltage controlled current source, more specifically the term "operational" comes from the fact that it takes the difference of two voltages as the input for the current conversion. The ideal transfer characteristic is therefore

$$I_{\text{Out}} = g_m(V_{\text{In}+} - V_{\text{In}-}) \quad (1)$$

or, by taking the pre-computed difference as the input,

$$I_{\text{Out}} = g_m V_{\text{In}} \quad (2)$$

with the ideally constant transconductance g_m as the proportionality factor between the two. In reality the transconductance¹ is also a function of the input differential voltage and dependent on temperature, as we will later see.

To summarize, an ideal OTA has two voltage inputs with infinite impedance (i.e. there is no input current). The common mode input range is also infinite, while the differential signal between these two inputs is used to control an ideal current source (i.e. the output current does not depend on the output voltage) that functions as an output. The proportionality factor between output current and input differential voltage is called transconductance.

¹The term "transconductance" comes about because the ratio of the output current over the input voltage, g_m , has the unit of a conductance if looked at "across the amplifier". The proportional factor of output vs. input for an amplifier with current input and voltage output has the unit of a resistance and such an amplifier is called a transresistance amplifier.

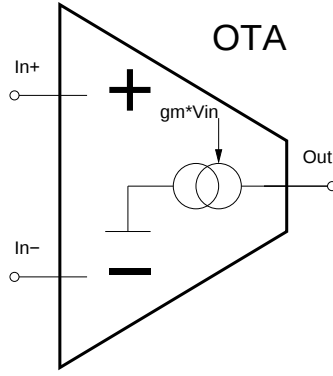


Figure 1: Ideal OTA

Any real OTA will thus have circuitry to process the input voltages with low input current over a wide common mode input range, to produce an internal representation of the input differential voltage and to provide a current to the output that is relatively independent of the output voltage. Since an OTA can be used without feedback, the maximum output current and with it the transconductance can often be adjusted.

2.1 The bipolar OTA

The most simple bipolar OTA consists of a differential pair to convert the input voltage difference to two currents I_+ and I_- . These two currents are then mirrored to the output so that their difference becomes the output of the OTA, while the rest of the OTA is made up of bias circuitry. The truly great feature of this “long-tailed differential pair” as it is often called is that the tail current, which is a necessary part of the biasing, can be used to control the transconductance as we will see in a moment.

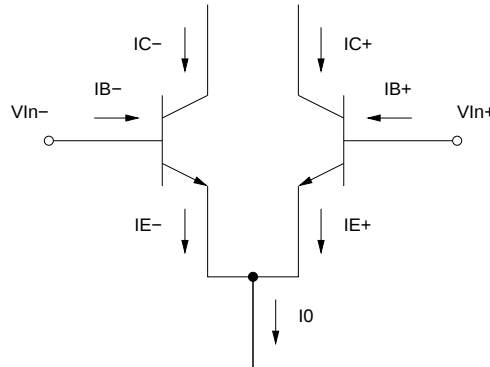


Figure 2: Bipolar Differential Pair (with npn transistors, biasing not shown)

2.1.1 The bipolar differential pair

The collector current of an npn transistor is (with some simplifying assumptions) related to its base-emitter voltage V_{BE} by

$$I_C = I_S \exp \frac{V_{BE}}{V_T}, \quad (3)$$

with the temperature voltage (k is the Boltzmann constant and q the elementary charge)

$$V_T = \frac{kT}{q}. \quad (4)$$

The collector current can also be expressed as a multiple of the base current, viewing the transistor as a current amplifier with a gain of β ,

$$I_C = \beta I_B, \quad (5)$$

which makes the emitter current²

$$I_E = -(I_C + I_B) = -(\beta + 1) I_B. \quad (6)$$

The tail current I_0 of the differential pair is composed of the emitter currents of the individual transistors.

$$I_0 = I_{E+} + I_{E-} \quad (7)$$

$$I_0 = \frac{\beta_+ + 1}{\beta_+} I_+ + \frac{\beta_- + 1}{\beta_-} I_- \quad (8)$$

and finally with $\beta_+ \gg 1$ and $\beta_- \gg 1$ this simplifies to

$$I_0 = I_+ + I_- \quad (9)$$

This simply means that as long as β is sufficiently high, its exact value is not at all important. Note however, that the β of a bipolar transistor is dependent on the collector current and therefore exact matching of β_+ and β_- can only occur at zero differential input voltage. Furthermore at low tail currents the error made in the simplification from (8) to (9) becomes quite noticeable as β drops off from its maximum value. Nevertheless for now we'll stick to the simplified equations and proceed to combine (3) and (9) to

$$I_0 = I_{S+} \exp \frac{V_{BE+}}{V_T} + I_{S-} \exp \frac{V_{BE-}}{V_T}. \quad (10)$$

When the transistors are matched and at the same temperature this results in

$$I_0 = I_S \left(\exp \frac{V_{BE+}}{V_T} + \exp \frac{V_{BE-}}{V_T} \right), \quad (11)$$

which can be solved for I_S to

$$I_S = \frac{I_0}{\exp \frac{V_{BE+}}{V_T} + \exp \frac{V_{BE-}}{V_T}}. \quad (12)$$

The output current of the OTA is the difference of the two collector currents in the pair

$$I_{Out} = I_+ - I_-, \quad (13)$$

and using (3) and (12) this gives the rather unwieldy expression

$$I_{Out} = I_0 \left(\frac{\exp \frac{V_{BE+}}{V_T}}{\exp \frac{V_{BE+}}{V_T} + \exp \frac{V_{BE-}}{V_T}} - \frac{\exp \frac{V_{BE-}}{V_T}}{\exp \frac{V_{BE+}}{V_T} + \exp \frac{V_{BE-}}{V_T}} \right), \quad (14)$$

which can be simplified to

$$I_{Out} = I_0 \left(\frac{1}{1 + \exp \frac{V_{BE-} - V_{BE+}}{V_T}} - \frac{1}{1 + \exp \frac{V_{BE+} - V_{BE-}}{V_T}} \right) \quad (15)$$

²Traditionally all currents for a single transistor are directed towards the crystal, hence the minus sign. The positive counting current direction in a circuit is often different for various reasons.

and further with $V_{In} = V_{BE+} - V_{BE-}$ to

$$I_{Out} = I_0 \left(\frac{1}{1 + \exp - \frac{V_{In}}{V_T}} - \frac{1}{1 + \exp \frac{V_{In}}{V_T}} \right). \quad (16)$$

You'll notice that the dependence on I_S is gone, thanks to the matching of both transistors and keeping them at the same temperature, but we're still not having an explicit and compact dependence on the input voltage. This is exactly what we'll develop next and we start by extending to the common denominator³:

$$I_{Out} = I_0 \left(\frac{\left(1 + \exp \frac{V_{In}}{V_T}\right) - \left(1 + \exp - \frac{V_{In}}{V_T}\right)}{\left(1 + \exp \frac{V_{In}}{V_T}\right) \left(1 + \exp - \frac{V_{In}}{V_T}\right)} \right) \quad (17)$$

which reduces to

$$I_{Out} = I_0 \left(\frac{\exp \frac{V_{In}}{V_T} - \exp - \frac{V_{In}}{V_T}}{2 + \exp \frac{V_{In}}{V_T} + \exp - \frac{V_{In}}{V_T}} \right) \quad (18)$$

which does not seem to look much better, but in fact this is

$$I_{Out} = I_0 \frac{2 \sinh \frac{V_{In}}{V_T}}{2 + 2 \cosh \frac{V_{In}}{V_T}} \quad (19)$$

which we find to correspond to (e.g. in [1])

$$I_{Out} = I_0 \tanh \frac{V_{In}}{2V_T}. \quad (20)$$

This puts us into a much better position to find out what g_m really is. The differential definition of the transconductance is:

$$g_m = \frac{dI_{Out}}{dV_{In}} \quad (21)$$

and with (20) we find

$$g_m = \frac{I_0}{2V_T} \operatorname{sech}^2 \frac{V_{In}}{2V_T}. \quad (22)$$

Thus we can finally show that the transconductance is anything but constant, depending both on temperature and input voltage as has been stated earlier. The second term is a bell shaped curve that equals 1 at zero input, falling off rapidly at both sides to asymptotically approach zero. The practical input range depends on how much error⁴ one is willing to tolerate, but seldom exceeds 20 mV. In fact, using (22) we find that to keep the linearity error below one percent (or -40dB below the signal) the input range is limited to $\pm 0.2V_T$ or 5 mV at room temperature. The maximum input range is approximately $\pm 5V_T$, 125 mV at room temperature or equivalently 28 dB of overdrive beyond the linear input range. Beyond this more than 99% of the tail current flows through just one of the two transistors and no changes in the output can be effected. The limiting action is comparably smooth, so overdriving an OTA from the input can be musically quite useful.

The temperature voltage in the argument of that term conspires to make the bell shape wider at higher temperature, which means that the linear input range of the OTA is smaller at low temperature as the g_m drops off more rapidly from its maximum value. Often you'll find just

³For the next steps you need to take a deep breath because I have to pull a stunt on you that I always hated when my math professors did it on me, because you sort of have to know what the result is before you can find the way to get there. An attempt at explaining of why and how to do this has been deferred to the appendix, along with some alternative derivations.

⁴Of concern would typically be the absolute error in the instantaneous output current for CV processing (after I-V-conversion) and total harmonic distortion (THD) for audio signal processing.

the first part of the above expression as the transconductance, accompanied by some mumbling about small input voltages. The transconductance is however strictly proportional to the tail current, which provides the function of a two-quadrant multiplier. This is typically used to set and modulate the transconductance, which is useful for instance for building VCO and VCF in analog synthesizers.

Making the tail current proportional to absolute temperature (which can be done using a ΔV_{BE} -Arrangement) gets rid of the the temperature dependence in the first part of the expression. Of course this just makes the transconductance for zero input a constant and thus does not compensate the temperature dependence for any useful circuit.

2.1.2 Input Diode Linearization

Making a better OTA involves flattening the transconductance characteristic to achieve a wider input range and of course removing the temperature dependence. Flattening the transconductance curve generally reduces the peak transconductance for any given circuit, however. Both objectives can be achieved by connecting a “differential pair” of diodes to the inputs, fed by another current source. In short, the diodes in connection with a resistive input network will provide a compression of the input voltages to the differential transistor pair which expands them into a current, while through their matching to the input transistors the temperature dependence of the inputs is also canceled.⁵

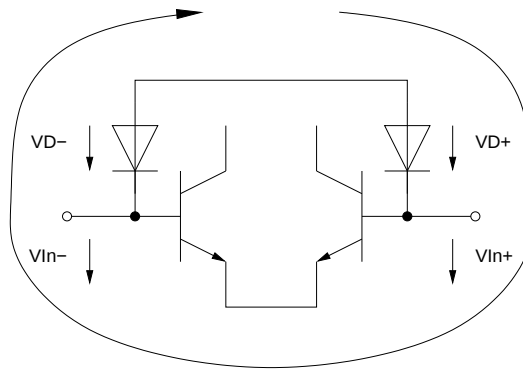


Figure 3: Principle of Input Diode Linearization

Let's look at the loop made up of the linearizing diodes⁶ and the base-emitter diodes of the differential pair. Now for the voltages in that loop⁷

$$V_{D+} + V_{I+} = V_{D-} + V_{I-} \quad (23)$$

holds with some reordering and expressing it in terms of the currents this becomes

$$V_T \ln \frac{I_+}{I_{S+}} - V_T \ln \frac{I_-}{I_{S-}} = V_T \ln \frac{I_{D-}}{I_{S,D-}} - V_T \ln \frac{I_{D+}}{I_{S,D+}}. \quad (24)$$

When all elements are matched, the saturation currents are identical and with some further simplification we get

$$\ln \frac{I_+}{I_-} = \ln \frac{I_{D-}}{I_{D+}}, \quad (25)$$

⁵This is an example of a translinear circuit, whose principle is that the input-output relations are linear even though potentially none of their internal nodes bear any linear relationship with the inputs or outputs.

⁶In an integrated circuit these diodes generally will be transistors with the base and collector shorted. Diode connected transistors have a diode characteristic that is close to ideal over a wider current range and provide better matching than simple diodes.

⁷The linearizing diodes can also be put in parallel to the base-emitter diodes (like it is done in the CA3280). The operating principle is not changed by that modification – all equations from (25) on are indeed identical, but the biasing requirements are different.

which means that the current ratios must also be equal:

$$\frac{I_+}{I_-} = \frac{I_{D-}}{I_{D+}}. \quad (26)$$

With (9), (13) and

$$I_D = I_{D-} + I_{D+} \quad (27)$$

$$I_{In} = I_{D-} - I_{D+} \quad (28)$$

(which again assumes $\beta \gg 1$) we can rewrite the currents

$$I_+ = \frac{1}{2} (I_0 + I_{Out}) \quad (29)$$

$$I_- = \frac{1}{2} (I_0 - I_{Out}) \quad (30)$$

$$I_{D+} = \frac{1}{2} (I_D - I_{In}) \quad (31)$$

$$I_{D-} = \frac{1}{2} (I_D + I_{In}) \quad (32)$$

and simplify further to

$$\frac{I_0 + I_{Out}}{I_0 - I_{Out}} = \frac{I_D + I_{In}}{I_D - I_{In}} \quad (33)$$

$$(I_0 + I_{Out})(I_D - I_{In}) = (I_D + I_{In})(I_0 - I_{Out}) \quad (34)$$

$$I_0(I_D - I_{In}) + I_{Out}(I_D - I_{In}) = I_0(I_D + I_{In}) - I_{Out}(I_D + I_{In}) \quad (35)$$

$$I_{Out}(I_D - I_{In} + I_D + I_{In}) = I_0(I_D + I_{In} - I_D + I_{In}) \quad (36)$$

and finally arrive at

$$I_{Out} = \frac{I_0}{I_D} I_{In} \quad \text{where } |I_{In}| < I_D. \quad (37)$$

Looking at the last equation we find of course that we have a current amplifier⁸ rather than a transconductance amplifier as the independent variable is now a current instead of a voltage. On the positive side, the temperature dependence of the transconductance is compensated. Of course one can use a resistor in front of each input for the voltage to current conversion, which should be dimensioned so that the maximum input current does not exceed the diode bias current at the maximum input voltage. It can also be observed that the maximum transconductance is achieved for vanishing diode biasing. While it appears at first that the transconductance can be made infinitely large, this is not the case as the input range is also zero at that point. We know of course that for vanishing diode bias current the OTA reverts to its non-linearized form.

When driven by voltage signals, resistors can be used to provide voltage to current conversion (the potential at the bases of the input transistors is almost constant). With equal input resistors the transconductance becomes

$$I_{Out} = \frac{I_0}{R_{In} I_D} V_{In} \quad \text{where } |V_{In}| < R_{In} I_D, \quad (38)$$

which also means that compensating for temperature is not as easy as it looked at first, depending on how you produce the currents for the tail and diodes. Overdriving a linearized OTA at the input more or less just clips the signal. Changes in the input potential that are effected by changes in either I_0 or I_D produce common mode inputs and are thus suppressed at the output as long as the common mode input range is not exceeded. The driving stage should be designed with careful consideration of the comparatively low and non-constant input impedance of a linearized OTA.

⁸The equation just derived may look familiar: it is the very same as for the famous Gilbert cell, where gain is the ratio of inner to outer current.

2.2 The FET OTA

An OTA could obviously also be implemented in CMOS technology by replacing the current mirrors and the input differential pair with their FET equivalents. Assuming ideal current mirrors and current sources again, the only real change is the switch to a FET differential pair.

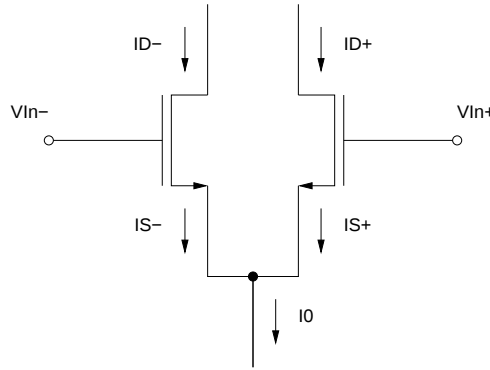


Figure 4: FET Differential Pair (with nMOS enhancement FET, biasing not shown)

Even though to the best of my knowledge there is no IC that implements a single, dual or quad FET OTA, these OTA are probably the most common analog circuit in existence - almost all continuous-time analog circuitry in modern CMOS integrated circuits is based on OTA building blocks.

2.2.1 The FET differential pair

We notice that the input resistance is infinite and hence there is no input current. Also, the source and drain currents are equal if we neglect leakage currents. The drain current of the nMOS enhancement FET with a threshold voltage of V_{th} in pinch-off regime is with some simplifying assumptions

$$I_D = I_{Dsat} \left(\frac{V_{GS}}{V_{th}} - 1 \right)^2 \quad | \quad V_{GS} \geq V_{th} \quad (39)$$

Thus,

$$I_0 = I_+ + I_- = I_{S+} + I_{S-} = I_{D+} + I_{D-} \quad (40)$$

$$V_{in} = V_{GS+} - V_{GS-} = V_{in+} - V_{in-} \quad (41)$$

and with the transforms

$$i_* = \frac{I_*}{I_{Dsat}} \quad (42)$$

$$v_* = \left(\frac{V_*}{V_{th}} - 1 \right) \geq 0 \quad (43)$$

$$\hat{i}_* = \frac{I_*}{I_0} \leq 1 \quad (44)$$

the equations

$$I_+ = I_{D+} = I_{Dsat+} \left(\frac{V_{GS+}}{V_{th}} - 1 \right)^2 \quad (45)$$

$$I_- = I_{D-} = I_{Dsat-} \left(\frac{V_{GS-}}{V_{th}} - 1 \right)^2 \quad (46)$$

can be written more simple (and hopefully more clear) as

$$i_{D+} = v_{GS+}^2 \quad (47)$$

$$i_{D-} = v_{GS-}^2 \quad (48)$$

Transformation and substitution into (41) yields under the assumption of matched transistors

$$v_{in} + 1 = \sqrt{i_+} - \sqrt{i_-} = \sqrt{i_0} \left(\sqrt{\hat{i}_+} - \sqrt{\hat{i}_-} \right). \quad (49)$$

Writing out the output current and using the identity $(\sqrt{a} - \sqrt{b})(\sqrt{a} + \sqrt{b}) = (|a| - |b|)$ together with (49) provides

$$\begin{aligned} i_{Out} &= i_+ - i_- \\ &= (v_{in} + 1) \left(\sqrt{i_+} + \sqrt{i_-} \right) \\ &= (v_{in} + 1) \sqrt{i_0} \left(\sqrt{\hat{i}_+} + \sqrt{\hat{i}_-} \right) \end{aligned} \quad (50)$$

The maximum input range is therefore $\pm \sqrt{i_0} V_{th}$, the signal is clipped beyond that point as the tail current flows through just one transistor in the differential pair and the other is closed. Recalling that

$$\hat{i}_0 = 1 = \hat{i}_+ + \hat{i}_- \quad (51)$$

we can substitute

$$1 = \sin^2 x + \cos^2 x \quad (52)$$

and use trigonometric identities to observe

$$\sqrt{\hat{i}_+} + \sqrt{\hat{i}_-} = \sin x + \cos x = \sqrt{2} \sin \left(x + \frac{\pi}{4} \right) \quad \left| \quad x \in [1, \sqrt{2}] \right. \quad (53)$$

$$\sqrt{\hat{i}_+} - \sqrt{\hat{i}_-} = \sin x - \cos x = \sqrt{2} \sin \left(x - \frac{\pi}{4} \right) \quad \left| \quad x \in [-1, 1] \right. \quad (54)$$

Through substitution of (54) into (49) we solve for

$$x = \arcsin \left(\frac{v_{in} + 1}{\sqrt{2i_0}} \right) + \frac{\pi}{4} \quad (55)$$

With (50), (53) and the identity $\sin(x + \frac{\pi}{2}) = \cos x$ we can finally express the output current as a function of input voltage

$$i_{Out} = (v_{in} + 1) \sqrt{2i_0} \cos \left(\arcsin \frac{v_{in} + 1}{\sqrt{2i_0}} \right) \quad (56)$$

$$\begin{aligned} I_{Out} &= 2I_0 \frac{V_{in}}{V_{th} \sqrt{2i_0}} \cos \left(\arcsin \frac{V_{in}}{V_{th} \sqrt{2i_0}} \right) \\ &= 2I_0 z \cos(\arcsin z) \quad \left| \quad z = \frac{V_{in}}{V_{th} \sqrt{2i_0}} \right. \\ &\approx 2I_0 z \quad \left| \quad z \ll 1 \right. \end{aligned} \quad (57)$$

which gives

$$g_m = \frac{dI_{Out}}{dV_{In}} = \frac{dI_{Out}}{dz} \frac{dz}{dV_{In}} \approx \frac{\sqrt{2I_0 I_{Dsat}}}{V_{th}} = \frac{I_{Dsat}}{V_{th}} \sqrt{2i_0} \quad (58)$$

This means that the g_m of a FET OTA is not proportional to the tail current⁹ as for the bipolar OTA, but rather to its square root. As long as one wants exponential control, it is sufficient to double the scale factor. Then each octave of transconductance translates into two octaves of tail current. The square law characteristic of the FET is not nearly as precise as the exponential characteristic of a bipolar transistor, so it is challenging to maintain tracking over many octaves.

For linear control, one could conceivably rig up a circuit with another matched FET to deliver a current proportional to the input voltage (the biasing may be somewhat tricky). Also, the input range of the FET OTA varies considerably with the transconductance, to keep linearity to one percent the input range again has to be in the Millivolt range.

⁹The same result is more laboriously arrived at via developing the full expression into a power series, differentiating that and truncating to the linear term. The quadratic term is slightly more than 1.5 times larger than that of the equivalent power series for the sech^2 part of the expression for the bipolar OTA.

3 Applications

Please refer to the datasheets for the various commercial OTA IC for a number of interesting circuits. In this (still incomplete) section the principle behind some of these circuits can hopefully be clarified somewhat.

3.1 The OTA as voltage controlled resistor

If you have a resistor that is referenced to the virtual ground of an operational amplifier, then it is easy to use an OTA to make that resistance voltage controlled. The resistor is replaced by a voltage divider to the real ground so that the divider puts out about 5 mV, which gets connected to the positive input of the OTA. The negative input is connected to ground as well, while the output of the OTA goes into the virtual ground of the operational amplifier. The apparent resistance can then be controlled by adjusting I_0 accordingly.

4 OTA IC

To facilitate easier analysis of the schematics in the various datasheets, the current mirrors are shown as ideal elements. Unfortunately most of these IC are discontinued as of May 2005.

The bipolar current mirrors come in two flavors: the most simple one is named after late (and legendary) Robert J. (Bob) Widlar and uses just two transistors. The base current of the transistors is not compensated for, so this mirror requires a relatively high transistor beta to work precisely enough. The second one, named after George Wilson, uses another transistor to compensate for the base current and improve dynamic output impedance at the expense of output voltage range. Actually there is another variant of the Wilson mirror that adds a fourth transistor that works even better at high current levels.

4.1 The SSM2040

The SSM2040 is actually a quad-section $g_m C$ filter chip, but it has the simplest OTA cell possible. There is just the differential pair, the tail current source (here with an exponential V-I converter to facilitate V/octave scaling) and a single Widlar current mirror.

This arrangement, while simple produces a significant output level shift with varying tail current and the output voltage range is not symmetric inbetween the supply rails. Therefore this structure is used only when a discrete OTA is built where the number of individual devices is of utmost importance or as building block inside an IC where the input and output levels can be well controlled.

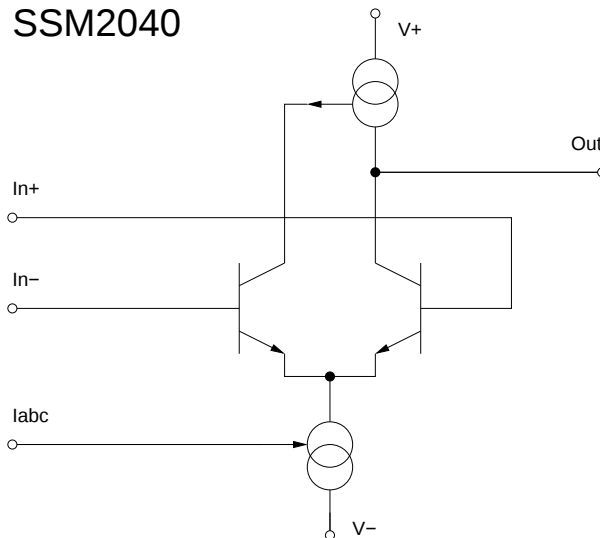


Figure 5: SSM2040

4.2 The CA3080

The CA3080 is probably the most simple standalone bipolar OTA that you can find. It consists of only the input differential pair and the current mirrors that bias the input transistors and produce the output current.

In particular, the mirror for the tail current is a simple Widlar type and emitter degeneration cannot be used as the tail current can vary widely. It is therefore important to keep the differential and current inputs at the same potential, otherwise the transconductance gets modulated by the common mode input voltage. Unfortunately the datasheet does not show the circuit for measuring the CMRR, but it appears that the common mode amplitude was low for the test and the input potentials about the same.

The output current mirrors are all Wilson type, the pnp mirrors also use a Darlington pair for the cascode transistor to get around the low beta of the pnp transistor in this process.

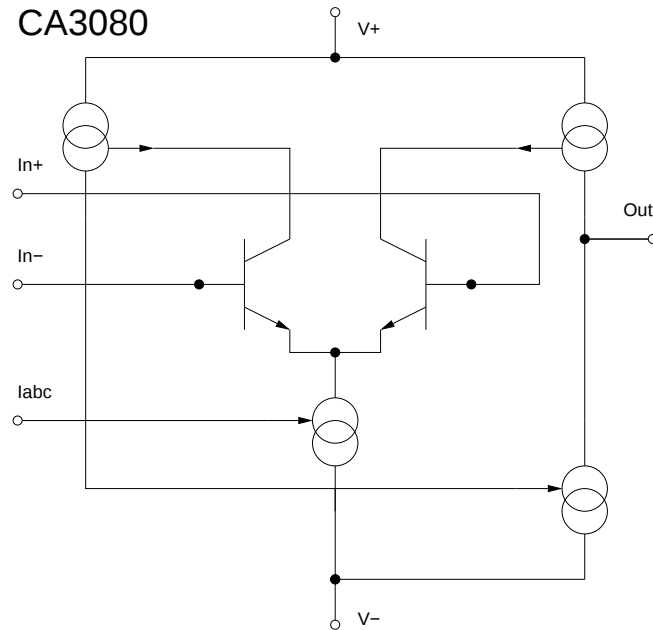


Figure 6: CA3080

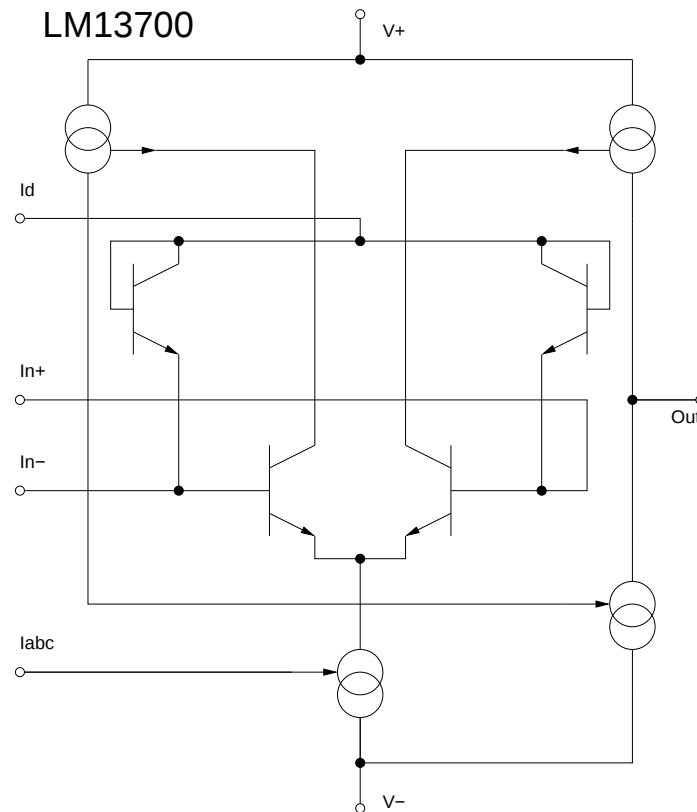


Figure 7: LM13600/LM13700 OTA section

4.3 The NE5517

The OTA section of the NE5517 is identical to the LM13700. The buffer bias is almost constant, only somewhat varied with the tail current, presumably to compensate the change in output impedance of the OTA section. The datasheet consequently claims “constant impedance buffers”. Since all figures in the datasheet are identical I originally suspected that the missing bias network in the datasheet of the LM13700 looks the same. However, as detailed below the biasing circuit of the LM13700 buffer is now known and it is different from the one shown in the NE5517 datasheet. Unless you don’t use the output buffers at all, these differences may be important in your circuit, so you should be wary of distributors replacing one type for the other.

4.4 The LM13600/LM13700

The LM13700 improves upon the CA3080 by adding linearization to the OTA inputs. While this improves the linear input range greatly, it lowers input impedance and changes the distortion properties. It uses a Wilson mirror also for the tail current. Since a Wilson mirror needs more voltage headroom, the common mode voltage range is reduced on the negative rail and the potential for the tail current input is increased in comparison with the CA3080, which may become important in certain applications.

The LM13600 and the LM13700 differ only in the way the bias current for the buffer (which is not shown here) is produced. The LM13700 uses a constant bias current according to the datasheet, while in the LM13600 the bias is a mirrored copy of the tail current. This can lead to CV feedthrough to the output when the tail current is changed rapidly. However the datasheet for the LM13700 does not show any biasing of the buffer at all, so one can only speculate how it is achieved. What is clear is that there must be some biasing and the only hint one can find of that is some mumbling about “controlled impedance buffers”. Meanwhile Don Sauer, one of the “fathers” of the LM13600 design has posted the missing details on his website [2], so it no longer remains a mystery.

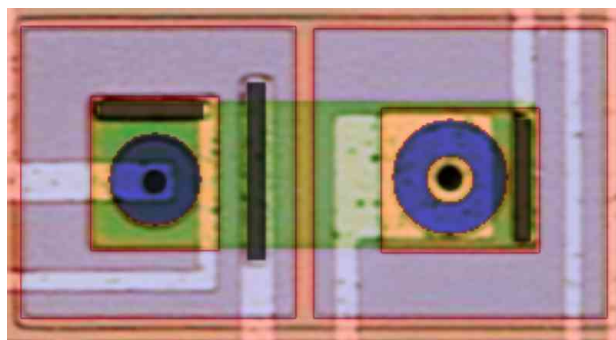


Figure 8: An npn (left) and pnp (right) transistor

Don Sauer also gave permission to use his chip micrograph of the LM13600 (he even sent a larger version of the picture). This and the fact that the transistors used in this technology are very large compared to contemporary technologies offers the opportunity to see how the circuit maps to the layout. The bond pads (where the wires to the pins will be attached) and the output transistor of the buffers are relatively easy to find and that provides us with enough information to label the pads with their pin numbers on the package. You could probably trace the entire circuit with just the schematic and this much information in hand (there are a few surprises, but you can work around those), but it is a lot easier if you know one or two things about the technology used.

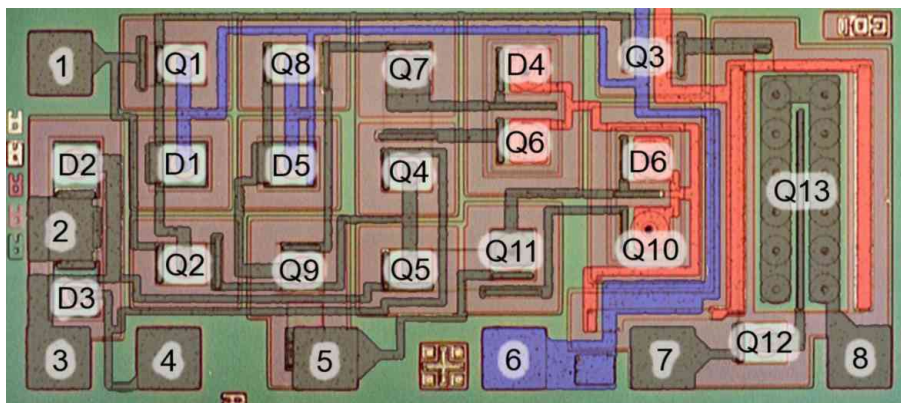


Figure 9: One half of the LM13600

The LM13600 is done in a planar bipolar technology with junction isolation and a single level of metallization (the “wires”). This type of technology starts with a lightly p-doped silicon substrate¹⁰. Then an n-type diffusion is created in certain places¹¹ and the surface is overgrown with an n-type epitaxial silicon, creating buried regions of high n-type doping that are called n-buried layer. These regions are isolated by a p-type diffusion with high doping from the surface through the n-epitaxial layer into the substrate, this forms a pn-junction and hence the name junction isolation. If the same p-type diffusion¹² is done inside a buried n-region, it will not reach the substrate and is therefore completely isolated by the n-tub it sits in. Another n-type diffusion, yet more highly doped but again more shallow is used to either make contact to the n-tub¹³ (in places where there is no second p-diffusion) or create another n-region inside the isolated p-region. Once this is all done, you make contacts to a) the n-tub and call it the collector, b) the isolated p-region and call it the base and the n-region inside the p-region and call it the emitter – and there is your npn transistor. In fact you have created many npn transistors, since each isolated region will have

¹⁰A much more detailed step-by-step explanation with cross-sections is available in chapter 1 of [3].

¹¹This involves opening a window in the oxide on top of the silicon through which the dopant can enter.

¹²In some technologies the isolation diffusion is done separately.

¹³To reduce the resistance from this contact to the n-buried layer, in some technologies another n-type diffusion called n-sinker will be employed for this function.

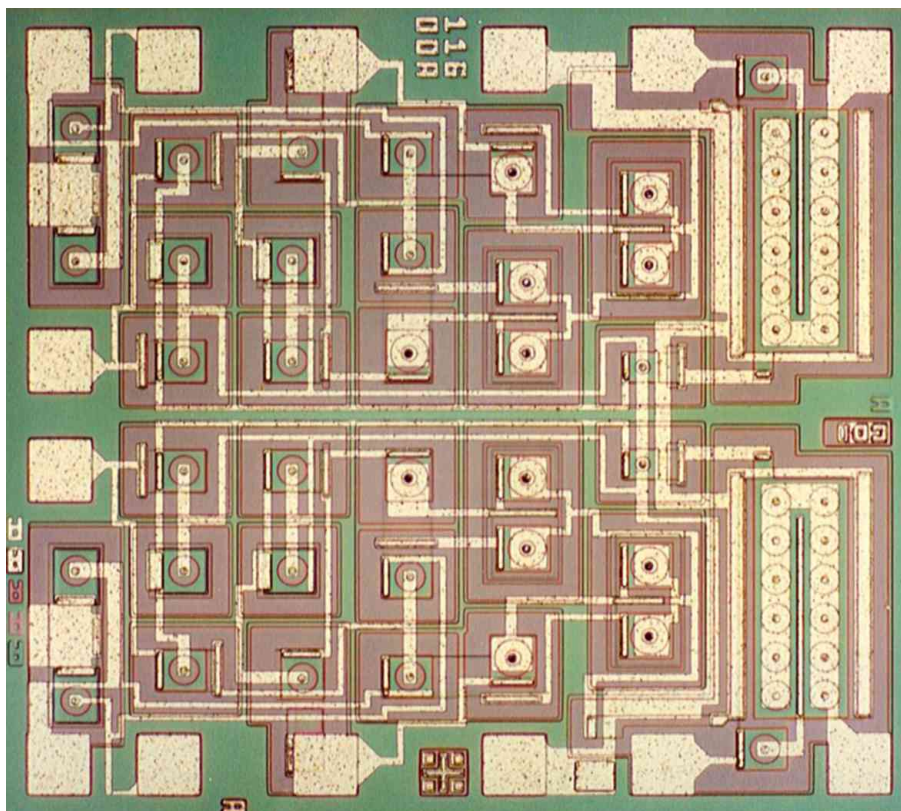


Figure 10: LM13600 chip micrograph (with permission of Don Sauer)

one. Since you have many transistors and not just one, to maintain isolation between them you must also make a contact outside them to the p-isolation or substrate and put this to the most negative potential anywhere in the circuit to keep the isolation junctions in reverse bias. Alas, there's no "real" pnp transistor (that would require to reverse all the doping polarities) – but if you place two isolated p-type regions very close to each other, they will act as a pnp transistor. This is called a lateral pnp transistor, but these aren't as good as the npn; in particular they are far slower and have less current gain. Since all diffusions are accompanied by etching of windows into the oxide already present as well as more oxidation to drive in the dopant and the amount of doping varies the oxide thickness too, each region of the chip will have a different oxide thickness depending on which doping it has received. Due to interference of incident and reflected light, thin transparent films have a color depending on their thickness, so each of these regions will have a different appearance¹⁴.

So, in the chip micrograph the greenish regions are p-type, the brownish frames are npn collectors or pnp bases, the greenish stuff inside those frames is the npn base or the pnp collector and the brownish circles are the npn bases. If you look very closely you'll see that for the pnp transistors these aren't circles, but donuts. Inside you'd find another greenish p-region acting as the pnp emitter, but you cannot see this because there's always metal on top. You'll also note a lot of things that look like depressions and they are – this is where silicon dioxide has been etched away to either let a diffusion take place or to make contact to the underlying silicon. In figure 8 an npn and pnp transistor are shown side-by-side (they are Q4 and Q7 in the datasheet). The window for the buried n-diffusion is tinted green, p-diffusion light red, n-epi light blue and n-emitter as dark blue, while contact windows are dark grey. This pair of transistors (as well as Q5 and Q11) uses a buried connection for connecting the collector of the npn to the base of the pnp, which is why they share the buried n-diffusion. In figure 9 the first half of the chip is shown and the wiring has been colored: red tint for positive supply, blue tint for negative supply and grey for everything else. Take note of the substrate contact next to the negative supply pin and the buried wiring for the

¹⁴You can find tables which detail the color vs. thickness in [2].

positive supply that crosses under the negative supply as well as the connection of the bias current source for the Darlington buffer¹⁵. Finally figure 10 shows the complete chip in all it's glory. A few things to note: while the two OTA sections are almost mirror copies of each other, there is a slight asymmetry in the power connection. Also the label on the chip actually says "11600A" instead of "13600" as there were several grades of this chip tested to different specifications, but produced from the same die. The set of letters between pad 1 and 2 are the revision letters for all the five layers that are needed in this process (all at their first revision). There is an alignment mark between pad 5 and 6 and a resolution or measurement target between the buffer output transistors (CD probably stands for "critical dimension" and the barely visible structure next to it would then be where the resolution gets checked).

¹⁵This transistor incidentally is smaller than the other npn transistors and hence the current mirror ratio is less than one - the datasheet neglects to mention that. With the information given in [2] the mirror ratio computes to $1 : 4\frac{1}{6}$.

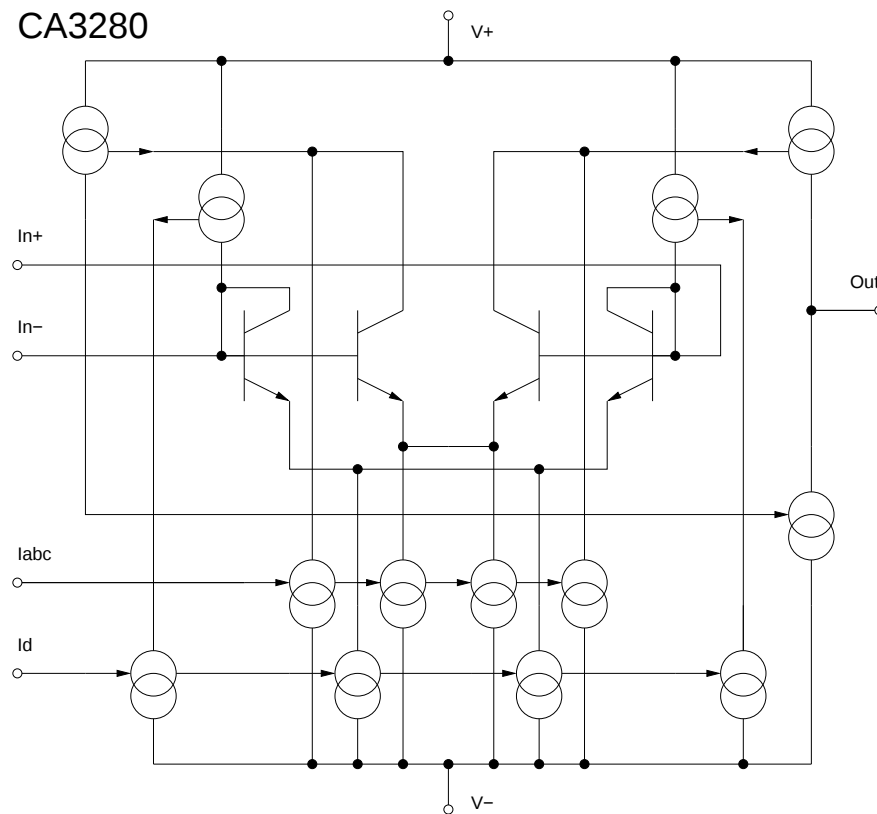


Figure 11: CA3280

4.5 The CA3280

The CA3280 also adds linearization, but in a slightly more complicated way than the LM13700 that maximizes the common mode input range when the linearization diodes are used. It also uses a Wilson mirror for the tail current. The output mirrors are not shown in detail on the datasheet. While it's safe to assume they're Wilson types, it is hard to know exactly if they use Darlington pairs. The relatively wide bandwidth leads me to assume that they're plain pnp transistors like the LM13700, however.

4.6 The “Diamond Transistor” OPA660

The OPA660 has a different tack on the OTA theme. The negative input is a low impedance terminal, in effect becoming both an input and a (differential) output. Burr-Brown touted the device as an “ideal” transistor.

4.7 The MAX435 / MAX436

The MAX435 is an OTA with differential outputs and a gain setting network, the MAX436 drops the differential output and has a different internal gain factor. The maximum output current is controlled by a set current like the conventional OTA. It is unclear whether these OTA could be used without the gain setting impedance and if the transconductance would then be controllable through the set current.

5 Appendix

5.1 The hyperbolic half-argument conundrum

Deriving a closed form expression for the transconductance requires two steps: step one is to recognize that there's a single hyperbolic function to uncover and step two is to see that the argument of that function ought to be half of what you've been dealing with up to now. Certainly both of these steps can be introduced at various points along the derivation, thus yielding different versions. All these versions have in common that at some point you'll need add a "nutritious zero" or multiply a "nutritious one", which can often be motivated by symmetry considerations.

The trigger for step one is that whenever you see something that looks like it could be brought into the form $e^x \pm e^{-x}$ (this can be quite hard to see, even though it comes up surprisingly often), you can save yourself lots of work by re-writing your equation in terms of hyperbolic functions and then working on these using a set of convenient equivalencies and relations between hyperbolic functions, which can be looked up in any decent book on higher mathematics. Just like the normal trigonometric functions the hyperbolic trigonometric functions have special relations to each other when the argument is multiplied or divided by integers, these are especially useful for double or half the argument. Looking up these equivalencies at the right time saves you the bother of actually carrying out a large part of the otherwise protracted derivations.

5.2 Direct introduction of the half-argument

If you knew in advance that you need the half-argument, this alternative (and a bit shorter) derivation of (20) (provided by Ian Fritz) results:

$$\begin{aligned}
 I_{\text{Out}} &= I_0 \left(\frac{\exp \frac{V_{\text{In}}}{2V_T}}{\exp \frac{V_{\text{In}}}{2V_T} + \exp -\frac{V_{\text{In}}}{2V_T}} - \frac{\exp -\frac{V_{\text{In}}}{2V_T}}{\exp -\frac{V_{\text{In}}}{2V_T} + \exp \frac{V_{\text{In}}}{2V_T}} \right) \\
 &= I_0 \left(\frac{\exp \frac{V_{\text{In}}}{2V_T}}{\exp \frac{V_{\text{In}}}{2V_T} + \exp \frac{V_{\text{In}}}{2V_T} \exp -\frac{2V_{\text{In}}}{2V_T}} - \frac{\exp -\frac{V_{\text{In}}}{2V_T}}{\exp -\frac{V_{\text{In}}}{2V_T} + \exp \frac{V_{\text{In}}}{2V_T} \exp \frac{2V_{\text{In}}}{2V_T}} \right) \\
 &= I_0 \frac{\exp \frac{V_{\text{In}}}{2V_T} - \exp -\frac{V_{\text{In}}}{2V_T}}{\exp \frac{V_{\text{In}}}{2V_T} + \exp -\frac{V_{\text{In}}}{2V_T}} \\
 I_{\text{Out}} &= I_0 \tanh \frac{V_{\text{In}}}{2V_T}.
 \end{aligned}$$

The different signs of the multiplicands can be motivated by symmetry considerations.

5.3 Substituting One

The second alternative derivation comes from the lecture notes on analog multiplication by Paul Junor. It starts off with a slightly different reduction of the common denominator, while the introduction of the half-argument can again be motivated by symmetry considerations.

$$I_{\text{Out}} = I_0 \frac{\left(1 + \exp \frac{V_{\text{In}}}{V_T}\right) - \left(1 + \exp -\frac{V_{\text{In}}}{V_T}\right)}{\left(1 + \exp \frac{V_{\text{In}}}{V_T}\right) + \left(1 + \exp -\frac{V_{\text{In}}}{V_T}\right)}$$

Substituting the identity

$$1 = \exp -\frac{x}{2} \exp \frac{x}{2}$$

gives

$$I_{\text{Out}} = I_0 \frac{\left(\exp -\frac{V_{\text{In}}}{2V_T} \exp \frac{V_{\text{In}}}{2V_T} + \exp \frac{2V_{\text{In}}}{2V_T}\right) - \left(\exp -\frac{V_{\text{In}}}{2V_T} \exp \frac{V_{\text{In}}}{2V_T} + \exp -\frac{2V_{\text{In}}}{2V_T}\right)}{\left(\exp -\frac{V_{\text{In}}}{2V_T} \exp \frac{V_{\text{In}}}{2V_T} + \exp \frac{2V_{\text{In}}}{2V_T}\right) + \left(\exp -\frac{V_{\text{In}}}{2V_T} \exp \frac{V_{\text{In}}}{2V_T} + \exp -\frac{2V_{\text{In}}}{2V_T}\right)},$$

which enables the following extraction

$$= I_0 \frac{\exp \frac{V_{In}}{2V_T} \left[\exp -\frac{V_{In}}{2V_T} + \exp \frac{V_{In}}{2V_T} \right] - \exp -\frac{V_{In}}{2V_T} \left[\exp \frac{V_{In}}{2V_T} + \exp -\frac{V_{In}}{2V_T} \right]}{\exp \frac{V_{In}}{2V_T} \left[\exp \frac{V_{In}}{2V_T} + \exp -\frac{V_{In}}{2V_T} \right] + \exp -\frac{V_{In}}{2V_T} \left[\exp -\frac{V_{In}}{2V_T} + \exp \frac{V_{In}}{2V_T} \right]},$$

dropping the terms in brackets gives

$$= I_0 \frac{\exp \frac{V_{In}}{2V_T} - \exp -\frac{V_{In}}{2V_T}}{\exp \frac{V_{In}}{2V_T} + \exp -\frac{V_{In}}{2V_T}},$$

which interpreted as hyperbolic function reads

$$= I_0 \frac{\sinh \frac{V_{In}}{2V_T}}{\cosh \frac{V_{In}}{2V_T}} \\ I_{Out} = I_0 \tanh \frac{V_{In}}{2V_T}.$$

5.4 Yet another go

Tim Stinchcombe had yet another proposal (borrowed in part from [4]), starting with developing an expression for the individual collector currents via (3) and with (9) – or we can simply take it out from the first part of (16):

$$I_+ = \frac{I_0}{1 + \exp -\frac{V_{In}}{V_T}},$$

motivated by the fact that with no signal each branch of the differential pair sees half the tail current we pull this out as the scaling factor

$$= \frac{I_0}{2} \frac{2}{1 + \exp -\frac{V_{In}}{V_T}}$$

and substitute the boring 2 with something more creative

$$= \frac{I_0}{2} \frac{\left(1 + \exp -\frac{V_{In}}{V_T}\right) + \left(1 - \exp -\frac{V_{In}}{V_T}\right)}{1 - \exp -\frac{V_{In}}{V_T}} \\ = \frac{I_0}{2} \left(1 + \frac{1 - \exp -\frac{V_{In}}{V_T}}{1 + \exp -\frac{V_{In}}{V_T}}\right)$$

and via one of the addition theorems we find

$$I_+ = \frac{I_0}{2} \left(1 + \tanh \frac{V_{In}}{2V_T}\right)$$

and due to symmetry

$$I_- = \frac{I_0}{2} \left(1 - \tanh \frac{V_{In}}{2V_T}\right)$$

and finally we arrive via (13) at

$$I_{Out} = \frac{I_0}{2} \left(1 + \tanh \frac{V_{In}}{2V_T}\right) - \frac{I_0}{2} \left(1 - \tanh \frac{V_{In}}{2V_T}\right) \\ I_{Out} = I_0 \tanh \frac{V_{In}}{2V_T}$$

References

- [1] I. N. Bronstein and K. A. Semendjajew, *Taschenbuch der Mathematik*. BSB Teubner, 24th ed., 1989.
- [2] D. Sauer, *The LM13600 Story*, May 2003. <http://www.idea2ic.com>.
- [3] H. Camenzind, *Designing Analog Chips*. Virtualbookworm.com Publishers, 2005. <http://www.designinganalogchips.com>.
- [4] T. H. Wilmshurst, *Analog Circuit Techniques with Digital Interfacing*. Newnes, 2001.
- [5] Intersil Americas Inc., *CA3080, CA3080A: 2MHz, Operational Transconductance Amplifier (OTA)*, Aug. 2004. FN475.6.
- [6] H. Wittlinger, *Applications of the CA3080 High-Performance Operational Transconductance Amplifier*. Intersil Americas Inc., May 2002. AN6668.2.
- [7] Intersil Americas Inc., *CA3280(A) Dual, 9MHz Operational Transconductance Amplifier (OTA)*, May 2002. FN1174.6.
- [8] Intersil Americas Inc., *An IC Operational Transconductance Amplifier (OTA) with Power Capability*, Oct. 2000. AN6077.3.
- [9] National Semiconductor Corporation, *LM13600 Dual Operational Transconductance Amplifiers with Linearizing Diodes and Buffers*, May 1998. DS007980.
- [10] National Semiconductor Corporation, *LM13700 Dual Operational Transconductance Amplifiers with Linearizing Diodes and Buffers*, Aug. 2000. DS007981.
- [11] Philips Semiconductors, *NE5517(A)/AU5517 Dual Operational Transconductance Amplifier*, Dec. 2000. Doc.No. 9397 750 10796.
- [12] Burr-Brown Corporation / Texas Instruments Incorporated, *Wide Bandwidth Operational Transconductance Amplifier and Buffer*, 1990. PDS-1072F / SBOS007.
- [13] Burr-Brown Corporation / Texas Instruments Incorporated, *Dual, Wide Bandwidth Operational Transconductance Amplifier*, 1991. PDS-1129F / SBOS011.
- [14] Burr-Brown Corporation / Texas Instruments Incorporated, *Macro Models for RF Op Amps are a powerful Design Tool*, 1993. AN-189 / SBOA074.
- [15] Maxim Integrated Products, *Wideband Transconductance Amplifiers*, 1993. 19-0042.
- [16] E. M. Zumchak, *A Short Discussion of the Operational Transconductance Amplifier (OTA)*, Feb. 1999. <http://www.uni-bonn.de/ uzs159/ota3080.html>.
- [17] J. Patchell, *Secrets of OTAs*, May 2003. WWW.
- [18] R. L. Geiger and E. Sánchez-Sinencio, “Active filter design using operational transconductance amplifiers: A tutorial,” *IEEE Circuits and Devices Magazine*, vol. 1, pp. 20–23, Mar. 1985.
- [19] B. Gilbert, “The multi-tanh principle: A tutorial overview,” *IEEE Journal of Solid-State Circuits and Systems*, vol. 33, pp. 2–17, Jan. 1998.
- [20] U. Tietze and C. Schenk, *Halbleiter-Schaltungstechnik*. Springer, 12th ed., 2002.
- [21] M. Seifart, *Analoge Schaltungen*. Verlag Technik Berlin, 3rd ed., 1989.

Acknowledgements

Thanks go to Ian Fritz, Paul Junor, Ryan Williams, Tim Stinchcombe and Tim Davis for discussions and spotting some typos and errors.

A big thank you to Don Sauer for clarifying that nagging buffer biasing question and letting me use the LM13600 chip micrograph.