Simple Analog Circuits – A Primer

Sam Roweis

1 Basics

1.1 Current vs. Voltage Mode

The whole idea of building circuits is to take some input data and do some useful computation to give some output data. But how is this data presented to the circuit and collected from it afterwards ? We typically represent information in circuits in one of two ways: we either use the voltage at some node in the circuit or we use the current flowing in some wire. These two ways of representing information can be thought of as two *data types* with analogy to computer science. Current is easy to understand: we just make an imaginary cut through a wire with a plane and count how many positive charges flow across the plane per second. We call that number the current in that wire. It can be positive or negative depending on which direction the *net* flux of charges across the plane was compared to the direction we definied the current in the wire.

Voltage is a little trickier: the voltage of a node *with respect to some reference node* is the energy it takes us to move a single positive charge from the reference node to the node itself. It can be positive or negative depending on whether or not is costs us energy or we get some energy back in moving the charge. Sometimes people just talk about the voltage *at* a node; they generally mean with respect to one of the power supplies, usually the lower one which is often ground.

As it turns out, certain kinds of computations are easier to perform when the input or output signals are either currents or voltages. Circuits that take there inputs as voltages and produce voltage outputs are the most common, these are known as volage mode circuits. Similarly, there are current mode and mixed mode circuits. We will first look at some basic computations in pure voltage and current modes and the circuits that perform them and then move on to more complicated but more useful circuits in all modes.

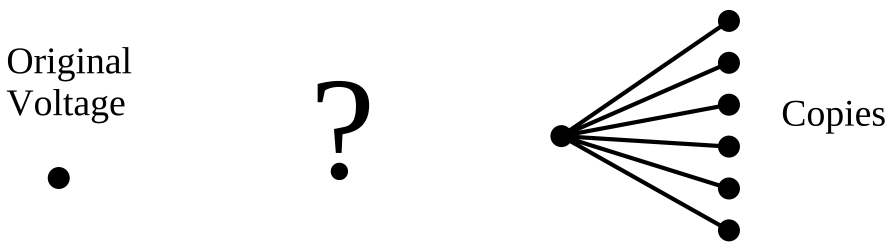
1.2 Linearity and Lumped Elements

Now is a good time to remind ourselves about a few of the basic assumptions that are inherent in the standard way we analyze circuits. The most critical assumption that we make is the use of the so called *lumped element* model for circuits. This means we pretend that all the changes to all the voltage and current signals in our circuit occur at magic localized spots where there is "an element" and that *nothing* changes anywhere else no matter how long or oddly shaped the connections between these elements are or how they are arranged around each other. Practically what this means is that when we do our analysis, we assume that the voltage and current do not change (spatially) along a wire.

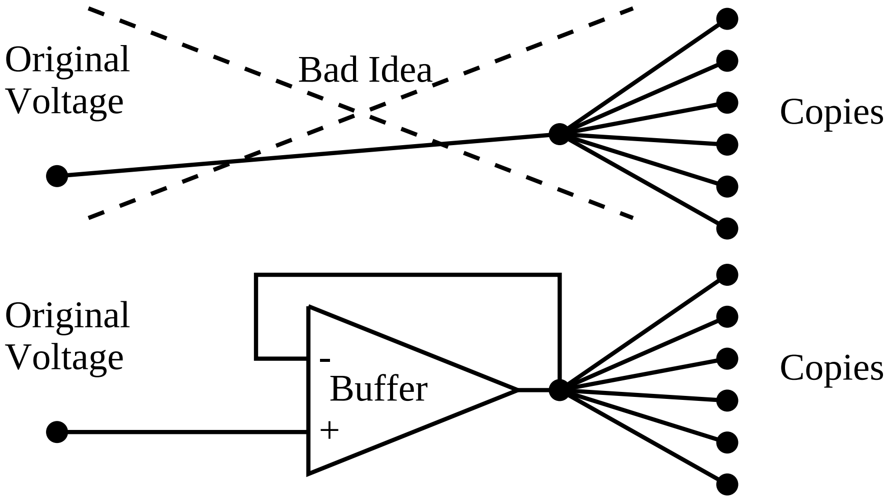
The other important assumption that is typically made is that *passive* components in circuit behave linearly. This means that resistances and capacitances (and inductances if we could make them on chips which we cannot) are assumed to follow the nice simple models you all have seen in college.

Neither of these assumptions is exactly true. On a real chip, "wires" are made by laying down strips of real materials like metal or polysilicon which have finite nonzero resistances. So the voltage drops slightly as you move along a wire in the direction of current flow. Also, because the substrate of the chips are semiconductors there is always a certain amount of *capacitive coupling* between any two places due to *parasistic capacitances*. However, we typically ignore these effects in our design analysis and deal with them as secondary effects later if they are prolematical.

1.3 Identity, Negation, Sum, Difference

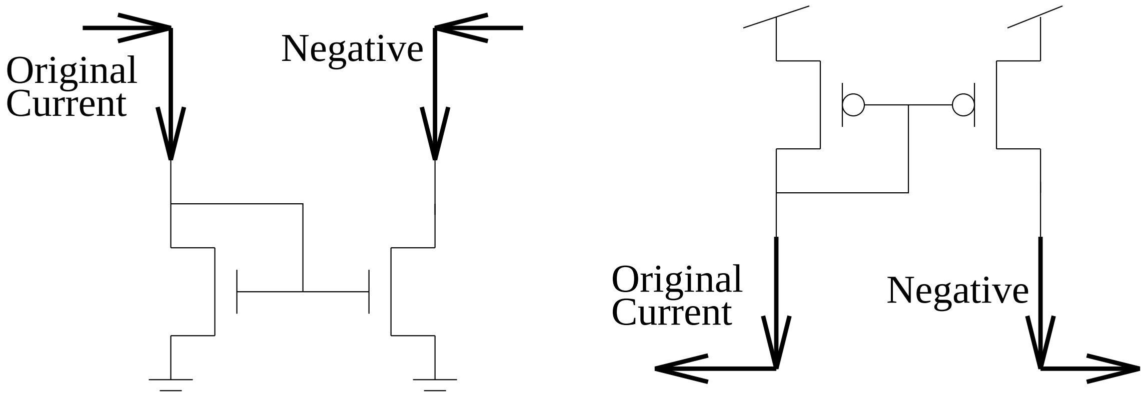


To make a copy of a voltage signal at one node on another node, in theory all we need to do is hook up the second node to the first with a wire. In practice this is usually an unwise thing to do because the “copy node” may be connected to other things which will then become connected to the original node and thus might mess up its voltage. So what we do is we feed the first voltage to a “buffer” which is a strong amplifier with a gain of one. What’s that you say, an amplifier with a gain of one ? Some amplifier! Indeed the whole point of a buffer is to sample a voltage at a sensitive node using by connecting it to a “nice” input that does not mess things up and then to output that same sampled voltage to one or more other “nasty” nodes which normally would mess up the voltage they were connected to. However, the enery we put into the amplifier is used to drive the voltages at these output nodes to the correct value.

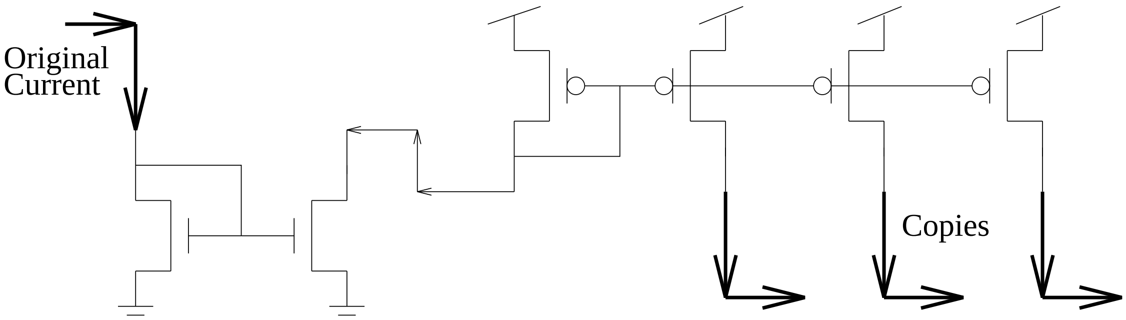


Negating a voltage is either easy or senseless depending on what kind of voltage you are working with. For a single node voltage defined with respect to one of the supplies, negation has no real meaning since we typically do not operate with voltages outside the power rails. For a voltage defined between two nodes, as in the voltage *from* node A *to* node B, negation is easy: hook node A to node B' and node B to node A' and then you have the negative voltage from A' to B'.

Copying a current is a little tricky, so we will first explain how to negate a current. We use the principle that if the drain voltage of a transistor is far enough from the source, then the current depends only on the gate to source voltage. So we feed the current we want to copy into a diode connected transistor and then we connect its gate and source to the gate and source of another transistor. This other transistor now has the same gate to source voltage as the first and so will draw the same current. This arrangement is called a *current mirror*. It can be used to produce one or more negated copies of a current. However there is a subtle point to be made here. If we have a wire which is a current *source* then we hook it up to a mirror made from two native type transistors and we get a wire that *sinks* the same current. But if the input wire itself begins to *sink* current then the output cannot *source* the negative current. To achive this we must we a mirror made from well type transistors. (This one-sided behaviour of current mirrors is important however, in the construction of rectification circuits so don’t completely ignore it.)



We can now easily copy a current signal by using one mirror to negate it and then a second mirrror (or set of mirrors) of the opposite type to negate again and recover a copy of the original current.



Adding and subtracting differential voltages is a simple matter. To add a voltage from node A to B and a voltage from node C to D simply connect node B to node C and then the sum of the voltages appears from node A to node D. Subtraction is just the addition of the negative: connect node B to node D or node A to node C and the difference $AB - CD$ or $CD - AB$ will appear across AD.

Adding and subtracting single ended voltages is a well defined operation as long as the resulting sum or difference still lies within the power rails, but is very difficult. Typically one is forced to use an operational amplifier circuit consisting of tens of transistors. For now we will assume that this computation is better done in current mode or with fully differential voltages.

Summing currents is truly a simple operation. We simply connect together all the wires on which the currents to be summed are running. The resulting wire carries the sum of the converging currents. This works sometimes but not all the time. For example, say we have two wires. If one sources a current I_1 and the other sinks a current I_2 then if we connect the wires we get a wire that sources $I_1 - I_2$ (equivalently it sinks $I_2 - I_1$). But if we want access to the sum $I_1 + I_2$ then we must mirror one of the currents before we hook them up together. Similarly, if the two wires both source or sink currents then the sum is easy but the difference requires one mirror.

2 Differential Pair

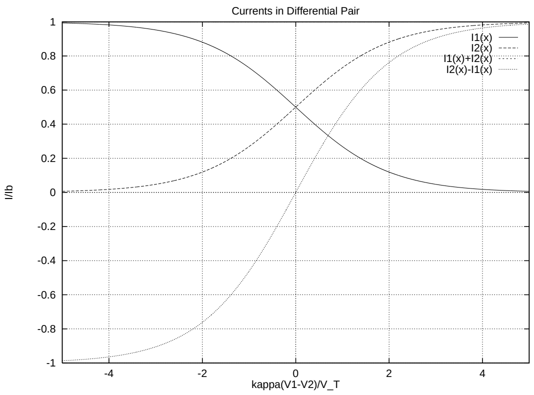
The differential pair is a simple three transistor circuit which is at the heart of many of the more complex analog circuits we will examine. Figure 5.1 shows the native transistor circuit for a differential pair. It can be easily understood if we keep in mind what it is supposed to do: *the differential pair uses the difference between two input voltages to partition a fixed current between two branches.* How does the circuit work ? First we set the bias voltage V_b which controls the total current in the circuit I_b. The sum of the currents in the two input transistors $I_1 + I + 2$ is always equal to the current in the bias transistor. Now as we change the input voltages V_1 and V_2, the node V common to all three transistors moves up and down until its voltage is exactly right so that the sum of the currents in the two input transistors is the right total.

We will now do a more quantitative analysis of the behaviour of the circuit in the subthreshold bias region. The above threshold analysis is quite complicated and the reader is referred to Sanjoy Mahajan's note for a detailed description and elegant solution technique. We know that in subthreshold, the current in each transistor is exponentially related to its source-gate voltage, hence:

$$I_1 = I_0 e^{\kappa V_1 - V} \qquad I_2 = I_0 e^{\kappa V_2 - V}$$

$$\Rightarrow e^{-V} = \left(\frac{I_b}{I_0}\right) \frac{1}{e^{\kappa V_1} + e^{\kappa V_2}}$$

What do the currents in the differential pair look like compared to the bias current ?



This is really true, check out the data that Amit and I took in figure D1.

If we take the difference between these two currents, we get an amazing result:

$$I_1 - I_2 = tanh \frac{\kappa(V_1 - V_2)}{2V_T}$$

What is so amazing here ? Well, we have obtained an exact rendition of one of the most popular soft nonlinearities used in neural computation models [1] directly from the device physics of this simple circuit. This is the kind of thing that gets people really excited about analog VLSI. In a digital simulation you would have to work out a hyperbolic tangent of some argument by grinding through some clever fast convergin power series or using some lookup table tricks or whatever. But in a circuit it gets computed for free, and damn fast.

[1] Remember all those tanh's in Hopfield's class?

3 Transconductance Amplifiers

3.1 Basics

Since we have noticed that the difference of the branch currents in the differential pair is so exciting, why don't we add a bit of extra circuitry to compute this difference ? No problem, look up our handy-dandy current mirrors from the sections above and off we go. Stick the current I_1 into a mirror to reverse it and then add the reversed current (called I_4 in the diagram) to I_2 by connecting the two wires into one. Let us call the difference $I_1 - I_2$ we just computed a new current I_{out}. Now we have a circuit that computes a hyperbolic tangent. It is called a transconductance amplifier, and its schematic is shown in figure 5.3. The ideal output characteristic is shown in figure 5.5. And it really works: see more data we took in figure D2.

The only slight trick is that the input variable is a difference of two voltages and the output variable is a current. The first problem is really an advantage. We often want to work with the differences of things to compare the results from two other computations in the circuit, or compare inputs from some detector or whatever. And when we just want to take the simple tanh of one variable, all we have to do is ground one of the inputs to set it to zero and then the output current will just be the tanh of the single variable input. The second problem is something we cannot do anything about. This circuit's output data type is current, and we have to find some way of working in current mode and not voltage mode until the next chance we get to convert back. But it is important to remember that *the transconductance amplifier converts from voltage type input to current type output.*

We were very excited about the fact that the tanh we a nonlinearity, but us circuit geeks also often get very excited about the fact that the tanh is very linear for small arguments. In the case of the simple transconductance amplifier, the output current varies linearly over a differential input range of a few thermal voltages. More formalyy, if you extrapolate the linear part of the output characteristic until it hits the output asymptotes, it will intersect at points separated by $4V_T/kappa$ volts. This gives a "linear range" in each direction of $2V_T/\kappa$ or about 70mV. The slope of this linear range is know as *the transconductance* of the amplifier since it describes the conductance which transfers input to output. It denoted with the letter G and its value is given by the expression:

$$G = \frac{\partial I_{out}}{\partial V_{in}} = \frac{I_b}{2V_T/\kappa}$$

which makes sense since we would expect to have to deliver more bias current in order to be able to get more gain.

3.2 Improvements and Subtle Points

This covers the very basics of transconductance amplifiers in subthreshold. There are a lot more issues involved and the entire above threshold regime of operation to analyze, but I will not over them fully here. Instead, I will just give you the punchline for some things that are important if you want to think more about this, and refer you to the references if you want a detailed analysis.

It would be nice if the output current were independent of the voltage at the output node. This is roughly true, with a few small caveats. At very high output voltages (more than V_{DD}, one of the transistors in the current mirror begins pumping current the wrong way, and so the output current is sucked into the power rail and drops off to zero.

At output voltages lower than the minimum of the inputs minus the bias voltage another complicated problem occurs and the output current skyrockets to large values. This last effect is known as the V_{min} problem. These two effects are shown in figure 5.6. Introducing two extra current mirrors can correct the V_{min} problem, this new circuit is called a *widerange transamp* and its circuit is shown in figure 5.11.

The wide range transconductance amplifier has the advantage that the output can swing from rail to rail whereas the regular amplifier is limited by the V_{min} problem at low input voltages. The wide range circuit has much lower output conductance and about an order of magnitude more gain since we can make the output transistors with very long channels. It pays for these advantage by taking a much larger chip area and dissipating more power. Since there are more devices present, the offsets will tend to accumulate, making the average offset for such devices worse than for the regular transconductance amplifier.

Also in both kinds of transamp, because of the Early effect of the mirror transistor, there is a finite output conductance effect on the output current which causes it to drop slightly as V_{out} is increased even inside the allowed range of V_{min} to V_{DD}.

It is also possible to use the transconductance amplifier as a voltage-in, voltage-out circuit if we connect the output node to some place that does not draw any current (such as the gate of a transistor). This forces I_{out}, (the variable we previously measured) to be always zero. Now instead, we look at the voltage that the output node takes on by itself (the variable we previously set at our convenience as long as it was above V_{min} and below V_{DD}. A difference in the input voltages normally causes a difference in the output currents I_1 and I_2 which we measure. However, we are now not drawing away any current to make I_{out} nonzero and so the difference needs to be absorbed by one of the three mechanisms that changed the output current in figure 5.6: the early effect, the V_{min} problem, or the V_{DD} shunting. For V_1 just a bit less than V_2, the output voltage goes a bit below V_{min} and the difference current $I_1 - I_2$ is supplied by the V_{min} effect. For V_1 just a bit smaller than V_2, the output voltage goes very near V_{DD} and difference current $I_2 - I_1$ is shunted to the power supply. For a very small range where V_1 and V_2 are nearly the same, the output voltage moves from V_{min} to V_{DD} because of the Early effect. The ideal curves for this effect are given in figure 5.8. Data for this effect that we took is in figures D3 (regular transamp) and D4 (widerange transamp without V_{min} problem). You can think of these curves as versions of figure 5.6 turned on its side.

The final thing to mention is a pointer to some notes and a paper by Rahul Sarpeshkar which discusses some tricks for extending the linear range of the transamp. Instead of a hundred millivolts or so, Rahul manages to get a linear range of about one volt in subthreshold using some neat circuit tricks called backgate input, sourcedegeneration and gate degeneration. Amit and I have some data from these beasts showing their wide linear ranges in figres D5 and D6.

4 Similarity Measures

Computing the similarlity between two quantities is an important and nonlinear operation. Multiplication is the crudest cut at such a similarity measure: the product of two numbers (normalized by their sum) is maximized when they are equal. Carver came up with a knifty circuit called a *current correlator* which computes exactly this sum normalized product. The circuit is shown in figure 6.1. Its output current in subthreshold is given by:

$$I = I_0 e^{-V_s} \frac{e^{V_1} e^{V_2}}{e^{V_1} + e^{V_2}}$$

$$I = \frac{I_1 I_2}{I_1 + I_2}$$

The total current $I_1 + I_2 + I$ flows into ground through the node V_S. So we have a circuit that computes a nice similarity measure between two currents.

What if we want a measure of similarity between two voltages. Well, we could use our trusty differential pair circuit to convert two voltage inputs into two currents and then correlate these as shown in figure 6.2. Note that the correlator in this circuit is built upside down. This circuit is called the *bump circuit*, and its output current is given by:

$$I_{out} = \frac{I_b}{4cosh^2\left(\frac{\kappa(V_1 - V_2)}{2V_T}\right)}$$

Some of our data of the individual currents in this circuit as well as the output is in figures D7 and D8. We can see that it does a pretty good job at detecting when V_1 and V_2 are similar.

You might ask, could we just hook out voltages up directly to the gates of the transistors S in figure 6.1 ? The answer, is yes, if we do a small trick first. If we fixed V_c then simply increasing both voltages V_1 and V_2 would increase the output current even if we did not make them any closer to each other. We have to make sure that the node V_c moves up and down with on of V_1 or V_2. By introducing extra transistors as in figure 6.3 we force the node V_c to move with the higher of V_1 and V_2 since all three currents must sum to the bias current. Now when the voltages are the same, there is current in all three legs of the circuit. When the differential voltage increases, one of the transistors Q_1 or Q_2 shuts off; whichever is connected to the lower voltage since the node V_c tracks the higher voltage. Since they are in series, this shuts off I_{mid}. Increasing or decreasing both voltages together by the same amount will not change I_{mid} since the node V_c will track. This circuit is called the *bump-antibump circuit* since I_{mid} measures the similarity between the voltages and $I_1 + I_2$ measures their dissimilarity since $I_1 + I_2 = I_b - I_{mid}$. The output current is given by:

$$I_{mid} = \frac{I_b}{1 + 4cosh^2\left(\frac{\kappa(V_1 - V_2)}{2V_T}\right)}$$

which is almost the same as before. Some data is shown in figures D9 and D10. The current I_{mid} bumps when the voltages are similar and the sum $I_1 + I_2$ dips.

5 References

1. Carver Mead, *Analog VLSI and Neural Systems*
2. Sanjoy Mahajan, *Differential Pair Above Threshold*
3. Rahul Sarpeshkar et al., *A low-power wide-linear-range transconductance amplifer*
4. Tobi Delbruck, *Bump Circuits*