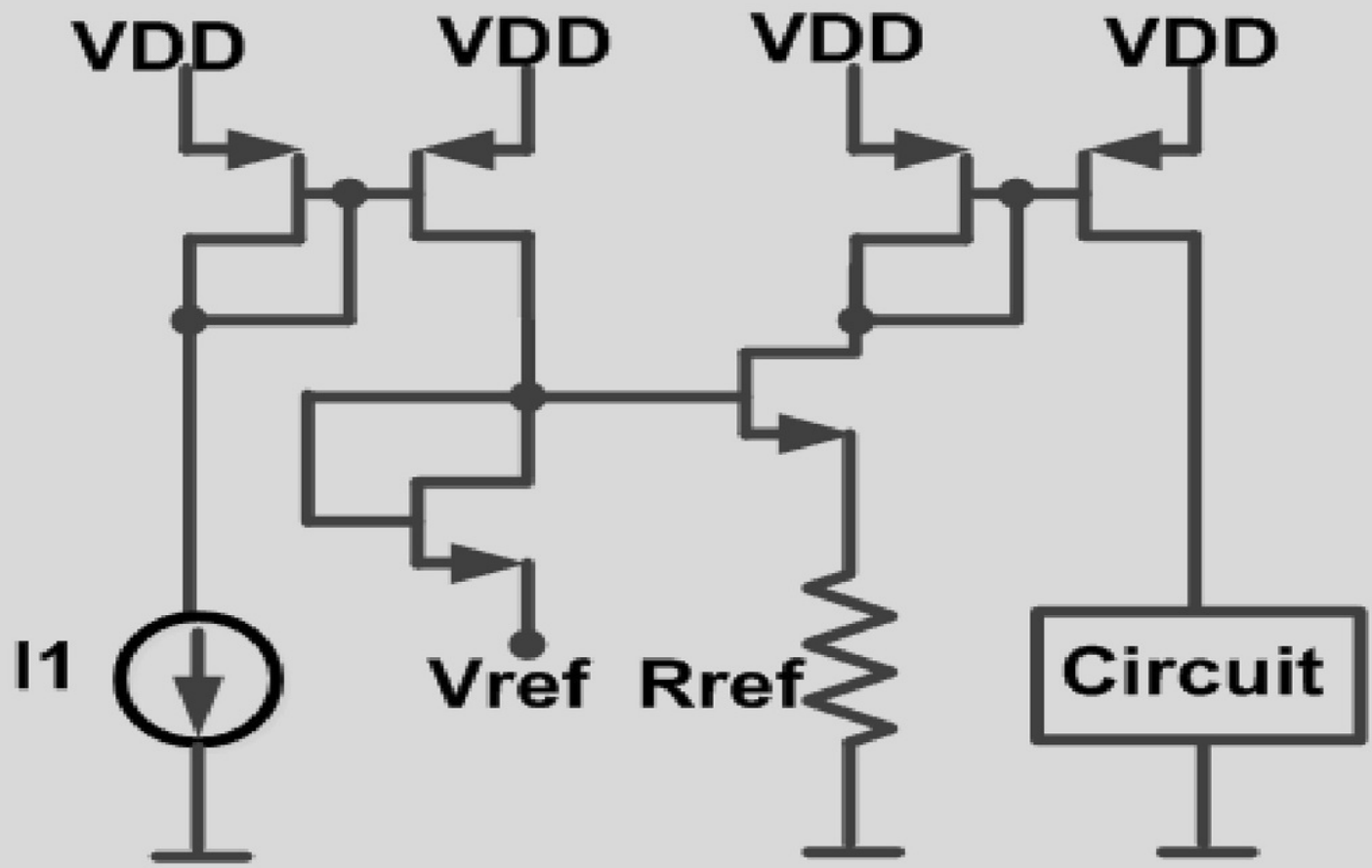
Zen of Analog Circuit Design



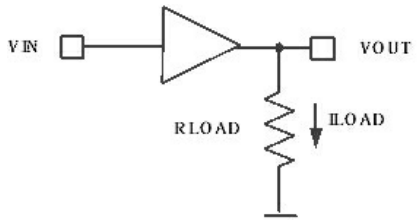
Anand Udupa

Contents

What this book is about

There are many different ways of teaching CMOS Analog circuit design. Most books start with an introduction to MOSFETs, their regions of operation and the I-V characteristics in each region. They present simple circuits and analyse how they work, gradually building up to more complicated circuit topologies. While such an ***analysis*** approach has its merits, the opinion of this author is that the true joy in learning Analog design comes from understanding the foundational principles behind how circuits are ***synthesized***, and not merely on how to analyse them. Analog design is an Art and the ***synthesis*** approach is what will help you pick up the topic similar to how a student picks up a new Art.

The book starts with a problem statement – *how do you realize an Analog buffer, a circuit whose output follows its input even when loaded*.

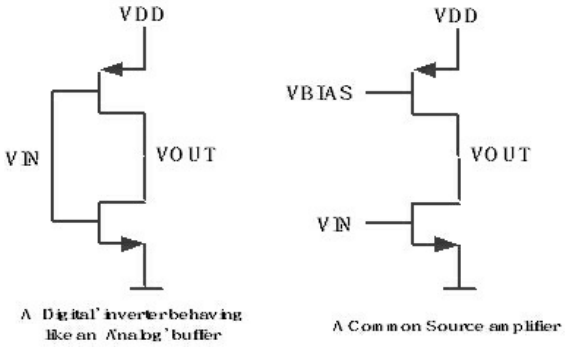


The intent of this book is that by the time the reader is done with reading it, he or she can not only build such a buffer (more accurately an approximation to such a buffer) but will also be able to appreciate the limitations of the circuit they have constructed. The journey to get to a more and more 'perfect' buffer will take us deeper into the concepts of Analog design – a journey the author intends to continue through a Part II of the same series.

To get started on answering the problem statement of building the Analog buffer, we revisit the basics of Electronics 101. We start with the simplest elements – Voltage source and Current source, and show graphically how a circuit involving them goes about establishing its operating point. We then extend the concept to Controlled Voltage and Current sources. Unlike 'Passive' elements like R,L, and C, the MOSFETs (or more generally 'Active' devices) have a controlling terminal which can alter their I-V characteristics. For a circuit involving such devices, the operating point of the circuit can be made to change based on the stimulus applied on the controlling terminal. So the introduction of Controlled sources brings in a new dimension to circuits – one in which the change in the operating point of

the circuit can be thought of as its ‘response’ to the signal, which is the stimulus applied to the circuit. The buffer we are looking to design is a type of circuit where the response is equal to the stimulus and follows it exactly.

We then briefly introduce the MOSFET but without getting into its detailed equations. The intent of this book is to show the ‘why’ without getting burdened by the ‘how’. We see enough of the I-V curve of the MOSFET to realize that in one of the regions of its operation, it behaves like a controlled source – more specifically a ‘voltage-controlled current source’. Going back to our analysis of operating points, we see that connecting two MOSFETs in a circuit leads to a situation akin to connecting two nearly-ideal current sources in series. It is tough to establish an operating point where both the transistors are operating in the intended region of operation. Even if such an operating point is established, a small change to the input leads to a large change in output, causing the transistors to shift from the intended region of operation. In fact, such a circuit is actually the familiar digital inverter attempting to behave like an analog buffer. A small tweak to the digital inverter circuit brings us to the topology of a Common source amplifier, which serves as our point of entry into the world of Analog circuits.



A ‘Digital’ inverter behaving like an ‘Analog’ buffer

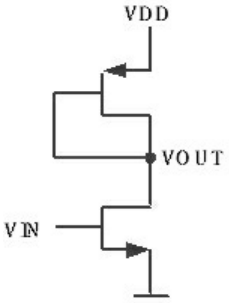
A Common Source amplifier

We find that it is not only difficult to establish a stable operating point for such a circuit but also the response of the circuit to even a small stimulus is high and not well controlled. This is because a small change in the input (stimulus) can potentially shift the operating point of the circuit by a huge amount. This finding leads us to the fact that such a circuit has the potential for high gain between its input and output, but is inherently very tough to get to operate in a reliable manner.

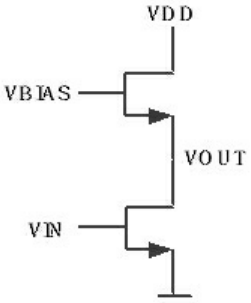
By this point, we have gotten to see the possibility of achieving high gain from a circuit involving Active elements. However, we are still some distance from realizing a buffer. We see that in addition to the operating point of our

initial high-gain circuit being extremely sensitive, it also suffers from not being able to maintain its operating point in the presence of loading. For example, connecting a resistor to its output can cause the operating point to shift quite a bit and cause the transistors to move away from their intended region of operation. This makes such a circuit further deviate from the attributes of a buffer, which is expected to be a circuit that maintains its operation even in the presence of loading.

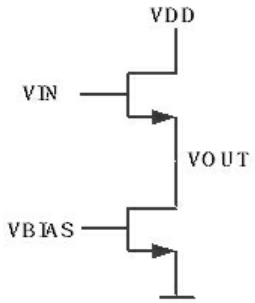
What we now need is a way to rein in the high gain properties in the circuit we have constructed and instead make it operate in a robust manner for a low gain. This involves making the transition from a circuit capable of high gain but having a sensitive operating point and an inability to be loaded, to a circuit with unity gain but with a stable operating point even in the presence of loading. We will see that to make this transition will require understanding one of the fundamental principles of Analog circuit design. This is the principle of ***Feedback*** – more specifically, Negative feedback. It will also reveal how the MOSFET can be more than what it initially appears to be – which is a voltage-controlled current source. Once we start to see how the MOSFET can change its behaviour based on what we impose on it, a plethora of circuits start to emerge. These include circuits like the common source amplifier with gm-load and with diode-connected load, and the Source follower.



A Common Source amplifier with a Diode-connected load

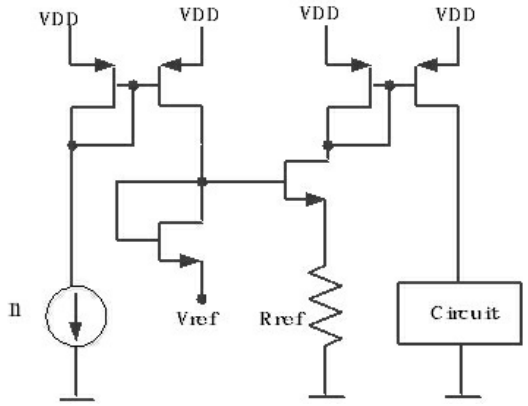


A Common Source amplifier with a 'gm' load



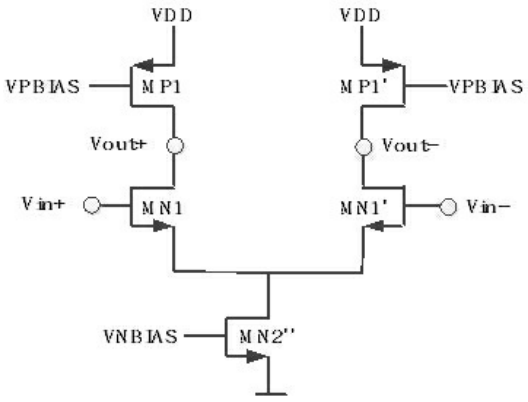
A Source follower buffer

Each of these circuits is introduced in a manner that illustrates the power of the 'synthesis' approach to circuit design – you first determine how exactly you want each MOSFET in the circuit to operate, and then connect it up in the manner that makes it behave the way you want it to. We extend the power of this approach to circuit analysis as well, and analyse intuitively (without resorting to equations), a fairly complicated circuit – the one shown below.

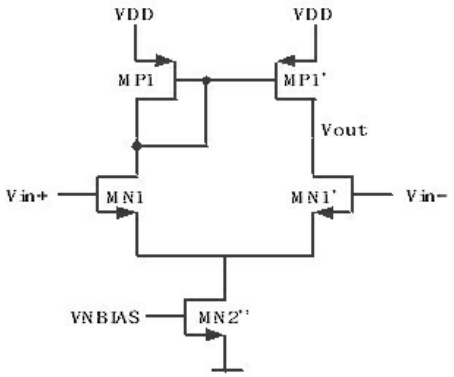


We will see how one of the circuits we have synthesized – the source follower – is actually a rather decent approximation to a voltage buffer. We do this by introducing the small signal parameters of the MOS transistor: gm and gds, and by showing that the source follower has a gain close to '1' and a small-signal output impedance that is relatively low. However, we realize that there is scope to do much better and realize a buffer that is a much closer approximation to an ideal buffer.

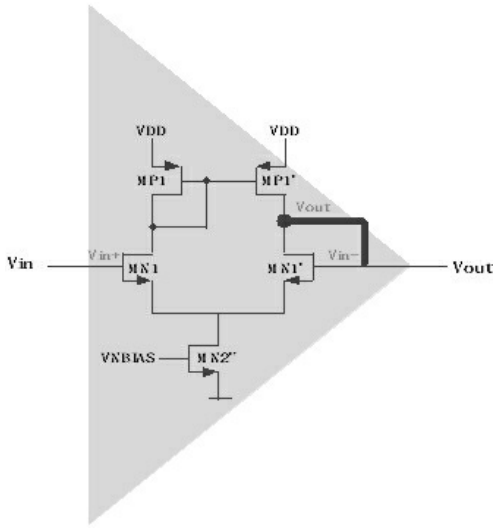
In getting to the source follower buffer, we realize that we have moved away quite a bit from the point where we started – which was the common source amplifier that had the potential to give a high gain. We pick up the Common source amplifier from where we left and address ways to solve the issues associated with it – namely a mechanism to establish its operating point in a robust manner. This leads us first to the topology of a Differential amplifier shown below.



We address the issues of operating point that continue to be there in the Differential amplifier. This then leads us to the topology of the Operational amplifier, the first approximation to which is a simple circuit built with five transistors as shown below.



Using the Operational Amplifier in a feedback configuration, we once again realize our buffer in the manner shown below.



We next revisit a very important aspect of the foundational principle on which all our circuits have been constructed – negative feedback. We look at conditions where feedback can go bad and cause undesired effects in our amplifier. We look at the circuit mechanisms that cause this and introduce a powerful technique that can mitigate these effects and restore the mechanism of negative feedback.

The quest for the ideal buffer

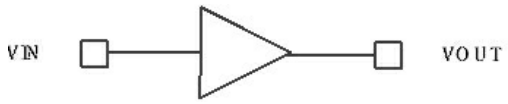
“I want to know God’s thoughts… the rest are details.”

- ***Albert Einstein***

The Holy Grail of Analog design can be summed up as a quest for an elusive component– the **Ideal buffer.**

But what exactly is an ideal buffer?

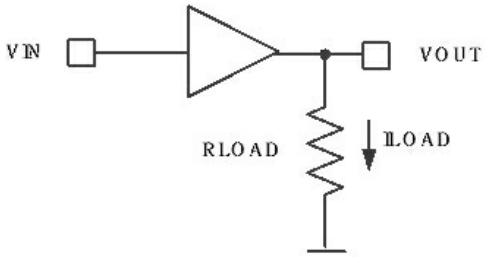
The simplest definition of an ideal buffer is one where the output ‘follows’ the input.



In other words, an ideal buffer is one where **VOUT=VIN**.

To put it more verbosely, an ideal buffer is one where the Response (Output) tracks the Stimulus (Input).

But there is one more attribute to an ideal buffer. The response (output) should continue to match the stimulus (input) even when the buffer gets loaded as shown below. Here the load is shown as a resistor drawing a DC current of ILOAD from the output of the buffer. But the load could also be a capacitor that draws a switching current – or the load could draw a combination of a DC and a switching current.



So an ideal buffer is one which responds in an appropriate manner to a stimulus – and this response is unaffected even if the buffer is loaded.

Now let us strike an analogy with one of the most important facet of human existence – our relationships.

Let’s say a couple has been contemplating taking a long overdue vacation.

The wife makes a fervent appeal to the husband to take a break so that they can plan a vacation to Bali – a place which serves as the setting for her favourite 9 PM soap opera. The husband agrees and works hard to get all his work completed so he can take off. The wife surfs the Internet, planning out every detail of the vacation. On the day the couple is supposed to fly out on their vacation, an important customer of the husband calls and reports a production-stop issue.

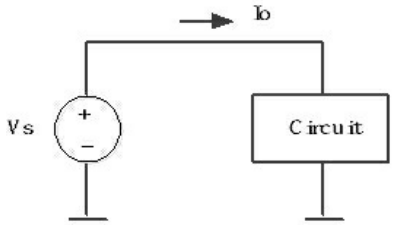
I suspect you get an inkling of where I am headed. So let me come straight to the point. The expectations from an ideal relationship bear some striking similarities to those from an ideal buffer. There is a stimuli - in the above example, it is the wife's request for a vacation. The response –the couple taking off for the vacation – was appearing to be an appropriate response to the stimuli. Until the point where the husband received that fateful customer call… Well, every relationship has to deal with several types of 'loading' – be it demanding customers, a painful boss, hyperactive kids, or interfering in-laws. And similar to our expectations from an ideal buffer, we can state an ideal relationship as one where every stimulus elicits an appropriate response - irrespective of the effect of the loading factors!

The Electronics 101 we need to refresh

To get started with the quest for the ideal buffer, we start with a recap of two of the most basic elements of Electronics. The first one is the ideal voltage source.

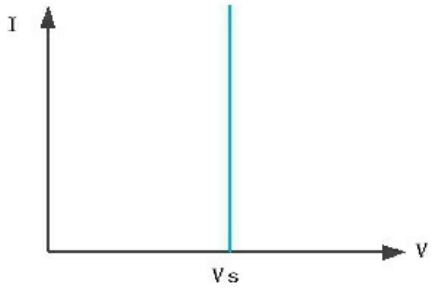
The ideal voltage source

I have shown below a voltage source, connected to a node in a circuit. It tries to impose the voltage (depicted by the symbol Vs) on the node it is connected to. The circuit in turn loads it with a current Io.

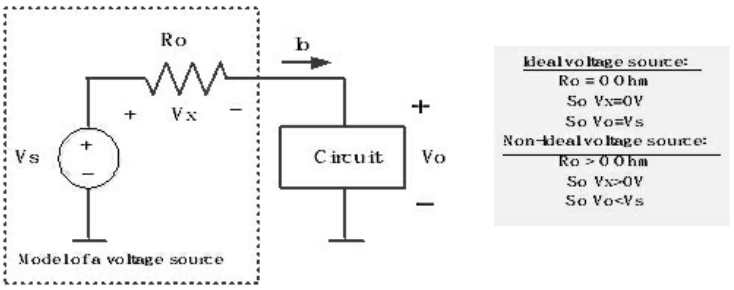


If the voltage source is ideal, then it continues to drive the node (of the circuit it is connected to) to the same value Vs irrespective of how much the current loading (Io) from the circuit is.

We can capture this graphically as an I-V curve as shown below.



Another property of the ideal voltage source is that it has zero output impedance. What this means is that if you load it with any current, however high, it will still impose the same voltage (equal to Vs) across the load. This is depicted below.

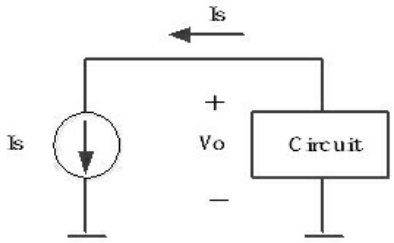


A non-ideal voltage source would have its output impedance, Ro to be non-zero. This would lead to the voltage Vo (the terminal voltage of the non-ideal voltage source) to keep reducing as more current gets drawn from the voltage source.

While the voltage source is a good starting point, we will use a different element to start off on our Analog journey.

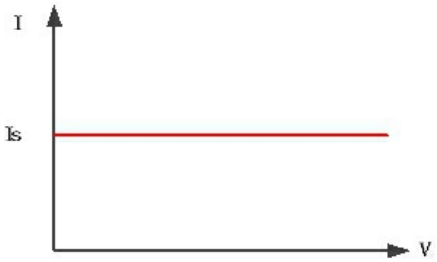
The ideal current source.

As shown below, the ideal current source is one that draws (or pumps) a constant current from (or into) the node of a circuit.



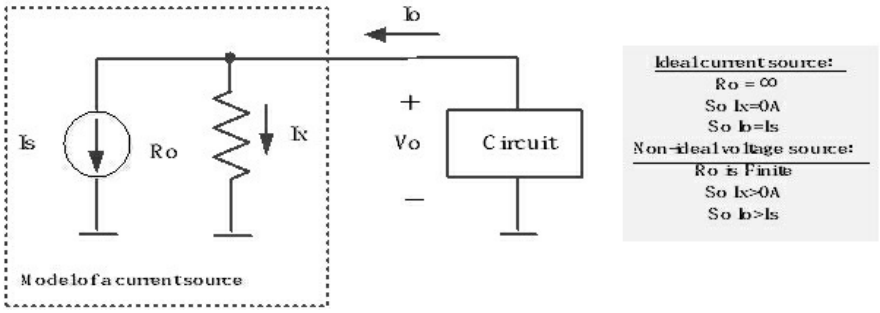
As before, the operative word is ‘ideal’. In this case, the current remains the same (equal to Is) irrespective of the voltage Vo imposed by the circuit across the current source. We will call Vo as the terminal voltage of the current source.

The I-V curve of an ideal current source is shown below.

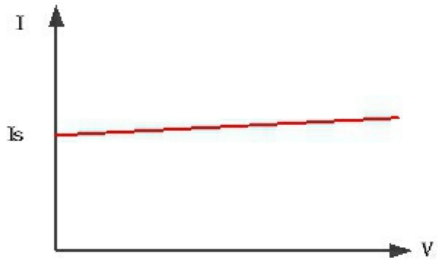


Another property of the ideal current source is that it has an infinite output impedance as shown below. What this implies is that irrespective of the

voltage across it, the full current Is will flow into the load. This is shown below.

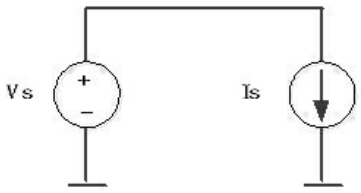


For a non-ideal current source, Ro would be finite. A finite value of Ro would cause the output current Io to have a dependence on the terminal voltage across it. The I-V characteristics of such a non-ideal current source is shown below.

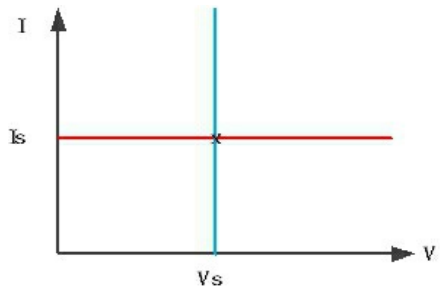


Usually, the current source value slightly increases as the terminal voltage across it is increased – this is consistent with the positive slope of the I-V curve shown above. This is because the finite Ro draws an extra current as the voltage across the current source increases.

Let us consider a case where an ideal voltage source is connected to an ideal current source as shown below.

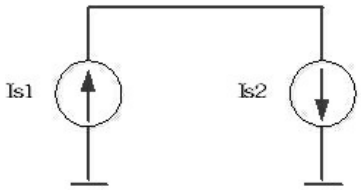


It can be readily concluded that such a circuit is not just a valid one but a ‘happy’ one. The current source imposes the current Is and the voltage source is happy to operate with any current through it. Similarly, the voltage source imposes the voltage Vs and the current source is happy to operate with any voltage across it. The I-V curves showing the point of intersection for a case corresponding to both the sources being ideal is shown below.

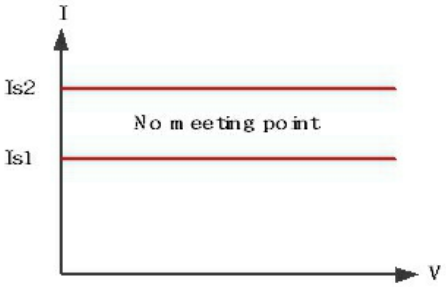


A voltage source and current source connected to each other forms a happy circuit and is able to establish its operating point easily.

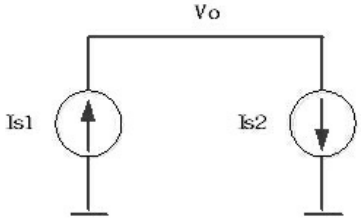
Now let's make at an interesting observation. Let us connect two ideal current sources in series as shown below.



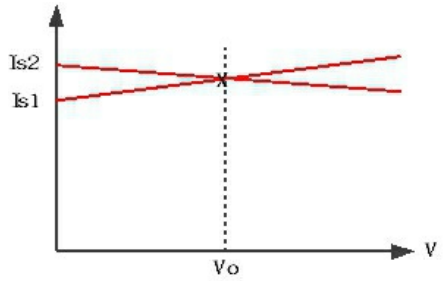
It is clear that Kirchoff would not be happy if the two currents were of unequal value. After all, where would the difference current flow? Such a circuit would be in violation of Kirchoff's Current Law. From a graphical perspective, the two ideal, unequal current sources would find no 'meeting point' as shown below. Graphically viewed, the source of the conflict is the lack of a meeting point.



Now, let us assume a case where the two current sources that are connected in series are non-ideal. In other words, the current changes just a little bit with the voltage across each current source.



By our definition of the terminal voltage of the current source, an increase in Vo increases the terminal voltage of Is2 while reducing the terminal voltage of Is1. So the two I-V curves slope in opposite directions as shown below.



Because the curves are not horizontal, the current sources are able to find a meeting point. This results in them being able to operate with a common voltage Vo across them.

The non-ideal current source will be the element we will use to establish the analogy with human relationships!

Having been thus introduced to the concept of the ideal and non-ideal voltage and current sources, we now look to strike the human analogy. Quite simply and without justification, let me state that a **current source** is akin to a person's **thoughts**. If you ask me why, then I would say that our thoughts 'flow' just like a current does. I am sure the reader can come up with an equally persuasive argument as to why the voltage source is a better representation of the human thought! But let's just go with *Current source = Human thought*…

Then what would the voltage source represent? How about…

Voltage source = Human action…?

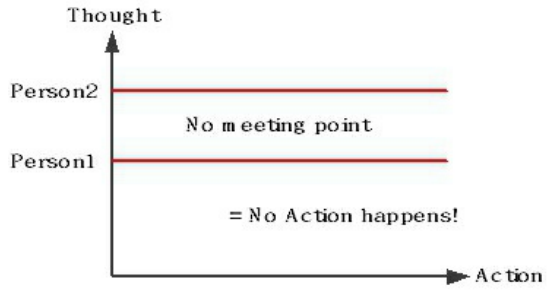
So our analogy is the following: the current source is representative of our thoughts, convictions, biases, prejudices, and the like- essentially the things that go on in our head when we deal with others. By contrast, the voltage source represents the actual action or outcome of our interaction with others.

Now with this analogy, what would an *ideal* current source represent? Quite simply, it would represent a person whose thought process is so rigid that it cannot be changed or influenced!

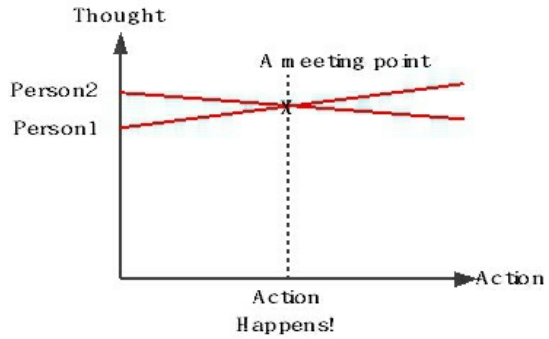
This analogy leads to some interesting results!

Let us consider an interaction between two people, each with extremely rigid

ways of thinking. They are debating on a matter that eventually requires some action to be taken. The likely scenario would be that each would hold on to his or her view so strongly and would not find a meeting point with the other! Imagine two extremely strong-willed individuals debating all day on what is the right thing to be done. As a result, no action happens. See the below plot if you do not believe me!

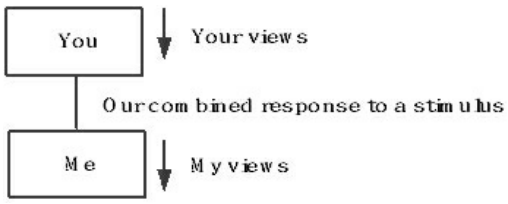


Now let us consider a case where you have two individuals who are akin to *non-ideal* current sources. Each is willing to yield a little bit in their thought process. Clearly a meeting point is possible – and the likely result is that they are able to decide on some mutually acceptable course of action. What results is the below.

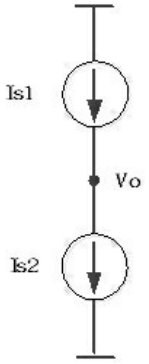


Let us now change the words a bit. What we called as Action, we will now refer to as a **‘response to a stimulus’**. Since we are talking about the specific case of relationships involving two people, we will qualify it further as a ‘*combined response to a stimulus*’.

Our “circuit” model for human interaction in a relationship therefore looks like the below.



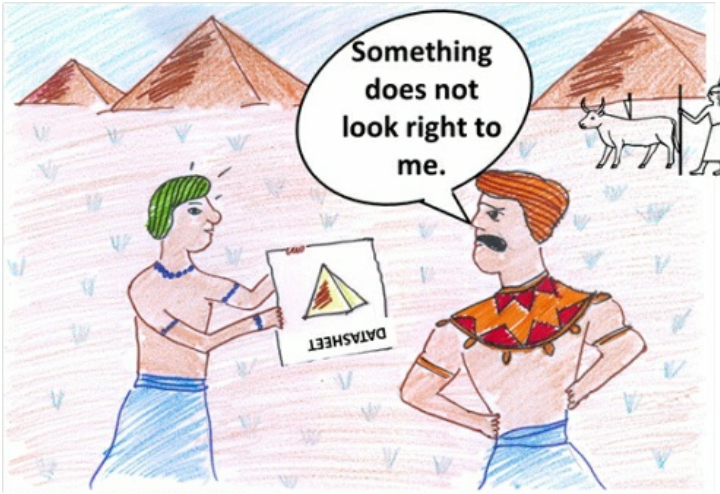
Does not the above “circuit” remind you of one we had looked at just a bit earlier – the one redrawn as below?



Here’s one more justification as to why the current source human analogy is not a misplaced one. As humans, we are usually OK with any outcome (and are capable of almost any response to a stimulus) as long as we think we are being right! Much like a current source that is OK to take any voltage across it provided it can impose its current…

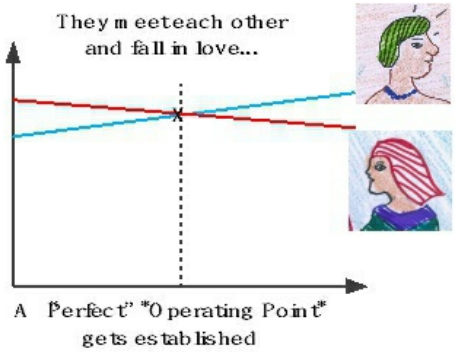
Let us use a story to reinforce the analogy. It is time to get introduced you to the cast of our story.

Our story starts in ancient Egypt 3300 years ago. Our hero is **Aman-Ra**, a civil engineer who works in the VLSM industry. That is short for *Very Large Scale Mummification*. In layman terms, Aman-Ra builds pyramids. His company’s clients are the hard-to-please Pharoahs. He has an equally hard-to-please boss called **Gamen-Ra**. Aman-Ra dreads the weekly meetings he has with his boss.



Aman-Ra is a happy and single but that is going to change soon – I meant the part about being ‘single’. The one destined for him is a beautiful girl called **Uman-Ra**. She works for the Pharoah’s wife in the role of a CFO (Chief Fanning Officer).

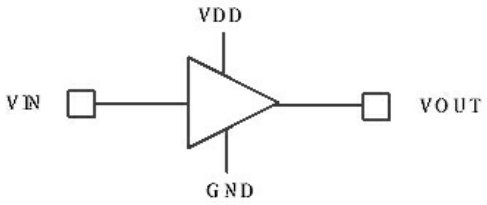
Aman-Ra and Uman-Ra met one day on the banks of the river Nile. The challenge of a relationship is how two people can come together, find a meeting point between them and be able to respond to a stimulus in a desired and predictable manner. Now, they were both very strong willed individuals, much like ideal current sources. But there is one force that can ‘bend’ the I-V characteristic of an ideal current source and make it non-ideal. The magic of **Love** can create between any two people the perfect meeting point. Technical jargon seems out of place in a romance, but we will take the liberty of using the term ‘*Operating point*’ instead.



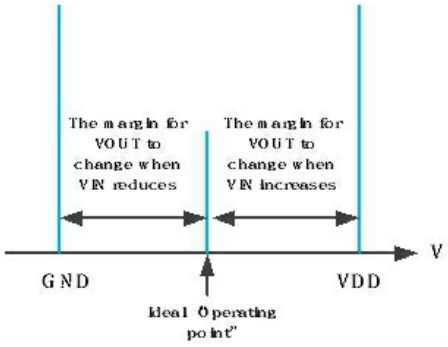
But what do we mean by a *perfect* Operating point? For that, we will have to jump back to the world of circuits.

Now, a realizable circuit always operates within 'bounds', the bounds usually being just the supply and ground terminals of the circuit. These are the terminals that provide power to the circuit. For most circuits of interest, we can assume that the voltages at all the nodes in the circuit are constrained to be in the range between the voltage of supply (VDD) and ground (GND, which is usually 0V).

In our buffer, these would be represented as below.



We can now visualize these bounds as below.



In the case of the circuit being the two current sources, a *valid* operating point would therefore be one where they are able to find a meeting point within the range between VDD and GND. An *ideal* operating point would be one where the meeting point is more or less centred between VDD and GND. What

makes it *ideal* is that the output now has ‘room’ to swing on either side of the operating point till the extremes of VDD and GND.

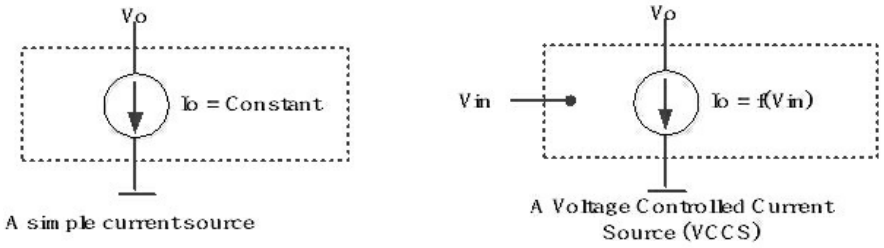
We will extend this concept to the human analogy but before that we will need to meet some new friends – the Lilliputs who rule the world of circuits.

An (Non-Mathematical) Introduction to NMOS and PMOS transistors

An introduction to circuit design usually starts with the building blocks R, L and C. These are referred to as **Passive** elements and there is quite a bit of stuff you can do with these elements. But if you want to do real magic with Analog, you need to start understanding about another set of building blocks. These are called … you guessed right … **Active** elements! Buried deep in some text book you will find the proper distinction between an active element and a passive one. But this book is for the rest of us, so let me give you my take on what I think an active element is.

I want you to think of an Active element as simply one that has *life*! I will elaborate on this shortly.

How do we model an active element? We model it using what we will refer to as a **Controlled source**. To begin with, we will look at a controlled element called the **Voltage-Controlled Current Source** (VCCS). Shown below is how it differs from a simple current source.

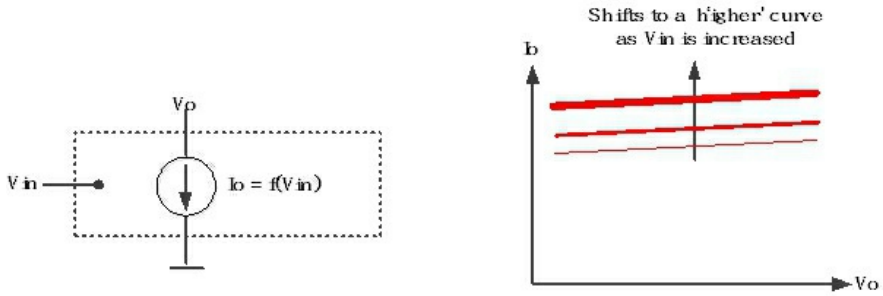


The symbol can be elaborated as follows – the VCCS behaves like a current source across its output terminals (in the above case the terminals marked as Vo and ground). However, the value of the current is a function of the voltage on Vin – depicted as f(Vin). For a given value of Vin, the element will behave at its output terminals as though it were a constant current whose value is set by f(Vin). In the above circuit, Vo is the ‘terminal’ voltage and Vin is the ‘controlling’ voltage.

The VCCS is not the only type of controlled source. The controlling parameter can be either a current or a voltage and the controlled parameter also can be either. This leads us to 4 combinations – VCVS, VCCS, CCVS and CCCS. But we will restrict our analysis for now to the VCCS - the reason will soon become clear.

A property of an active element that emerges with the above model is that it changes its terminal characteristics (Io versus Vo) when a stimulus (Vin) is applied! Again not a misplaced analogy – as humans, the state of our minds is constantly changed by the stimulus that keeps coming our way. A deadline here, a few harsh words from the boss – that's all it takes to swing our mental pendulum from one extreme to another!! This is what makes our relationships so interesting!

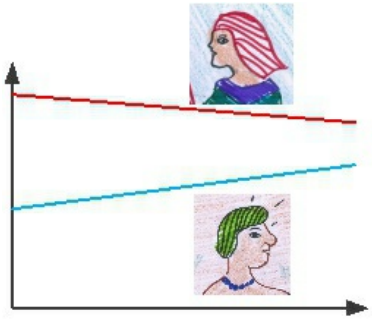
It is very useful to visualize the active element (as modelled by the VCCS) from a graphical perspective. This is shown below. The I-V curve of the active element is now not a single curve but a family of curves. For each value of Vin, there is an I-V curve that models the terminal characteristics of the element, namely the current(Io) that flows through the output terminal versus the voltage (Vo) of the output terminal. In the example shown below, the I-Vo curve shifts up as Vin is increased.



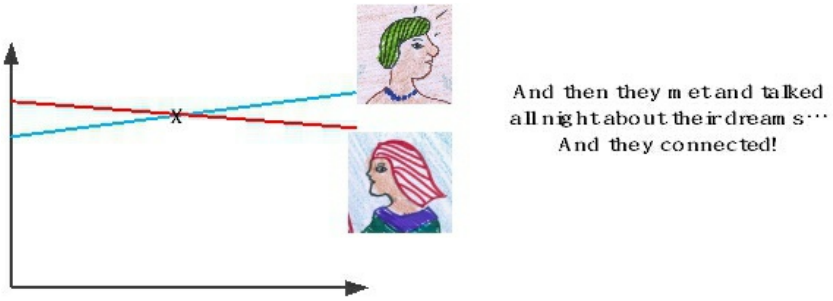
The fact that a change in Vin causes the I-Vo curve to shift can be succinctly captured as "a stimulus stirs a response from the active element". Contrast this with a passive element whose terminal characteristics remain constant (passive) irrespective of what signal is applied to them. For example, the resistor has a constant ratio between its terminal voltage and current (given by Ohms law) irrespective of how much voltage is applied across it. This is because it does not have any controlling terminal that can change its I-V characteristics. Unlike in active elements, passive elements have no 'knob' that can change their characteristics.

Coming back to our chief characters…

Before they set their eyes on each other, they had nothing in common.

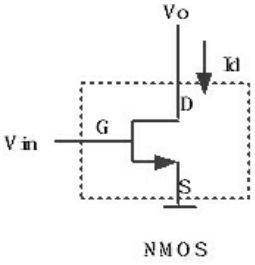


And then they met and as they got talking to each other, their “curves” shifted.

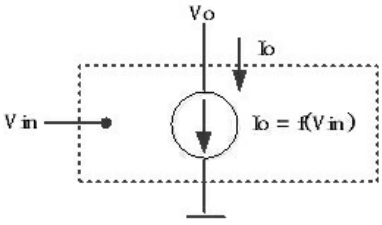


But without further ado, we must get to the MOS transistor. They are after all the Aman-Ra and Uman-Ra of our Analog story!

First we focus our sights on the **NMOS** transistor. Its three terminals G,D and S refer to **Gate**, **Source** and **Drain**.



The above symbol is akin to the VCCS.



In the connections to the NMOS shown above, the input Vin is applied to the Gate (G) terminal, which is the controlling terminal. The output node (with

voltage Vo) is indicated as connected to the Drain (D) terminal and we have shown the Source (S) terminal connected to ground. The current through the element flows from Drain to Source and is referred to as the Drain current, Id. Note that there is no current flow through the Gate – it is merely a controlling terminal. To terminology is expanded below.

Controlling voltage = V_{GS} : the voltage of the Gate relative to the Source

Terminal voltage = V_{DS} : the voltage of the Drain relative to the Source

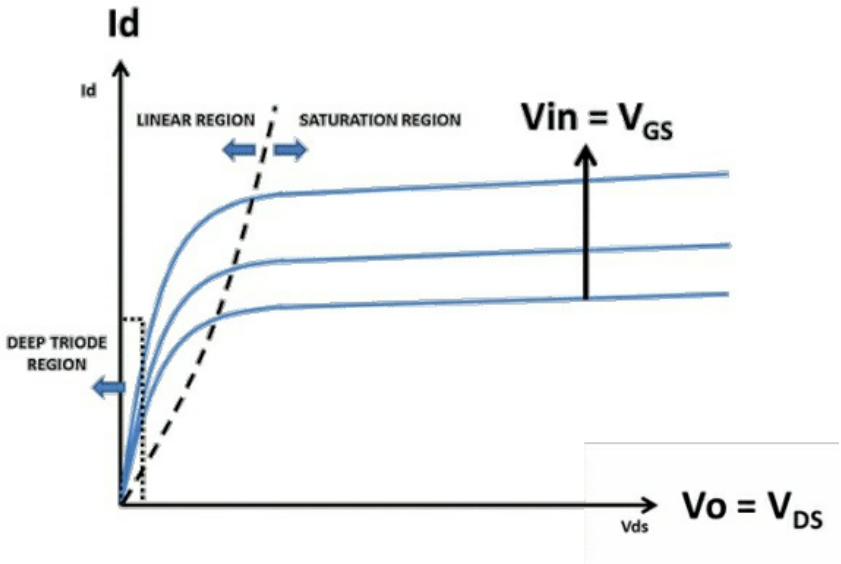
The equation that determines the drain current Id of the NMOS transistor is:

$$Id = K.(1+\lambda V_{DS}).(V_{GS}-Vt)^2/2$$

There are some constants like K, Lambda and Vt but we do not need to worry about them too much for now.

Note that Id has a dependence on both Vin (which is V_{GS}) as well as Vo (which is V_{DS}). While the dependence on Vin is obvious (Vin being the controlling terminal), the dependence of Id on Vo suggests that the MOS transistor behaves like a non-ideal current source. Its drain current has a weak dependence on the terminal voltage V_{DS}. The justification for using the term 'weak dependence' comes from two factors. While the dependence on V_{GS} (or Vin) is quadratic, the dependence on V_{DS} (or Vo) is linear. Also usually Lambda is <<1, making the dependence of Id on V_{DS} weak.

If we plot the I-V curve of the NMOS transistor , we get what is shown below.



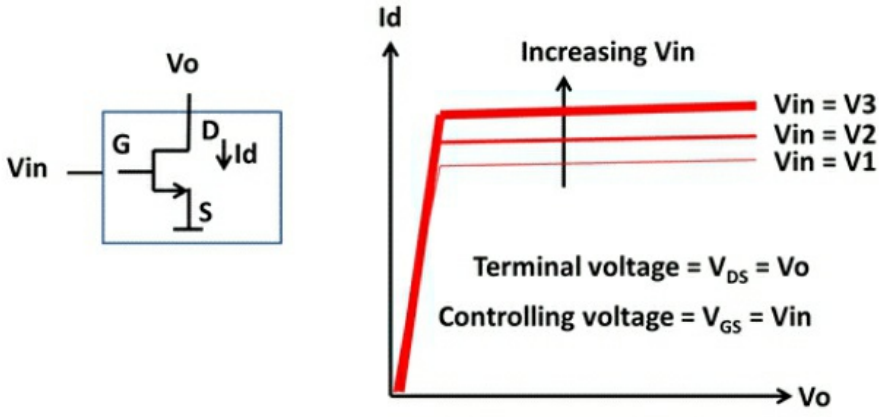
There are several points to note in the above figure:

1. A family of curves is shown. The curves show a plot of Id versus V_{DS} and each curve corresponds to a certain value of V_{GS}.
2. As V_{GS} increases, the Id-V_{DS} curve shifts to a higher one. Because of the quadratic dependence on V_{GS}, the curves start to diverge faster as fixed increments to V_{GS} are applied.
3. Three different regions are shown and labelled as Saturation region, Linear Region and Deep Triode region. What differentiates these regions is the range of V_{DS} voltage – or more specifically the relation of V_{DS} to V_{GS}. For now, we will just assume that if V_{DS} is high enough, the MOS transistor will be in the Saturation region. This is where our initial Analog design will happen.
4. It is also to be noted that the equation we wrote for Id only models its operation in the Saturation region. In fact in the Linear and Deep triode regions, Id has a strong dependence on V_{DS} as can be made out by the considerable slope of the I-V curves in those regions.
5. As can be observed from the curves, the MOS transistor behaves like a (non-ideal) VCCS in the Saturation region. It is in this region where the Current source-like behaviour of the MOS transistor shows up. Since this is commonly the region where we look to operate most MOS transistors in Analog circuits, it should

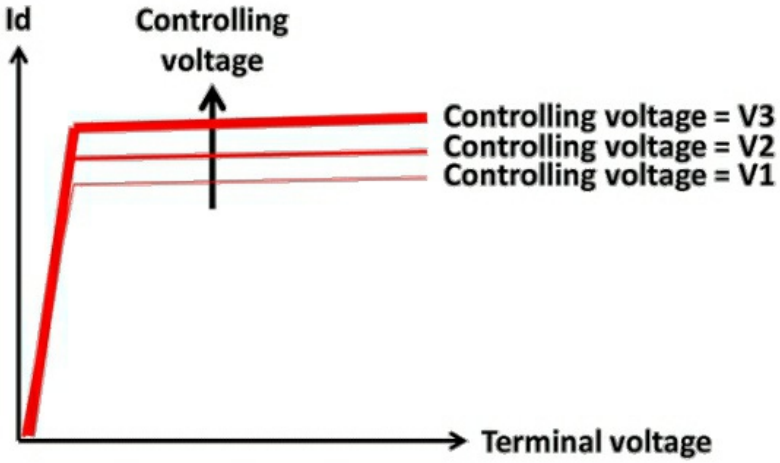
put to rest why we chose the analogy of human behaviour to a current source!

In summary, while operating in the Saturation region, the MOS transistor has a High sensitivity to V_{GS} (Vin) and a Low sensitivity to V_{DS} (Vo).

Let us simplify the MOS characteristics a bit as shown below. It is easier to draw and also gives a lot more insight into the part that really matters for now.



We can redraw the curves more generally as below.



We have shown above three curves of Id-Vo, corresponding to three different values of Vin (V1, < V2 < V3). As is expected from the equation, the Id-Vo curve shifts to a higher curve as Vin is increased.

Furthermore, let us choose V1, V2 and V3 as the following values:

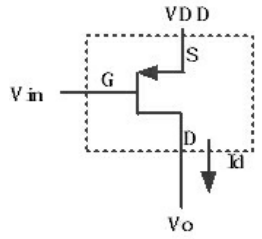
V2=VDD/2

V3=(VDD-V1)

Basically, we have taken an I-V curve that corresponds to a controlling voltage equal to VDD/2 (the middle of the supply-ground rails) and two other curves corresponding to controlling voltages spaced equally on either side of VDD/2. Note that we have drawn the curve corresponding to Vin=V1 with the thinnest line and the curve corresponding to Vin=V3 with the thickest line.

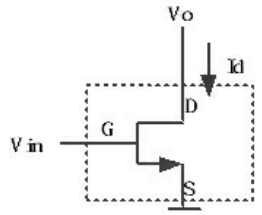
The reason why we chose these specific values of Vin will soon become apparent.

We are now ready to meet the **PMOS** transistor whose symbol is shown below.



PMOS

Keep a mirror on 'top' of the NMOS transistor's symbol - the PMOS transistor looks like its image!



NMOS

The Source terminal of the PMOS transistor is drawn at the top indicating that the Source is to be connected to the highest potential of the circuit, which is VDD, the power supply. Also note that just like we did for the NMOS transistor, we continue to apply Vin at the Gate and connect the output Vo to the Drain.

The important thing to keep in mind regarding the PMOS transistor is that the 'Terminal voltage' is V_{SD}, the voltage from Source to Drain.

The current flow direction is from Source to Drain.

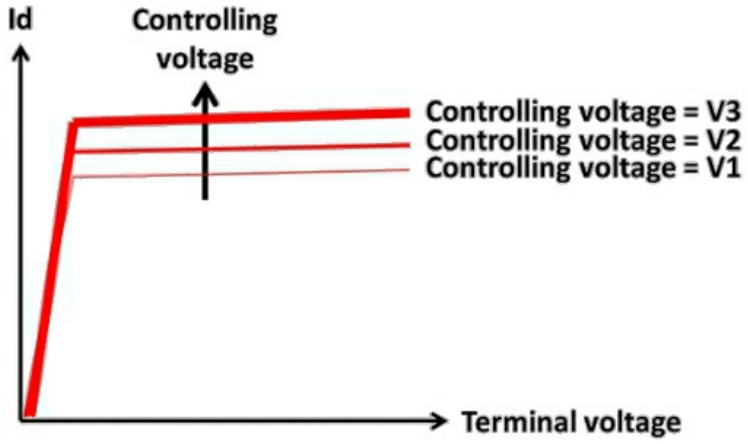
The controlling voltage is V_{SG}, the Source-to-Gate voltage. This is summarized in the table below.

	NMOS	**PMOS**
Terminal voltage of the transistor	V_{DS}	V_{SD}
Terminal voltage in the circuit	Vo	VDD-Vo
Controlling voltage of the transistor	V_{GS}	V_{SG}
Controlling voltage in the circuit	Vin	VDD-Vin

Current flow direction	Drain to Source	Source to Drain

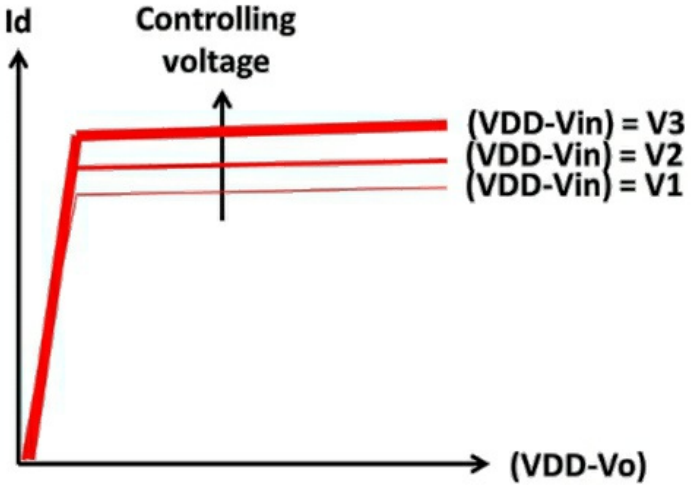
How does the I-V curve of the PMOS look when drawn based on its controlling voltage and terminal voltage?

Exactly similar to the NMOS! In other words, the below set of curves still hold even for the PMOS.



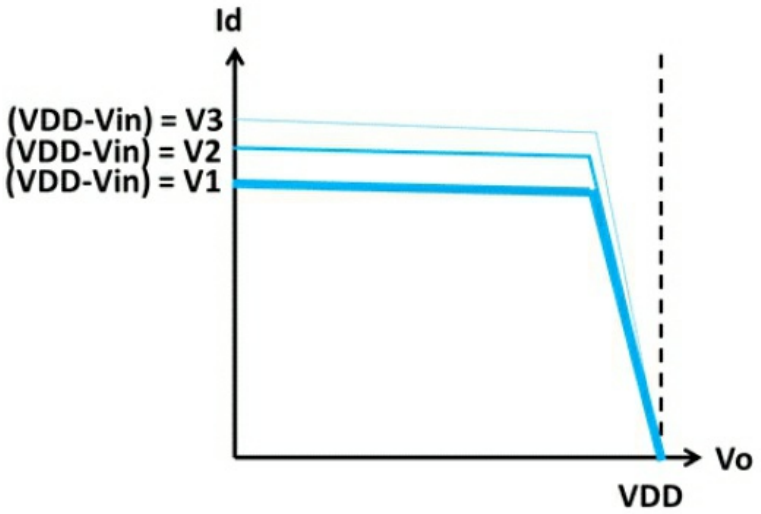
This tells us that the PMOS behaves exactly like the NMOS – one just needs to keep in mind that the terminal and controlling voltages just relate differently (with respect to the NMOS) to the voltages at the S, D and G terminals.

Expanding on the relationship of the terminal and controlling voltages of the PMOS in the circuit to Vin and Vo, we can redraw the above curves as:

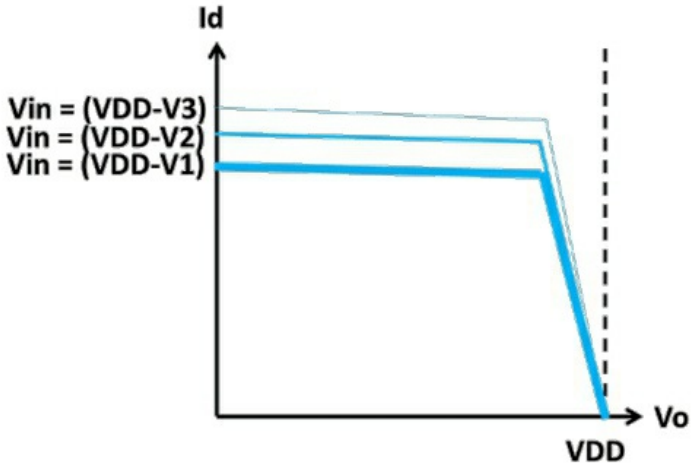


However, if we want the variables to be Vin and Vo instead, then we need to make some changes to the above curves.

First let us see what change is needed to make the x-axis as Vo. It is easy to see that the above curves merely need to mirrored about Vo=VDD. They look like below:



Next, if we have to associate the curves to Vin instead of (VDD-Vin), then it is easy to redraw as shown below:



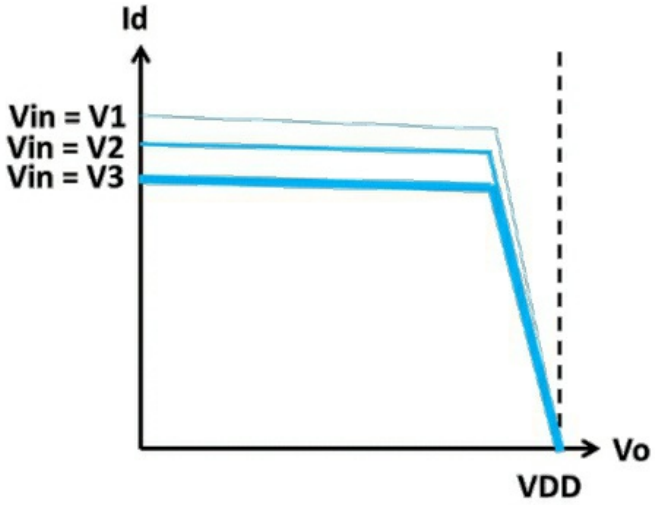
Now it is easy to see why we chose V2=VDD/2 and V1=(VDD-V3). For this choice:

(VDD-V2) = (VDD-VDD/2) = V2

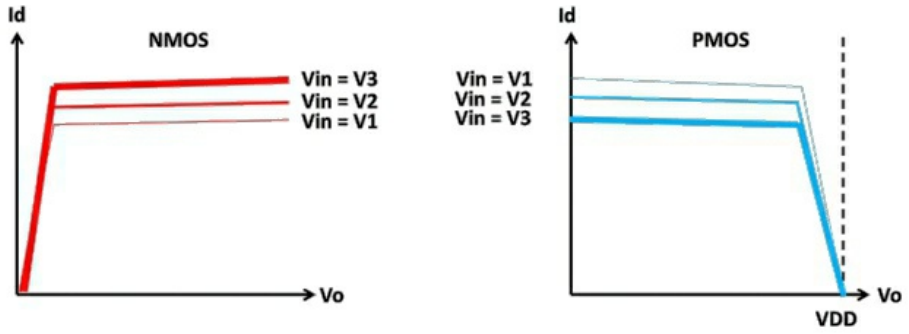
(VDD-V1) = V3

(VDD-V3) = V1

So the above set of curves looks as shown below:



For our choice of V1, V2 and V3, we can see how the NMOS and PMOS curves compare:



Essentially, there are two sets of mirroring that happen going from the NMOS set of curves to the PMOS set. The first is the mirroring of each curve horizontally. The second is the mirroring that happens vertically for all the curves around the centre curve (the one corresponding to Vin=VDD/2).

Note that for both the NMOS and the PMOS, we have drawn the curve corresponding to Vin=V1 with the thinnest line and the curve corresponding to Vin=V3 with the thickest line. This makes is easy to associate each PMOS curve with the appropriate NMOS curve!

Let us summarize some key points so that we can mentally assimilate the

PMOS:

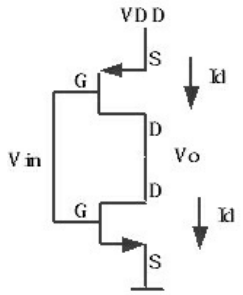
1. The Source is at a higher potential than the Drain for the PMOS. While the NMOS transistor source terminal is connected to GND (the lowest potential), the PMOS transistor source terminal is connected to VDD (the highest potential). Note this is not strictly a requirement but for the initial circuits we will build, this will be the case.
2. The controlling terminal for the PMOS is still the gate but the controlling voltage is the difference between the Source and the Gate. This is expected because the Source is the highest potential. That is the reason why the curves are shown shifting down as Vin increases (with increase in Vin, the controlling voltage V_{SG} reduces).
3. The current flows from the Source to the Drain for the PMOS. If we now define the PMOS drain current as the current from Source to Drain, then it is still positive. It is much easier to think in terms of positive current even for the PMOS, keeping the direction in mind.
4. The output terminal for the PMOS is still the Drain. However, the terminal voltage (the voltage between the positive and negative terminals of the current source) is now given by (VDD-Vo). So as Vo increases, the terminal voltage actually reduces. That is the reason the Id-Vo curves of the PMOS transistor becomes mirror images of the NMOS curves when plotted with the x-axis as Vo.

Take a few moments to digest the NMOS and PMOS curves.

We will now move on to designing our first Analog circuit!

The Digital Inverter as an Analog circuit

Now that we have understood the terminal characteristics of the NMOS and PMOS transistors, let us see what happens when we stack them up as shown below.



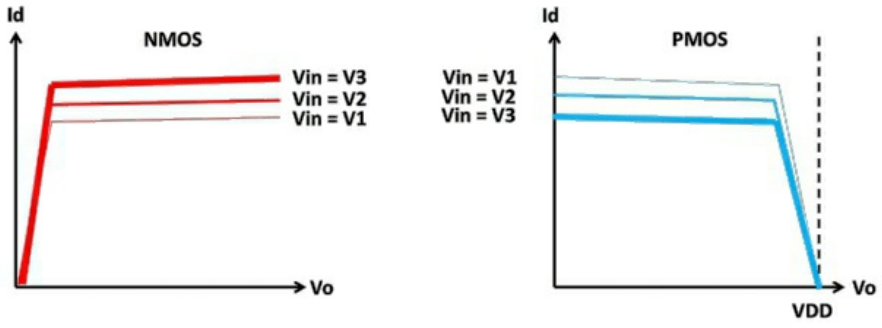
Wait… that looks like an inverter!

Yes, our plain old digital inverter is what is going to be our guiding beacon into the world of Analog!

Firstly, note that input Vin is applied to the Gates of both the NMOS and the PMOS.

The current Id through the PMOS (flowing from VDD to Vo) has to flow completely through the NMOS (from Vo to ground). There is no other path for it to flow to. That is the reason both the PMOS and NMOS currents are shown as Id.

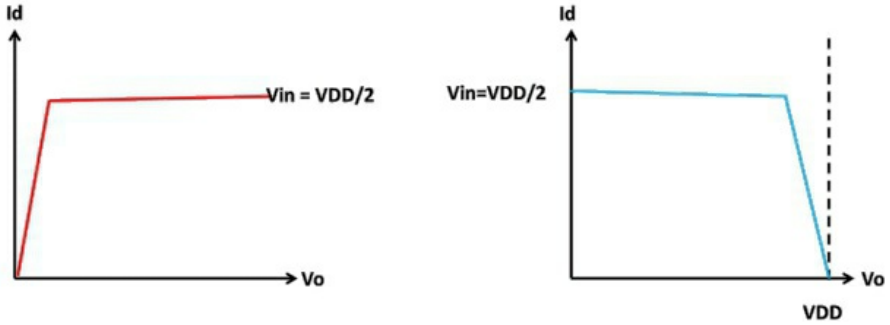
To recap how the curves look, we show below the NMOS and PMOS curves side by side. We have already plotted them using the parameters of the circuit which are shared between the PMOS and NMOS – Vin, Vo and Id.



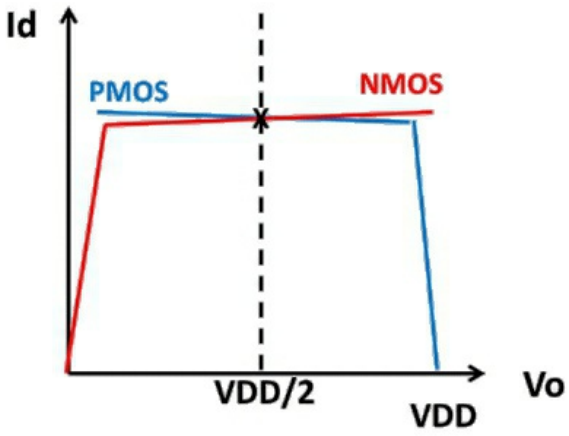
How would we now find out the output (Vo) corresponding to a certain input (Vin)? If you think carefully, we have drawn both the NMOS and PMOS curve with the same variable for the x-axis (Vo) and the same variable for the

y-axis (Id). So to find the 'operating point' of the above circuit, we just need to take the specific curves for the NMOS and PMOS corresponding to the same Vin and then we should find where they intersect. This point of intersection would determine the operating point for Vo and would also correspond to a value of Id that is same for the two elements.

Let us first consider the case of Vin=V2=VDD/2. In this case, the PMOS curve is the horizontal mirror of the same NMOS curve.



So for Vin=VDD/2, the intersection of the curves would be as shown below.



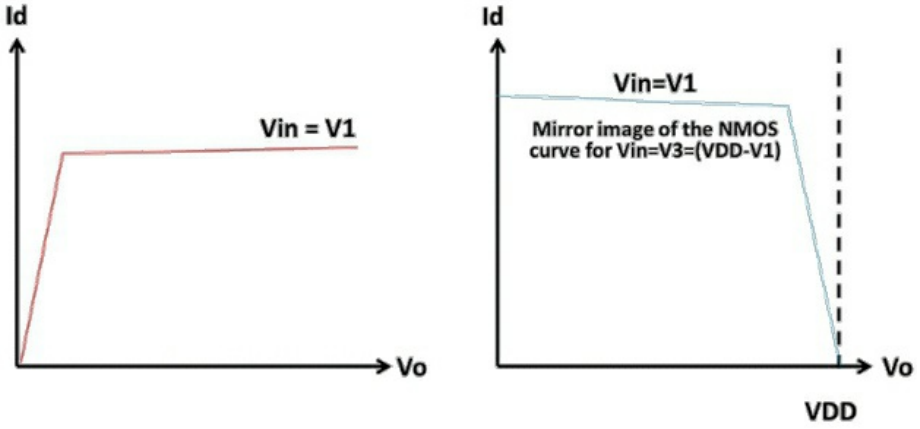
We see that for a Vin=VDD/2, the curves intersect at a value of Vo which is also roughly at the VDD/2 point. All this is approximate (in reality, the value of Vin at which the intersection happens at Vo=VDD/2 might be slightly off from a Vin=VDD/2). But the point is that if we carefully adjust Vin to a value which is roughly around VDD/2, we can get the operating point of Vo to be also be equal to roughly VDD/2. Also for the case shown above, both the NMOS and PMOS transistors are operating in their saturation region.

In the context of a buffer, this would be what we would consider as a

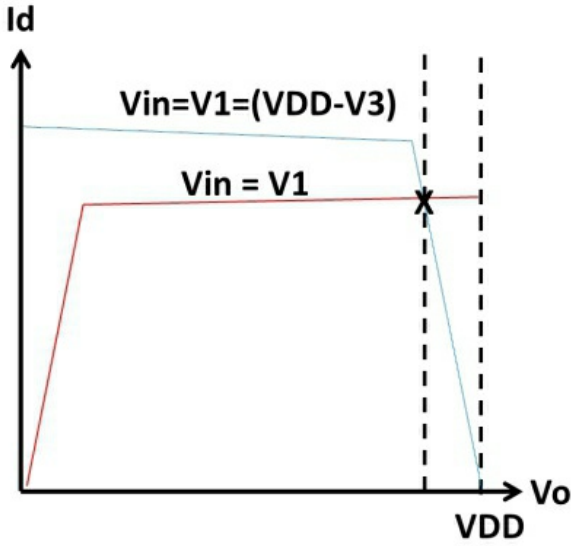
‘perfect’ Operating point, with Vo centred nicely in between the extremes of VDD and GND.

Now, let us move to a value of Vin=V1.

The NMOS and PMOS curves corresponding to Vin=V1 are shown below.



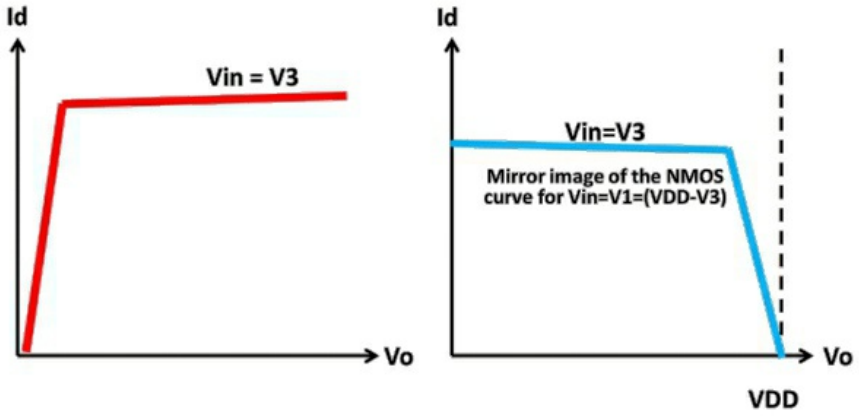
The operating point when Vin=V1 would be the intersection of the above two curves as shown below:



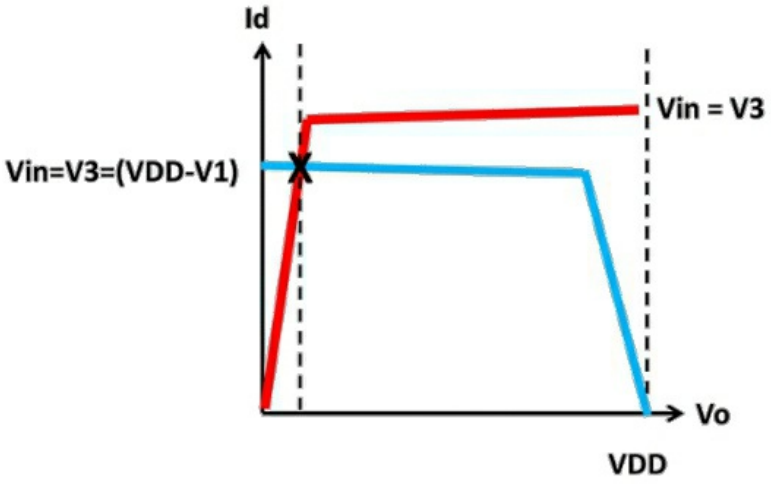
As can be seen, the operating point has shifted very close to one of the rails – VDD in this case. Note that V1 was a voltage less than VDD/2 and such an input has resulted in an output that is close to the VDD rail. It can also be seen that the point of intersection corresponds to the NMOS still operating in the saturation region whereas the PMOS has now shifted to operating in its linear region.

Now, let us consider a case where Vin=V3, a value slightly above VDD/2.

Here the PMOS curve would be the one got by taking the NMOS curve corresponding to Vin=(VDD-V3)=V1 and mirroring it horizontally.



The point of intersection would be as shown below.



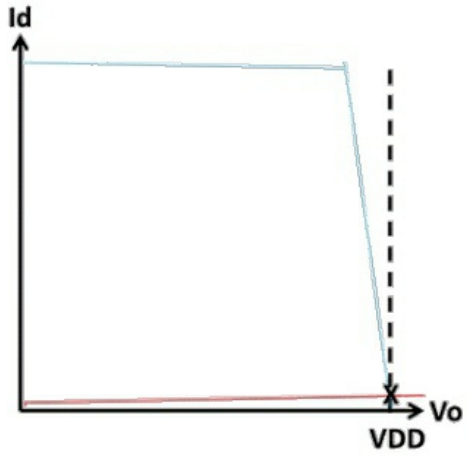
In this case, the operating point has shifted very close to ground. The PMOS is operating in the saturation region, whereas the NMOS has now shifted to the linear region.

The preceding analysis shows how sensitive the operating point is to the value of Vin. Any slight change around the optimum value – and the circuit lands up very close to one of the two rails!

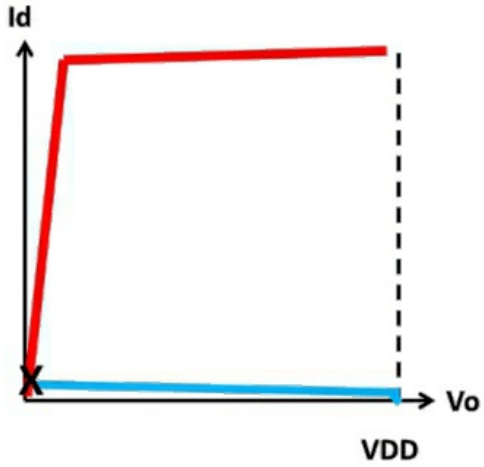
The above is not very tough to see. The circuit we have analysed is in fact the simple inverter. The properties of the inverter become apparent when Vin assumes values corresponding to ‘digital’ logic levels. (close to 0V for the logic ‘0’ and close to VDD for the logic ‘1’). The corresponding outputs also

correspond to voltage levels of the digital logic.

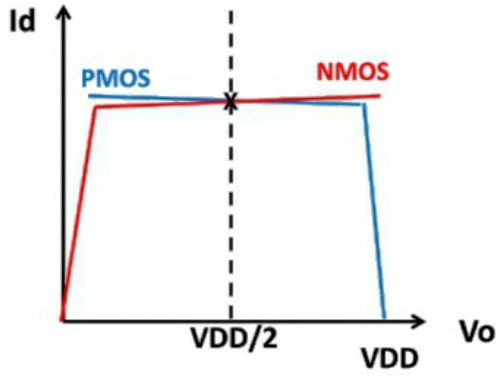
The intersection of the NMOS and PMOS curves for the case of Vin=0V (logic ‘0’) is shown below. For illustration, a value of Vin slightly higher than 0V is used. As is expected, the output is VDD (logic ‘1’). Also there is zero current (Id=0) through the inverter, which is one of the key properties of CMOS digital logic.



The case for Vin=VDD (logic ‘1’) is shown below. The output is 0V (logic ‘0’).

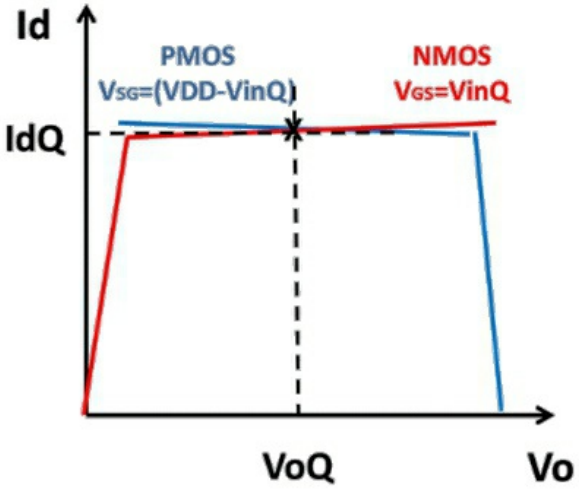


The digital inverter that we are so familiar with is in fact going to be our stepping stone into the world of Analog. Let us hold the below image in mind as we look to understand its Analog nature in greater detail.


Id
PMOS
NMOS
VDD/2
VDD
Vo

Synthesizing our first Analog buffer –the Common Source Amplifier

Let us consider the case where we have given an input VinQ such that the output VoQ is roughly at VDD/2 – what we have defined as our 'ideal' operating point. Earlier we had assumed that this happens at Vin=VDD/2 but as stated, it is only a simplification.



The suffix Q (that we added to Vin and Vo) stands for **Quiescent** point which depicts a state of sleep. What it actually signifies is that the circuit is in a state where it is sitting in wait to respond to a 'signal'. Here a 'signal' is viewed as a disturbance of the circuit around its quiescent state.

As we have seen in the previous chapter, the value or range of Vin over which the circuit is near its ideal desired operating point is a very small one. We saw that how shifting to one of the neighbouring Vin curves caused the operating point to move to either of the extremes. In fact, one of the key challenges of Analog design is to get each transistor to operate in its desired operating zone – usually the saturation region. The method to get all transistors to operate at their desired operating zones is what is called **Biasing**.

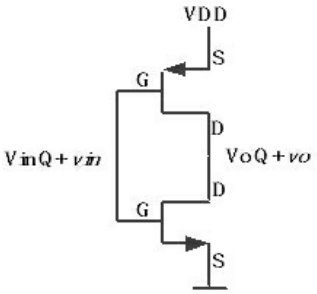
We will assume that we have somehow established the ideal operating point as shown above. Now let us assume there is a small signal applied around Vin=VinQ. This can be thought of as an increment *delta*(Vin) applied around VinQ.

Vin = VinQ+*delta*(Vin)

The change in Vin will cause a change in Vo. We will refer to this change as *delta*(Vo).

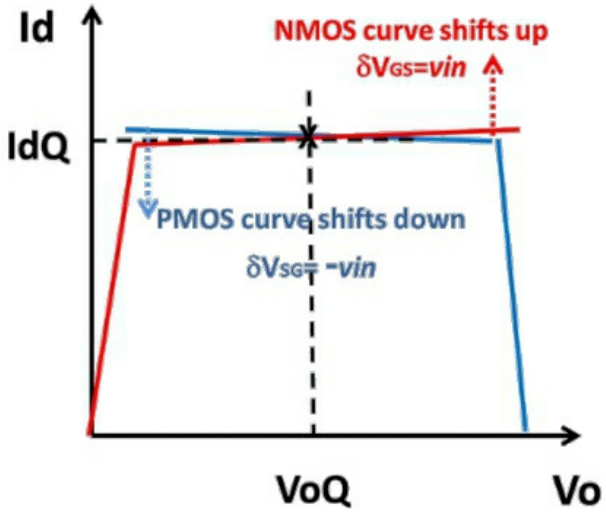
Vo = VoQ+ *delta*(Vo)

Let us replace *delta*(Vin) by *vin* and *delta*(Vo) by *vo*. Here the notation of capital letters (VinQ, VoQ) suggest Quiescent points and the small letters (*vin, vo*) stand for increments around these quiescent points. So *vo* is the incremental change in Vo around VoQ when there is an incremental change of *vin* for Vin around VinQ. This is depicted in the circuit shown below.

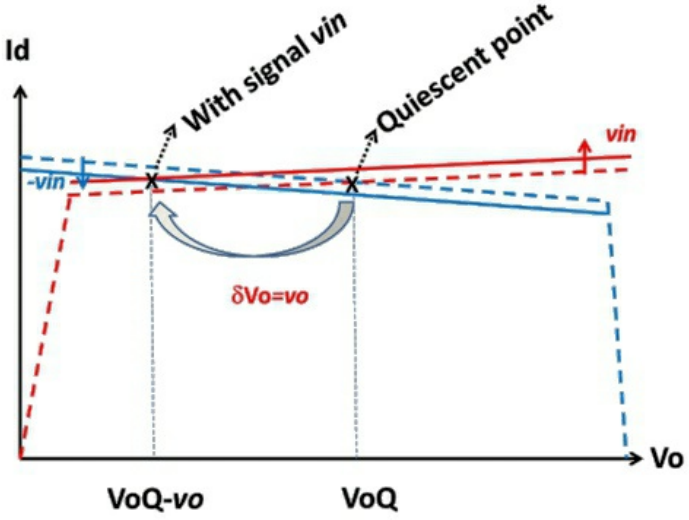


How would we go about figuring out *vo* for a certain *vin*?

First note that an increment in Vin (equal to *vin*) will cause the NMOS to shift to a slightly higher curve. The PMOS however would shift to a slightly lower curve because its controlling voltage reduces by *vin*. This is pictorially depicted as below.



The new intersection point shifts as shown below.



There is a lot happening in the above figure, so let's go through it step by step:

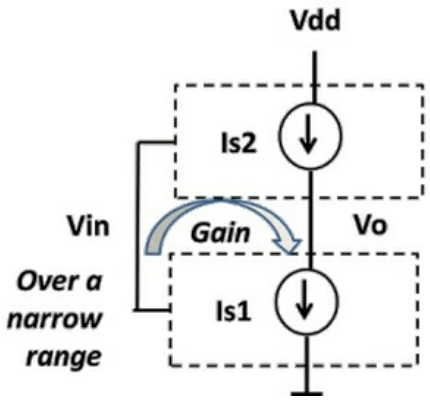
The original curves of the NMOS and PMOS at the Quiescent point are shown by the dotted lines. The new curves that the NMOS and PMOS shift to with a signal *vin* (an increment of Vin around VinQ) are shown by the solid lines. As explained before, the NMOS shifts to a higher curve whereas the PMOS shifts to a lower curve. The new intersection point (which determines what will be the new value of Vo because of the application of the signal) is given by the intersection of the solid lines.

As can be seen, even though we have assumed a small change in Vin (a small signal *vin*), the output has changed by a relatively large amount. This is suggestive of a high 'gain' between the input signal and the output signal. The active circuit has responded to the incremental stimulus (*vin*) with a gained up increment in the output (equal to *vo*)! Also note that the increment in Vin has resulted in a decrement to Vo. So the gain is high and has a negative polarity.

If you examine the curves closely, what has caused the gain to be high is the fact that the curves are nearly horizontal (low slope) around their original operating point. If you were to make the slope lower and lower (so that the saturation regions of the PMOS and NMOS curves start to become almost horizontal), you can verify for yourself that the gain will get larger and larger.

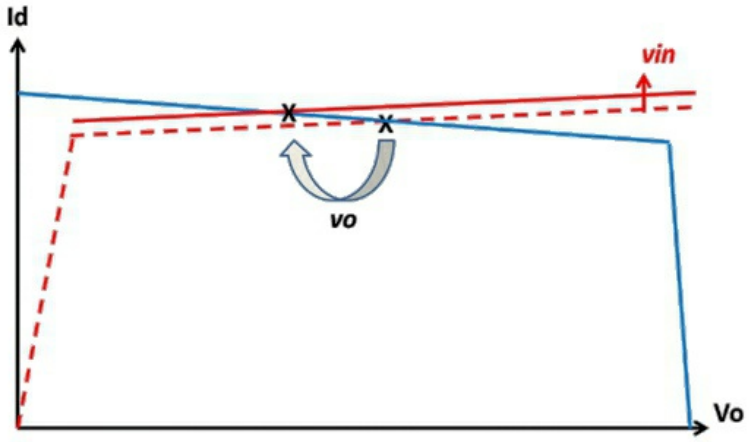
The other way of stating this is that the mechanism of gain is made possible by having two nearly ideal (controlled) current sources (the PMOS and

NMOS) connected to each other. We have earlier seen that it is tough to get such two elements to co-exist peacefully and establish a proper operating point. What we have analysed so far is how they behave when they somehow operate in the narrow region (range of Vin) where they are indeed able to co-exist and produce gain. This is pictorially depicted as shown below.



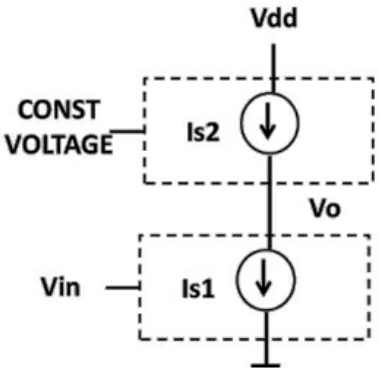
The Digital inverter is thus an element capable of operating like an Analog circuit that can provide a high gain between its input and output. This is not to say that this circuit is a feasible Analog circuit. There is a lot we will need to do in order to get to a circuit that has all the key attributes (a proper operating point and high gain around it) and is able to operate in a practically realizable manner.

Firstly, let us note that, to achieve gain, it is not required for both curves to respond (shift) to the input signal. For example, it would have been sufficient had only the NMOS curve shifted. This is shown below – the PMOS curve is shown unchanged with signal.

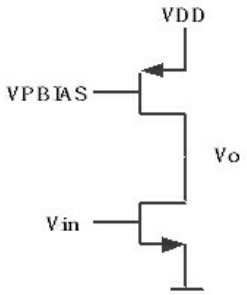


As can be seen above, signal amplification still happens (although only to

about half the extent as compared to when both the curves shifted). This scenario can be depicted by a constant current source (PMOS) connected to a controlled current source (NMOS). The way to make the PMOS behave like a constant current source independent of *vin* would be by making sure that the input signal does not change the controlling voltage of the PMOS. One way is by connecting its gate to a constant voltage. This is depicted below:



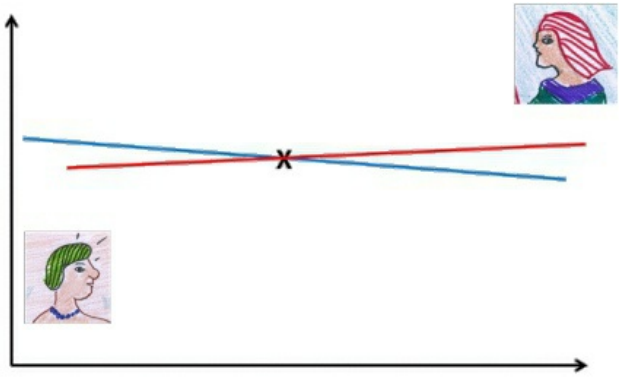
This leads us to our first “Analog” circuit!



The above circuit is called the **Common source amplifier**. The words “Common source” qualifies the fact that the source terminals are connected to a constant potential (for the NMOS, this potential is ground and for the PMOS it is VDD). In other words, both the input and output are referred to the common terminal which is the source.

Making one of the transistors (PMOS in the above case) behave like a constant current source makes things a bit easier (not a whole lot!) in terms of getting to the ideal operating point. For example, you can envision some kind of control that comes and adjusts the PBIAS voltage magically such that the point of intersection is at the desired VoQ. But we are getting ahead of ourselves. So we will now look at how our protagonists are doing.

As stated earlier, Aman-Ra and Uman-Ra have recently settled to their life of marital bliss. Remember their ‘operating points’ that we earlier saw?

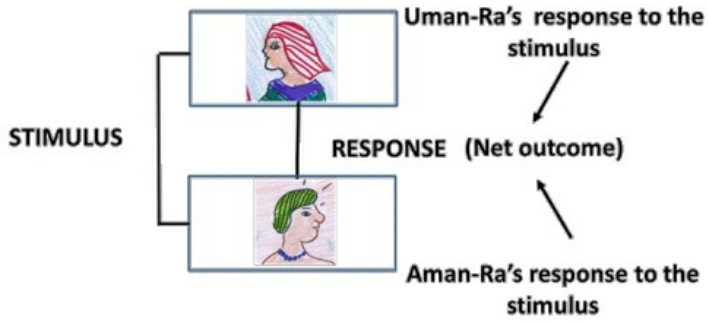


Now that we know that the right terminology to describe their meeting point is ‘Quiescent’ point, let us see what happens when we apply a ‘signal’. What could possibly constitute a ‘signal’ around their Quiescent point? The couple has rented a single bedroom apartment in downtown Cairo and have so far been living a life of undisturbed togetherness. What we need to mimic a ‘signal’ is therefore a ‘stimulus’ – what they would consider a ‘change’ relative to their routine.

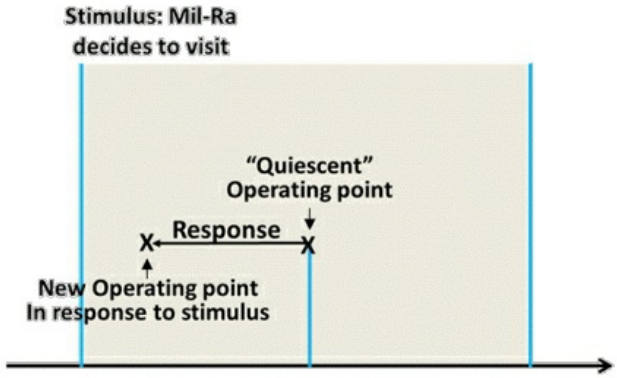
Enter Mil-Ra. She is the mother of Aman-Ra and therefore the mother-in-law of Uman-Ra. Even back in those days, I am, guessing there would have been the concept of ‘in-laws’, so the entry of Mil-Ra should not come as a huge surprise to the readers- neither should the fact that Mil-Ra is not really an easy person to please!

Mil-Ra has a laundry list of requirements that she expects Uman-Ra to fulfil daily for her.

The sudden entry of Mil-Ra can now be modelled as shown below:



Note that this stimulus is applied 'around' the Quiescent point – the equilibrium state that existed between husband and wife before the mother-in-law made her entry. The response to this stimulus is shown abstractly as below.



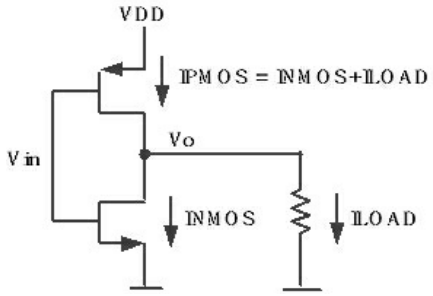
We have shown above a 'response' that falls within two 'boundaries' – again a vague representation. We shall elaborate on the 'boundaries' in a bit. But the point to be noted is that the response is shown to fall well within the boundaries. In that sense, it is a healthy response to the stimulus. Over the duration of Mil-Ra's visit, the couple takes a joint decision to wake up an hour earlier than usual. They both use the extra hour to do the preparations needed to satisfy all of Mil-Ra's requirements over the day. What made such a mature and balanced response to what could have been an unpleasant stimulus was the fact that to begin with, their operating point was nice and centred. A happy operating point lead to a healthy response.

We are yet to elaborate on what we mean by 'boundaries'. It is easier to explain it by taking two extreme situations. These would be akin to the way a digital circuit would be operated and how it would respond. The equivalent of a logic '1' at the output would be something like the below - here, Uman-Ra has given up her CFO job and decides to serve Mil-Ra instead.

What would be the equivalent of a digital ‘0’? The below perhaps…?

Why the Common source amplifier is not a good buffer – the problem of loading

Till now, there is one important factor we are yet to consider – Loading. Let's say we load our Inverter with a load by connecting a resistor to its output.

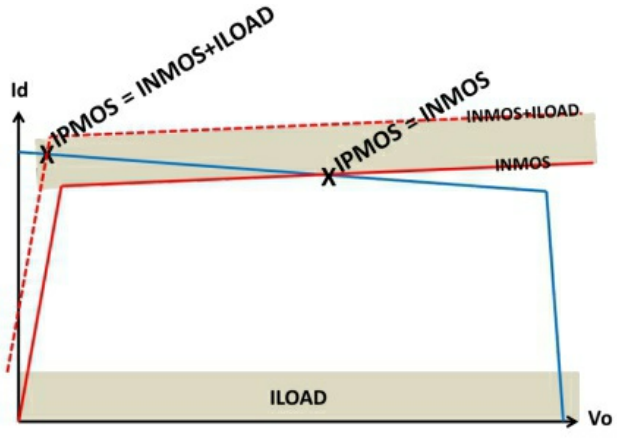


When we did not have the load resistor, the IPMOS and INMOS currents were equal. This was implicitly captured in the way we went about graphically figuring out the Quiescent point – we merely looked for the intersection of the PMOS and NMOS curves.

With the load, however, the PMOS current should be equal to the sum of the NMOS current and the load current.

IPMOS = INMOS+ILOAD

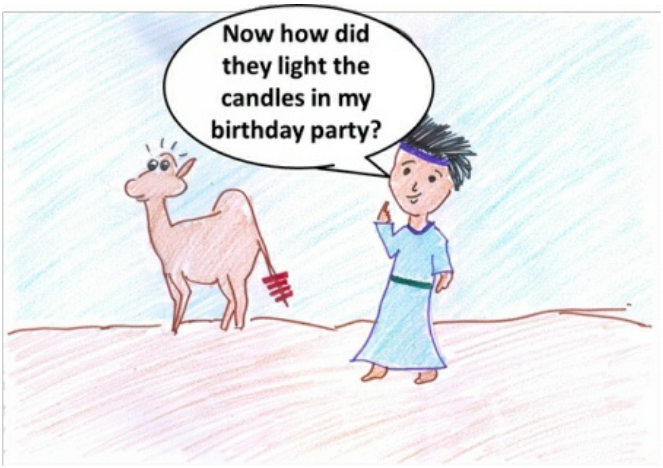
What this means is that we now have to first add ILOAD to the INMOS curve and then see where this new curve intersects the IPMOS curve.



This is shown above. The INMOS curve (shown as the solid red line), when shifted by ILOAD results in the dotted red line. The intersection of this dotted line with the IPMOS curve (solid blue line) gives the new operating point. In the illustration shown, the operating point is shifted hugely from its

ideal operating point. This is not hard to see why. The load has a similar effect on the operating point as a signal – it shifts the delicate balance between the two well-matched current sources and ends up disturbing the operating point quite a bit.

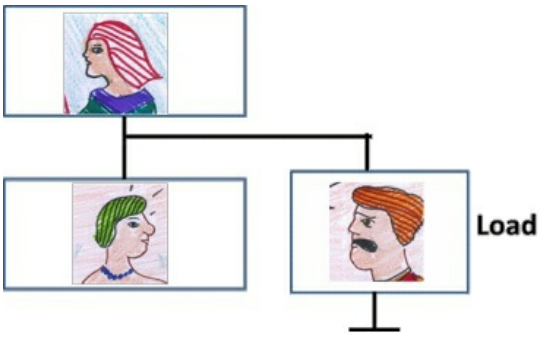
Let us return back to the world of Aman-Ra and Uman-Ra. It has now been seven years since they tied the knot. Some new factors of ‘loading’ have made their entry into their lives. One of them is Lil-Ra, their son, now six years old.



Tearing up Aman-Ra’s datasheets, hiding Mil-Ra’s pet snake inside his mother’s fan,… all in a day’s work for the hardworking Lil-Ra.

There are other serious loading factors that are testing the relationship. Aman-Ra’s boss Gamen-Ra has started to become a pain. One of Aman-Ra’s pyramids designs has collapsed on top of its intended inhabitant, and Aman-Ra is now under his manager’s close scrutiny. His project is being closely tracked by Gamen-Ra and he has to give him daily progress updates.

The below illustration depicts the additional loading from Gamen-Ra in the couple’s relationship.



It is with this backdrop that the stimulus returns. Mil-Ra has decided to pay another visit!

It is not difficult to see that the relationship is much more strained (with all the new loading factors) than earlier. It is unlikely Mil-Ra’s visit will be handled with the same equanimity by the couple this time!

The loading that shifted our Common source amplifier from its ideal operating point has much the similar effect on the relationship of Aman-Ra and Uman-Ra! Did I not tell you that Analog mimics the real world?!

The circuit trick that solves the problem of loading

Struggling to keep harmony in their lives, the couple has an unexpected visitor from a faraway land.

There are many legends about Ang-Lao. Some say he is a mere boy who had attained enlightenment while still a baby. Yet others say that his face lies about his true age - that he is several hundred years old and has been traveling the length and breadth of the world spreading his message of love and peace.

One stormy night, Ang-Lao visits the village of Aman-Ra. Seeing the baby-faced ascetic sleeping on the roadside under the pouring rain, Aman-Ra invites him home. The couple takes a huge liking to Ang-Lao and shower him with great care.

That night, Ang-Lao witnesses a stormy argument between Aman-Ra and Uman-Ra. He reflects on the fact that they each seem to be such nice and generous human beings. Yet, their relationship seems to be breaking under all the strain it is being subject to. The topic of the night's altercation is Mil-Ra's increasing demands from her daughter-in-law.

The next morning, as Ang-Lao gets ready to leave the village, he addresses the couple before parting…

"It is with great sadness that I witnessed your altercation yesterday night. I wish to tell you a secret about relationships."

It was a secret that had come to Ang-Lao whilst he lay contorted in the midst of a complex yoga postures – Pranalogasana. A disclaimer to the readers of this book is not to attempt to try this posture except under the guidance of an Analog Guru.

Ang-Lao reveals the secret…

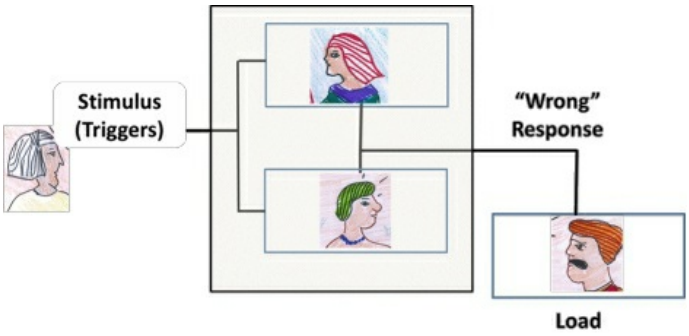
"In a relationship, you may not be able to react with the right response every time…"

Ang-Lao takes a long and dramatic pause and continues…

"***… but you always know!***"

And with these final parting words, Ang-Lao takes leave of the couple blessing them with peace and contentment.

The couple break their head over Ang-Lao's cryptic statement for several days. It triggers in them a process of self-reflection. They start with a drawing of their current state of affairs.

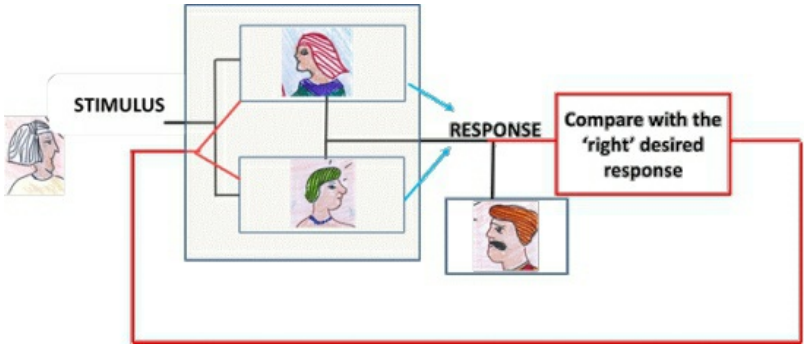


They then ask each other "*Do we really know what the 'right' response is?*" They reflect on the relatively hassle-free visits of Mil-Ra in the initial phase of their married life. They realize that they indeed know what is the 'right' way to respond to her idiosyncrasies. Yet they reflect that of late, they have been ending up reacting in all the wrong ways.

It is at this point that Aman-Ra and Uman-Ra take a giant leap in their relationship. They reason that if each reflected on the 'wrong' response (which has now become their automatic reaction), they would at once realize that it is a deviation from the 'ideal' response. What would then be required

to get to the right response would be a slight adjustment to the thought process of each (a slight shift in their ‘curves’) to be able to achieve the right response.

And thus, the couple stumbles upon one of the most important concepts of Analog design. It is contained in the additional red lines drawn below.

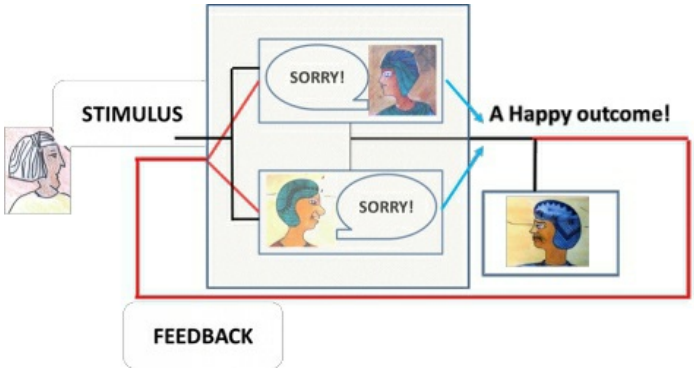


What the couple figures out is that two things are required from them in order to achieve the ‘right’ response. They codify it succinctly as:

1. *Thou shalt reflect*
2. *Thou shalt correct*

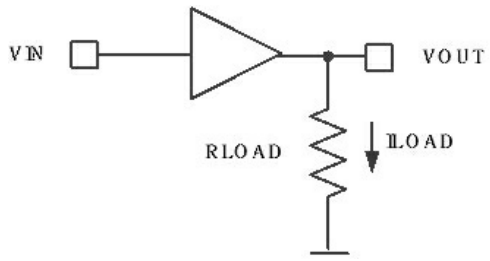
And thus our couple figures out the pivotal concept of Analog – **Feedback**!

In the particular case involving the reaction of each to Mil-Ra’s visit, a simple reflection on their initial behaviour followed by a simple correction to their thought process results in the below outcome…



We shall now see how the above concepts apply to our buffer.

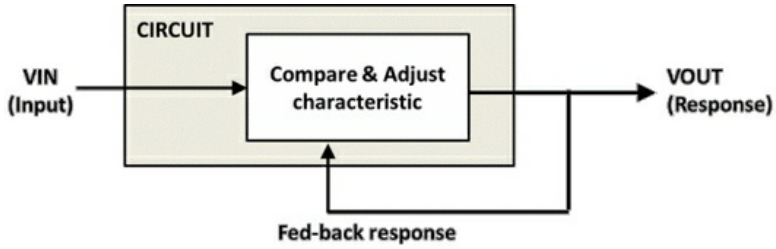
We had depicted a buffer as shown below:



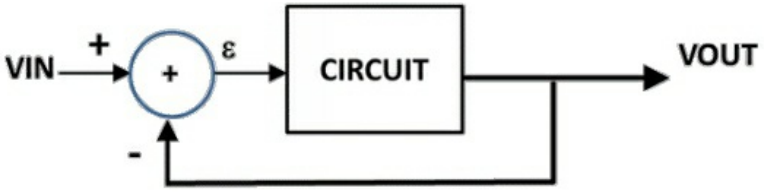
But does our buffer know what is the ‘right’ response to a stimulus?

In the case of the ideal buffer, the answer is straightforward. In fact we defined an ideal buffer as one where VOUT=VIN. So the ‘right’ response to a stimulus VIN is nothing but a response which is … **VOUT=VIN!**

From the insight we got from our protagonists, it is now possible to guess what is needed to realize an ideal buffer. It is basically a mechanism to look at VOUT and change the characteristics of the constituent elements so that VOUT becomes equal to VIN! A scheme is shown below:

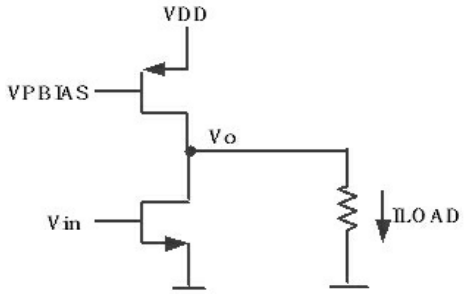


What we have drawn above is actually a more illuminating way of depicting how feedback actually works in the context of a circuit. The ‘textbook’ way is shown below for comparison. Here, Epsilon is the ‘error signal’ that causes the circuit to change its characteristics so that VOUT can match Vin.



We are now ready to build our first ‘practical’ circuit!

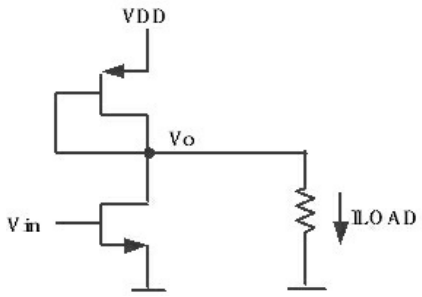
We earlier saw how the operating point of the Inverter was greatly affected when loaded. A similar analysis reveals that the operating point of the Common source amplifier shown below is also greatly affected when loaded.



If we recall our earlier discussion, the problem with loading was the following – the PMOS transistor behaved like a near-ideal current source. The NMOS transistor behaved like a voltage controlled current source and shifted to a different curve when a signal was applied on Vin (or) when Vo was loaded. As a result, the intersection point of the two current sources dramatically shifted away from the ideal operating point. We called the above circuit as a 'Common source amplifier'. But it can be described more accurately as a 'Common source amplifier with a current source load'.

Given our understanding of what feedback does, can we now come up with a small change in the above circuit such that the operating point does not shift much even with loading?

Clearly, a mechanism of *reflect and correct* is needed! Either the PMOS or the NMOS needs to look at the output and modify its characteristics so that the circuit is able to stay close to its ideal operating point. If you look at the available terminals that can 'look' at the output and change its characteristics, there is only one possibility – the Gate of the PMOS. Let us translate the requirement of '*look at the output*' quite literally as '*connect the Gate of the PMOS to the output*'. This leads us to the below configuration. We still do not know if it will work but it is the most literal interpretation of what we got out of Ang-Lao's wise words.



Here, the information of Vo gets passed to the Gate of the PMOS transistor. Since the gate is one of the controlling terminals of the PMOS (the other

being its source), the PMOS is now able to ‘react’ to the voltage on Vo and ‘shift’ its characteristics in response to it. In other words, we have now built in a mechanism of ‘*reflect and correct*’ into the circuit!

While it is clear that such a mechanism has now been built in, what is less obvious is the fact that the above configuration actually maintains a proper operating point in the presence of loading. For now, we will just state that in the above configuration, the PMOS transistor is guaranteed to operate in saturation (because its terminal voltage is equal to its controlling voltage and hence large enough to maintain it in saturation). Additionally, if the controlling voltage (which is equal to its terminal voltage since Gate and Drain are shorted) of the PMOS is not too large, then there would be enough voltage margin to keep the NMOS transistor also in saturation. So for now, we will assume that both the transistors are in saturation. That is a big part of the problem solved. Also the controlling voltage of the PMOS transistor is no longer fixed. It can change based on Vo and this gives the circuit a key additional knob to allow IPMOS to become equal to (INMOS+ILOAD).

The PMOS transistor connected in the above manner (with its drain and gate connected) is referred to as a **Diode-connected transistor**. So this topology is referred to as a ‘Common source amplifier with a diode-connected load’. With the drain and gate shorted, the PMOS has now become like a two terminal device and behaves somewhat like a diode – hence the name. By making this simple change to the circuit, we have established a circuit that will work without much fuss.

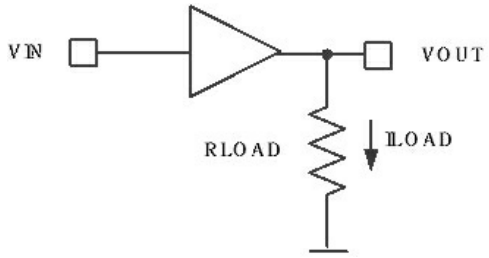
A human analogy is useful to share at this point. The PMOS transistor connected in the Diode connected manner is like a noble soul! It does not look to ‘impose’ its own current. Rather it adjusts its characteristics so that harmony is established in its relationship with the NMOS transistor – its behaviour is much like the selfless action that one associates with the mother of a child!

We stated without any justification that the above configuration can maintain a decent operating even in the presence of loading. The simple way of looking at it is that earlier (when the PMOS gate was connected to a fixed voltage PBIAS) the PMOS current was constrained to be closely around a certain value. However, now the PMOS current has one more degree of freedom to change itself since its Gate is now tied to a voltage (Vo) that can

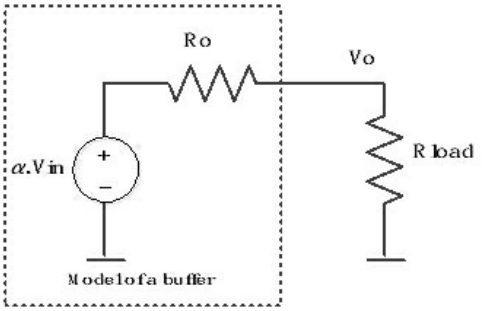
change. So by small adjustments to Vo it can react to an incremental load current.

When we dig deeper into the reason why this circuit is robust to loading, some very insightful facts will emerge.

Firstly, we need to look at what is the property of a buffer that will enable it to be loaded and still maintain its output voltage without much change.



The relevant attribute of such a buffer is **Low output impedance**. A buffer maybe modelled as shown below using a voltage controlled voltage source(□ .Vin)in series with an output impedance Ro. The ideal buffer is one where □ =1 and Ro=0.You can verify for yourself that with these values for □ =1 and Ro, the output Vo is equal to the input Vin irrespective of Rload.



The output impedance (Ro) of a non-ideal buffer is non-zero. The voltage at Vo can be thought of as a voltage division between Rload and Ro. Smaller the Ro in relation to Rload, the more will be the ability of the buffer to maintain its operating point in the presence of loading.

If a buffer is indeed able to maintain its operating point in the presence of loading, it should imply that its Ro should have been small. What that further implies is that the circuit should have a behaviour at the output which is more like a voltage source than a current source. Till now, we have been talking about the MOS as a current source, which by definition has a high output impedance. So how did our simple change (of shorting the PMOS gate and

drain) give it a voltage-source like property at the output?

Clearly, the NMOS transistor configuration is unchanged and it continues to behave like a voltage controlled current source (controlled by the voltage Vin). What about the diode-connected PMOS transistor?

Let us take a step back to see whether (and when) a MOS transistor can behave as a voltage source.

We have looked at the current equation of the MOS transistor. With small changes to the terminal and controlling voltages, the same equation holds true for the PMOS transistor as well.

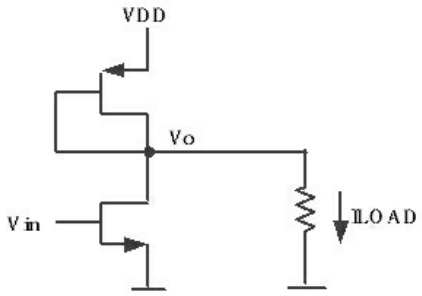
$$Id = K.(1+\lambda.V_{DS}).(V_{GS}-Vt)^2/2$$

Ignoring the weak dependence on VDS, we can rewrite the equation as:

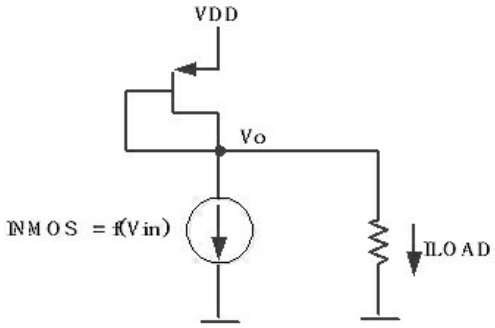
$$Id \sim K.(V_{GS}-Vt)^2/2$$

In more general terms, Id = f(V_{GS}).

Now, let us take a closer look at the circuit we came up with.



For the NMOS transistor, we are forcing its Gate and Source. So we are forcing its V_{GS} to Vin. Therefore, it is constrained to take a current of Id=f(Vin) provided of course that it is operating in the saturation region. The NMOS transistor in the above circuit therefore behaves like a (controlled) current source. We can depict the circuit as shown below:



What about the PMOS transistor? If we ignore the load for now, the NMOS transistor is more or less setting the current through the PMOS transistor. However, while the PMOS source is forced to VDD, its gate is free to adjust itself. So for the case of the PMOS transistor, the current Id is forced and the controlling voltage can be thought of as the parameter that will look to adjust itself to be able to carry that value of current.

So for the PMOS transistor:

V_{SG} = g(Id) where 'g' is the inverse function of 'f'. The function g(Id) is the one shown below (for the PMOS transistor).

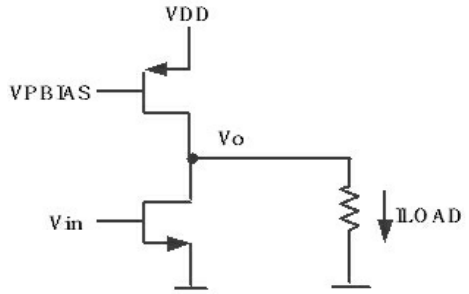
$$V_{SG} \sim Vt + (2.Id/K)^{1/2}$$

So for the PMOS transistor in the above case, the current Id forced through it dictates its V_{SG}. Since the source node of the PMOS is driven to VDD, the current is therefore 'dictating' the voltage on the Gate node, which is nothing but the output Vo! So the PMOS behaves like a **Current-Controlled Voltage Source** at Vo!

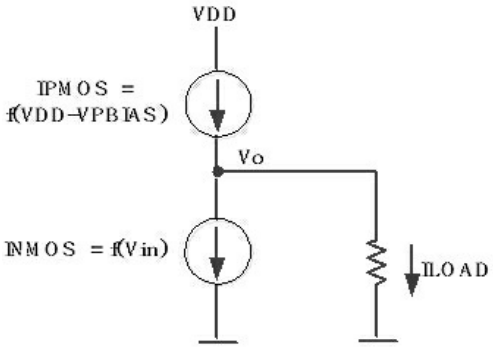
Whether a transistor in an analog circuit behaves like a current source or a voltage source is determined by whether what is forced is the controlling voltage or the current through it.

When V_{GS} is forced, the transistor behaves like a current source at its Drain. When Id is forced, the transistor behaves like a voltage source between its Gate and Source.

The reason why our original Common source amplifier (shown below) was so difficult to make functional was because the controlling voltage was forced in both the transistors – therefore both tried to behave like current sources!

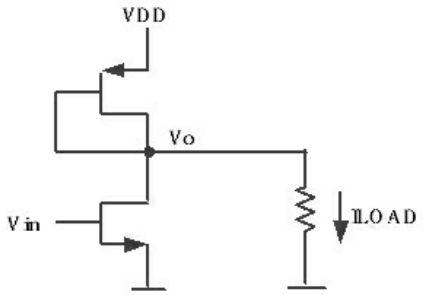


This is depicted below.

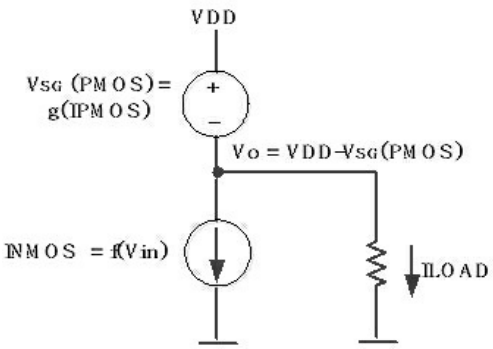


The conflict is obvious to see.

However, once we transformed the PMOS transisitor into a diode-connected transistor, its controlling voltage was free to adjust to the current that the NMOS transistor was trying to force through it.



As a result, what we ended up with was a circuit that looked like below:



Here IPMOS=ILOAD+INMOS is the current that gets forced through the PMOS and it develops a V_{SG}=g(IPMOS) where 'g' is the inverse function of 'f'. The PMOS is now what determines the output voltage.

Now let us analyse what happens to Vo when Vin changes by a small amount *vin*. Let us assume that the NMOS and PMOS transistors are identical. What we mean by this is that for the same current through both the transistors, their controlling voltages are the same- in other words, the V_{SG} of the PMOS transistor is equal to the V_{GS} of the NMOS transistor. If Vin increased by *vin*, then the current through the NMOS transistor would change. To be able to take the changed current, the V_{SG} of the PMOS would also have to increase by *vin*. This would require Vo to **reduce** by *vin*. In other words, the gain for an increment in input signal is -1. We will for now ignore the inverted polarity of gain and will consider this circuit as an approximation to an ideal buffer.

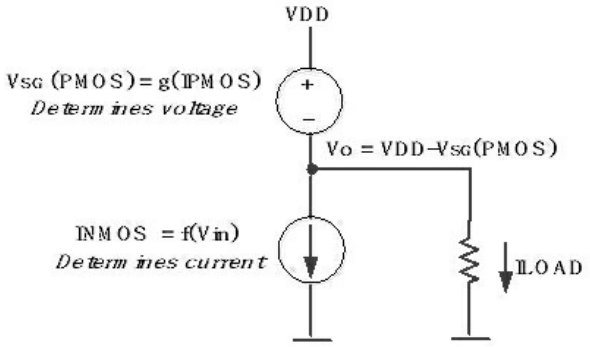
What makes the above circuit a practically realizable one is the fact that the 'voltage-source' like element (PMOS in this circuit) can happily co-exist in series with the 'current-source' like element (NMOS in this circuit). This leads us to two observations:

(i) This is what makes the above circuit robust in its operating point.

(ii) This is what makes the above circuit a good buffer – namely one which has a low output impedance (because of the voltage-source like attribute of the PMOS in this circuit)

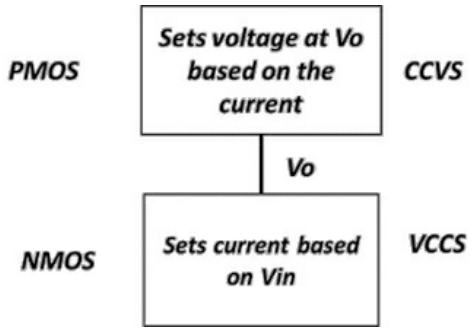
The analysis of when a transistor behaves like a voltage source and when it behaves like a current source can be a powerful synthesis tool to synthesize new circuit topologies. Let us take an example to illustrate.

If you recall the common source amplifier with the diode connected PMOS load, it looked like the below:

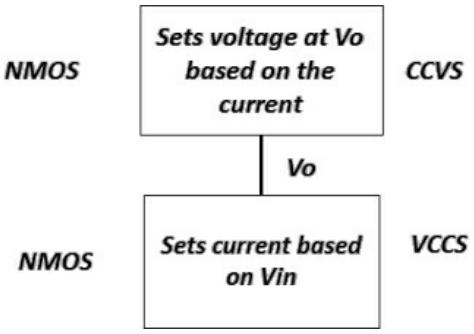


Here, the NMOS transistor was generating a current depending on Vin and the PMOS transistor was adjusting its controlling voltage (V_{SG}) based on the current.

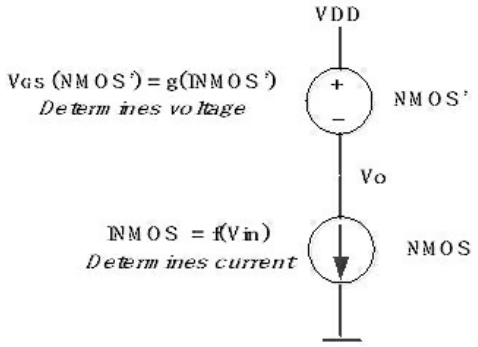
We can depict the role of the two transistors in this circuit as:



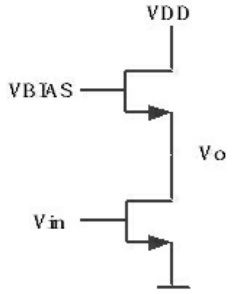
What other current source-voltage source topologies can we come up with? For example, can we realize the CCVS on top with a NMOS transistor – something like what is shown below?



Seen from the perspective of the controlling voltages and currents, the picture we want to get to is the one shown below:



While it is fairly obvious that the lower NMOS needs to be connected in the same manner as the previous circuit, how do we realize the upper NMOS? Clearly, it needs to behave like a voltage source; so we should make sure we keep at least one of its controlling terminals (either gate or source) flexible and not forced. You can reflect on the various possibilities and should be able to arrive at the one shown below.

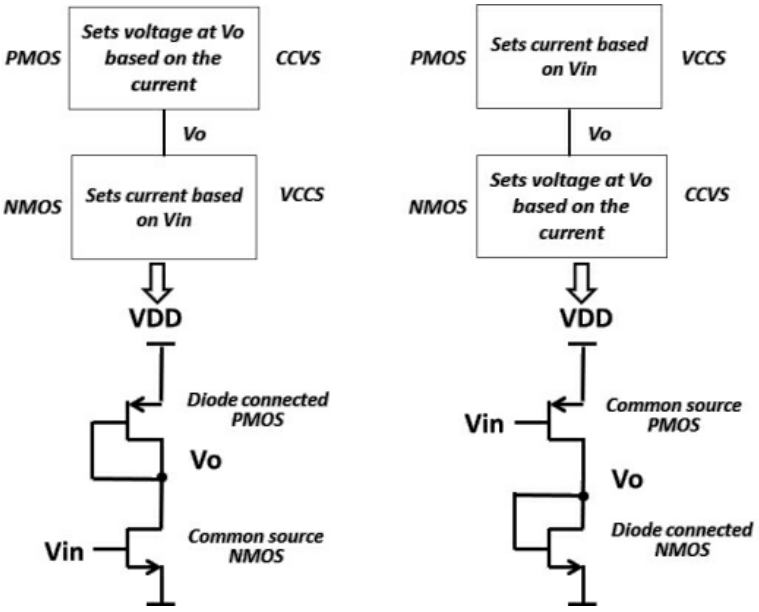


Here, the gate is forced to a constant voltage called VBIAS and the source is flexible to adjust its voltage. It should be easy to see that the top NMOS behaves like a CCVS and imposes the voltage on Vo.

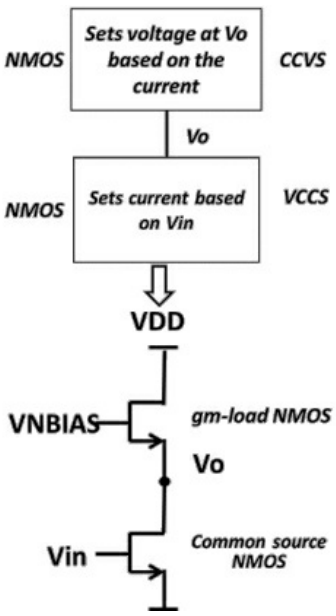
In this configuration, Vo will be equal to VBIAS minus the V_{GS} of the NMOS transistor. It is evident that the top transistor behaves like a voltage source (with the voltage being controlled in some way by the current flowing through it). But what determines the relationship of Vo to Vin? If you assume the two transistors are identical, then to a first order, their V_{GS} should be the same at all values of current. Now if Vin changes by a certain amount, it causes the V_{GS} of the bottom transistor to change by the same amount. This causes the current through both the transistors to change. For the top transistor to take the changed current, its V_{GS} change would have to track the V_{GS} change in the bottom transistor. This would require Vo to change by the same amount as Vin but in an opposite direction. For example, if Vin changes by *vin*, then Vo will have to change by *–vin* so that the V_{GS} of the two

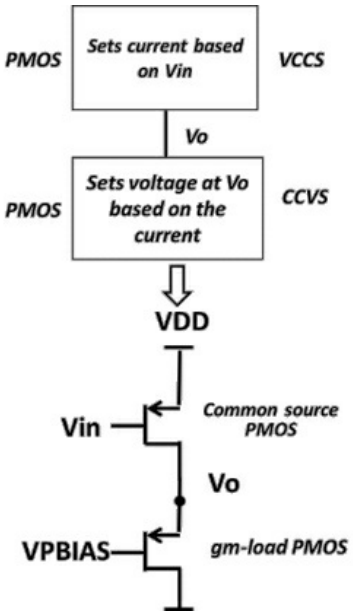
transistors track. So this circuit also has a gain of -1 between Vin and Vo. The NMOS transistor that we have stacked on top in this manner is referred to as a **gm-load**. We will see the significance of this terminology in a later part of this series. So this topology is referred to as a ‘Common source amplifier with a gm-load’.

In the two circuits synthesized so far, the roles of the top and bottom transistors could well have been reversed. The first one corresponds to the common source amplifier with the diode connected load as shown below.



The second one corresponds to the common source amplifier with the gm-load as shown below.


NMOS
Sets voltage at Vo based on the current
CCVS
Vo
NMOS
Sets current based on Vin
VCCS
VDD
VNBIAS
gm-load NMOS
Vo
Vin
Common source NMOS

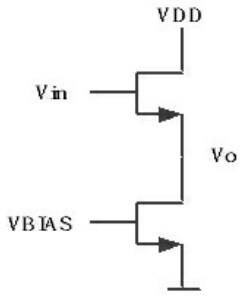

PMOS
Sets current based on Vin
VCCS
Vo
PMOS
Sets voltage at Vo based on the current
CCVS
VDD
Vin
Common source PMOS
Vo
VPBIAS
gm-load PMOS

We are now ready to build our first practical buffer – the Source follower

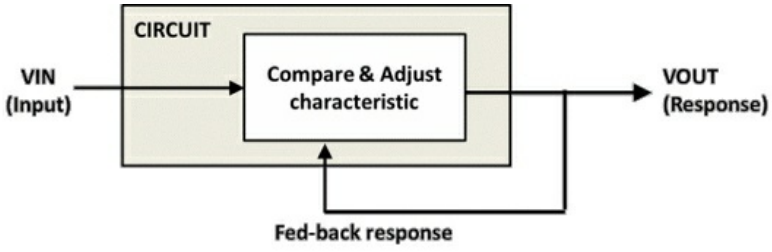
Let us try and synthesize one more topology. This one is a little less obvious.

This is a topology where the NMOS transistor below behaves like a constant current and we have another transistor sitting on top of it that changed its characteristics based on both Vin and Vo. The requirements of such a circuit would therefore be that the transistor on top should not have its *controlling voltage* imposed on it – rather be free to adjust itself to take the current set by the lower NMOS. A bit more thought will lead us to the fact that the transistor on top also needs to be an NMOS transistor.

Such a topology would result in the below circuit.



It is tougher to see feedback at play in the above circuit. Unlike the first circuit which had the diode connected PMOS, there is no explicit connection that indicates that Vo is being fed back. However, feedback is very much at play even here. Vo is the source potential of the upper NMOS and therefore a contributor to its *controlling voltage*. The upper NMOS is therefore able to change its characteristics based not only on Vin but also on Vo. If you reflect our model for feedback (shown below), it will become clear that this circuit also has feedback at play.



The above circuit is called the **Source follower** circuit. It is a very handy circuit and in fact, our first approximation to a buffer with a gain of +1. This

can be seen by looking at the relation between Vin and Vo.

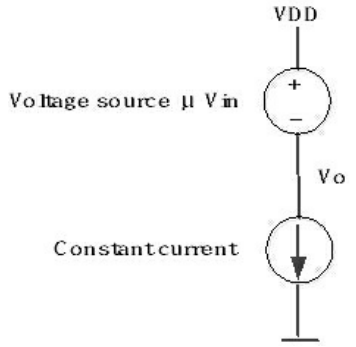
V_{GS}(Upper NMOS) = Vin – Vo

This can be rewritten as:

Vo = Vin – V_{GS}(Upper NMOS)

The lower NMOS has both its gate and source tied to constant voltages. So to a first order, it is a constant current source. Assuming that even with a change in Vin, the lower NMOS current is unchanged, the V_{GS} of the upper NMOS is also to a first order unchanged with Vin. Therefore Vo (the Source) 'follows' Vin (the Gate) with a fixed voltage difference – which gives the circuit the name 'Source follower'. In reality, when Vo follows Vin, it would cause a change in the V_{DS} of the Lower NMOS and would result in a slight modulation of its current. But to a first order, the above assumption that the source voltage *follows* the gate voltage holds.

The equivalent circuit of this topology would in fact look like below.

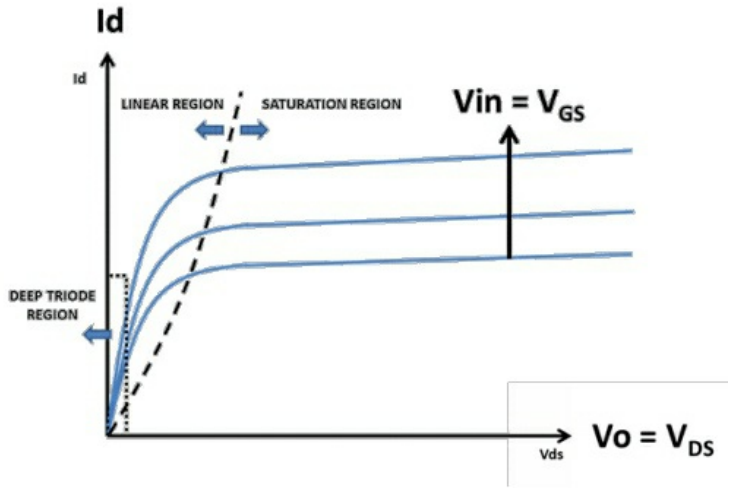


Again, the co-existence of a voltage source and current source is a happy one, and the Source follower operates robustly over a wide range of signal and loading.

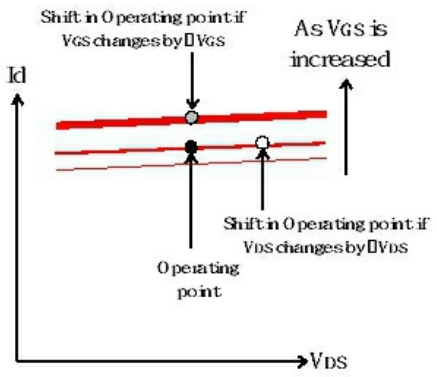
We have thus synthesized three simple circuit topologies that are practically realizable using the simple premise of getting a voltage source to co-exist with a current source. All these three circuits constitute reasonably good buffers. They have low output impedance (voltage source-like characteristics at the output) and their output follows the input with a gain of almost 1.

Analysis of MOSFET small signal parameters to quantify how good our buffer really is

Having synthesized a Source follower buffer, we need to next see how good a buffer it really is and what should be our target for improving on it. To see this, we need to first understand a couple of parameters related to the MOSFET. Let us go back and revisit the MOSFET I-V curves.



Assuming we are operating in the saturation region (where the MOSFET operates either like a Voltage controlled current source or a current controlled voltage source depending on how you connect it), there are 2 parameters of interest: ***gm*** and ***gds***. These 2 parameters are referred to as small signal parameters since they refer to the response of the MOSFET to a small change applied around the operating point.



From the illustration, we can see that if V_{DS} changes by a small amount, the change in current is very small – this is because of the very gentle slope of the Id-V_{DS} curves in the saturation region. This change is depicted by the small-signal parameter called ***gds***.

Delta Id = gds* (Delta V_{DS})

The above equation depicts the change in Id if only V_{DS} is changed (V_{GS} is kept constant). So the operating point moves along the same V_{GS} curve.

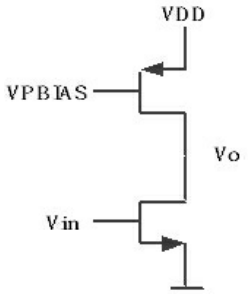
If V_{GS} changes (V_{DS} remaining constant), then the operating point would 'jump' to a V_{GS} curve corresponding to the new V_{GS}. Since Id has a strong dependence on V_{GS} in the saturation region, the change in Id due to a change in V_{GS} would be relatively high. This is depicted by the small-signal parameter called ***gm***.

Delta Id = gm* (Delta V_{GS})

Here again, the equation depicts the change in Id if only V_{GS} is changed (with V_{DS} remaining constant). So the operating point jumps vertically.

Without getting into numbers, it should be clear that for a MOSFET operating in saturation, gm should be much higher than gds. In fact, the ratio of gm/gds is a measure of the inherent 'gain' of the transistor. The exact value of gm/gds depends on the Silicon process and the type of transistor, but for now let us think of a number in the range of 10-20.

Let us revisit the common source amplifier and convince ourselves that the 'gain' from input to output is indeed set by the factor 'gm/gds'.



If Vin increases by a small amount, it directly increases the V_{GS} of the NMOS. So the NMOS tries to respond by increasing its Id (through the factor, gm). Let us for a moment assume that the PMOS is an 'ideal' current source that imposes a constant current on the NMOS. So the only knob left for the NMOS to be able to offset the current increase due to increase in its V_{GS} is by reducing its V_{DS} (this causes a current change in the opposite direction as set by the factor gds). However, as we have seen, gds is much smaller than gm. So a small increase in V_{GS} has to be offset by a large decrease in V_{DS}.

gm * (Delta V_{GS}) = -gds * (Delta V_{DS})

This leads us to:

Delta V_{DS} = -(gm/gds)*(Delta V_{GS})

Recognizing that Delta V_{GS} and Delta V_{DS} are nothing but the change in Vin and Vout, we get:

Delta Vout = -(gm/gds) * (Delta Vin)

Since (gm/gds) is a factor usually >>1, we can now see the gain that the Common source amplifier is capable of giving is in fact the inherent gain of the MOSFET.

A small detail we have left out is the role of the PMOS, which is not an ideal current source. Note that the V_{SG} of the PMOS is fixed, however its V_{SD} changes with a change in Vo. So a reduction in Vo would cause the V_{SD} of the PMOS to increase. What this means is that the V_{DS} of the NMOS needs to decrease by a slightly smaller extent than what would have been the case if the PMOS was a constant current source. A little bit of intuition will lead us to the more accurate result:

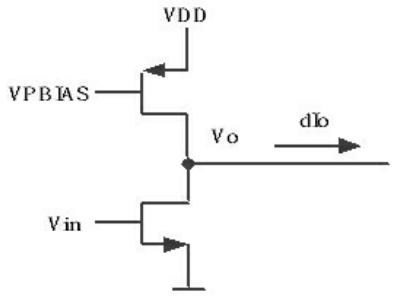
Delta Vout = -gmN/(gdsN+gdsP)

Where the N and P suffixes refer to the small signal parameters of the NMOS and PMOS respectively. We are still at a gain that is about half the intrinsic gain of a single MOSFET.

We have looked at the gain. But there is another more important parameter that characterizes a buffer. This is the small-signal ***output impedance***. What this impacts is the change in output voltage if there is an incremental current loading.

Output impedance = (Change in output voltage)/ (Change in load current)

For an ideal buffer, there should be no change in output voltage if there is a change in its load current. This would mean that the output impedance of an ideal buffer would be zero Ohms – that is consistent with our intuitive understanding that the ideal buffer should be akin to an ideal voltage source. Let us analyse the output impedance of the Common source amplifier. Note that we are referring to scenarios involving small changes around the operating point, so what we are actually talking about is the *small-signal* output impedance.



Let us assume that some external load demands an incremental current of (Delta Io) around the operating point. How would Vo have to change to be able to provide this extra load. Let us start with the assumption that the PMOS behaves like an ideal current source. So some extra diversion of current through the load would require the NMOS to now take lesser current. Since Vin is constant, the V_{GS} of the NMOS is unchanged. So the only way for the NMOS to reduce its current is by reducing its V_{DS}. The change in current would be therefore equal to:

Delta Io = -gds * (Delta V_{DS})

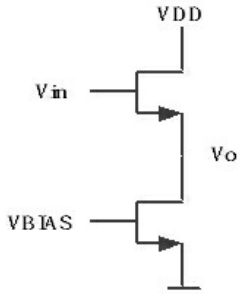
Note that Delta V_{DS} is the same as Delta Vo. So the output impedance as given by (Change in output voltage)/(Change in load current) is equal to (1/gds).

Again, a more detailed analysis involving the role of the PMOS will lead to the more exact result of Output impedance as being equal to 1/(gdsN+gdsP).

Since gds in general is a small number (small and large are relative, so we will keep it relative to gm), this points us to the fact that the Common source amplifier has a large output impedance. This should not be surprising – we have already intuitively seen that it is a poor buffer and not very robust to loading. This also explains why the operating point of the Common source amplifier changes by so much when it is loaded.

In summary, the Common source amplifier has a small-signal gain which is on the order of (gm/gds) and a small-signal output impedance which is on the order of 1/gds.

Now, let us see how the Source follower compares. First the small-signal gain.



The V_{GS} of the lower NMOS is fixed. So the only way it can change its current is by changing its V_{DS} (to which its current has weak sensitivity). So we can start with the assumption that the current of the lower NMOS (and thereby the upper NMOS does not change even with a change in Vin). So if Vin changes by Delta Vin and the current of the upper NMOS has to stay constant, then its V_{GS} has to stay roughly constant. The way this can happen is if the source 'follows' the gate. So to a first order, we can assume that the change in Vin reflects at Vo. So let us assume that Vin increase by Delta Vin and this causes Vo to increase roughly also by Delta Vin. But this will result in a reduction of the V_{DS} of the upper transistor by (Delta Vin). Starting with the assumption that the lower NMOS is an ideal current source, the change in current of the upper NMOS as triggered by a reduction in its V_{DS} (by Delta Vin) has to be offset by a slight increase to its V_{GS} (which will restore parity). What this means is that Vo cannot increase by exactly (Delta Vin) but by a slightly smaller amount so that there is a slight increase in V_{GS} to

compensate for the reduction in V_{DS}.

gm * (Delta V_{GS}) = -gds * (Delta V_{DS})

Noting that Delta V_{GS} is equal to (Delta Vin – Delta Vo) and Delta V_{DS} is equal to (-Delta Vo), we get:

gm * (Delta Vin – Delta Vo) = -gds * (-Delta Vo)

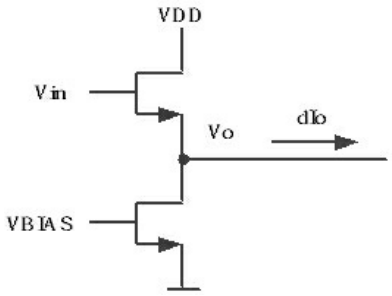
This gives us:

Delta Vo = gm/(gm+gds)*(Delta Vin)

So the 'gain' from input to output is equal to gm/(gm+gds), which is close to '1' but slightly lower. The extent to which it is lower than 1 is determined by the factor gm/gds – the larger it is, the closer to '1' is the gain of the source follower. For example, the gain would be 10% lower than 1 if gm is 10 times gds.

Again, the role of the lower NMOS (which has been neglected in the initial analysis) can be ascertained to change the gain to be equal to gmT/(gmT+gdsT+gdsB) where the suffixes T and B refer to the top and bottom NMOS transistors.

So we have seen that the Source follower has a gain close to '1', which is good for a buffer. What about its output impedance?



If an external load draws an incremental current of Delta Io, how will the source follower respond to this current? Since Vin is constant for this analysis, all it needs to do is to change its Vo by a small amount so that its V_{GS} adjusts to provide the incremental current Delta Io.

Delta Io = gm * (Delta V_{GS})

Note that Delta V_{GS} is equal to the negative of Delta Vo (since Vin is constant). So we get:

Delta Io = -gm * (Delta Vo)

The output impedance is therefore given by 1/gm. As we have seen before, gm is (relatively) large, so the factor 1/gm points to a ‘small’ output impedance. This fact again reinforces that the Source follower is a good buffer with a low output impedance that is capable of robust operation even under loading.

We can rewrite the small signal attributes of the source follower buffer as follows:

Gain = 1/[1+(gm/gds)]

Output impedance = 1/gm

By making the NMOS transistors larger and larger, the gm can be increased, thereby reducing the small-signal output impedance of the buffer. However, the gain accuracy (deviation from ‘1’) is limited by the inherent (gm/gds) of the transistor. This leads us to the direction for how we can improve on a source follower in terms of its approximation of an ideal buffer. Can we do better? For example, can we get to a topology that gives us something like the below?

Gain = 1/[1+(gm/gds)^2]

Output impedance = (1/gm)/(gm/gds)

What we have done in the above equations is to boost up the gain accuracy and lower the output impedance both by an additional factor of (gm/gds). Such a buffer would be an even better approximation to an ideal buffer. Or for that matter, what about something that has the following attributes:

Gain = 1/[1+(gm/gds)^3]

Output impedance = (1/gm)/[(gm/gds)^2]

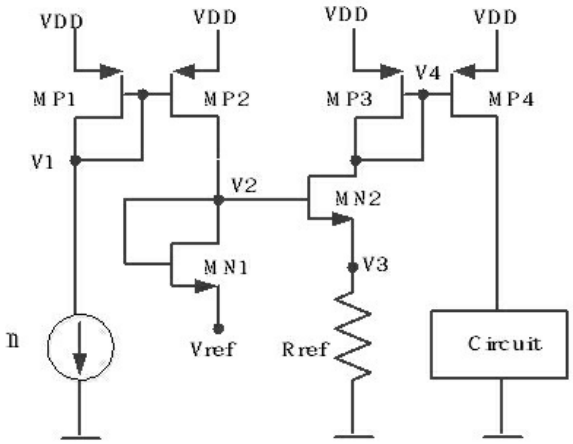
That would be super!

The quest to get to an even better buffer than the Source follower will lead us deeper into the world of Analog design. It will force us to go back to a high gain topology (like the common source amplifier), modify it achieve even higher gain, and to harness the high gain. It will also force us to deal with the complications inherent in negative feedback circuits. Aman-Ra and Uman-Ra will soon realize that not all is perfect with what seemed like the magic bullet

to solving all their relationship issues.

A powerful technique to analyse any circuit

The concept of when the MOS transistor behaves like voltage source and when it behaves like a current source can be used powerfully to analyse circuits. Let us use this technique to analyse the behaviour of the circuit shown below.



Starting from left to right, let us go through the role of each transistor.

For MP1, its current is forced to I1. However its controlling voltage is free to adjust itself in order to take the current I1. In fact, MP1 is a diode-connected transistor and as seen earlier, behaves like a current controlled voltage source. Its role is to ‘generate’ the proper controlling voltage needed to take the current I1.

For MP2, its controlling voltage gets imposed by the connections to its gate and source. It therefore tends to behave like a voltage controlled current source. In fact, since its controlling voltage is same as that of MP1, it is expected that MP2 will take the same current as MP1, which is I1. The manner of connection of MP1 and MP2 is what is called a **Current mirror** – a circuit that is used to ‘mirror’ currents from one circuit arm to another. All this is valid of course only provided MP2 is also in saturation. But for now, we will assume all transistors in our circuit are in saturation.

For MN1, the current gets imposed by MP2. However, while its source is forced to a voltage equal to Vref, its gate is free to adjust itself in order to take the current imposed by MP2 (namely I1). MN1 is also connected in diode connected manner, and like MP1, its function is to generate a V_{GS} voltage needed to take the current I1. As stated earlier, in a scenario like this,

the transistor behaves like a voltage source connected between its Gate and Source terminals. The potential at V2 will therefore be equal to (Vref+V_{GS1}) where V_{GS1} is the controlling voltage required for MN1 to carry the current I1.

Analysing MN2 is a bit more tricky. Its source is not forced. Hence it is reasonable to assume that it will have a voltage-source like behaviour. Let us assume that its V_{GS} is equal to the V_{GS} of MN1 (this is strictly true only if the currents through MN1 and MN2 are equal). In that case, the voltage that MN2 tries to 'impose' at V3 would be equal to Vref. This is because:

V2=Vref+V_{GS}(MN1)

V3=V2-V_{GS}(MN2)

For now, we assume that MN1 and MN2 have the same VGS.

So then V3 = Vref.

By imposing a voltage V3=Vref, the current through the resistor Rref would be given by **Vref/Rref**. Of course, if this current is different from I1, then our assumption that MN1 and MN2 have the same V_{GS} would not be valid. However, since the current has a high sensitivity to V_{GS}, to a first order we can go ahead with the assumption that the V_{GS} of MN1 and MN2 re nearly equal if the currents are in the same ballpark.

The configuration we have used therefore helps us generate a current that is proportional to a voltage Vref. The constant of proportionality is determined by Rref. What we have analysed in fact is a **Voltage-to-Current converter**, also referred to as a **V2I converter**.

Having generated such a current, we can now see how transistors MP3 and MP4 serve the role of current mirrors and are used to impose that current to flow into another circuit.

MP3 is diode connected and behaves like a voltage source, generating a controlling voltage needed to take the current Vref/Rref. MP4 behaves like a current source (more precisely a current mirror) and mirrors the same current into the circuit shown to the right.

The above analysis has implicitly assumed that all the PMOS transistors have the same dimensions and all the NMOS transistors have the same dimensions. This is what results in a current mirroring ratio of 1. However, if

the transistors (for example MP1 and MP2) are scaled with respect to each other, then it would result in a current mirroring ratio different from 1. However all the concepts would still hold true.

The analysis of when a transistor behaves like a voltage source and when it behaves like a current source can be a powerful analysis tool as shown in the analysis of the preceding circuit.

Taking stock of the ground we have covered thus far

In a few short chapters till this point, we have covered a lot of ground.

We had initially set out defining an ideal buffer as a circuit that provides an output exactly equal to its input – implying it has a gain of exactly 1. Along the way, we stumbled upon the Common source amplifier with a current source load. This was a circuit that gave a high gain between the input and output. But we realized that this is a circuit that is hard to practically realize because it had two current source-like elements in series. So we started synthesizing other configurations where one of the elements was made to have voltage source-like behaviour. We came up with three such topologies:

1. The Common Source amplifier with a diode-connected load
2. The Common Source amplifier with a gm-load
3. The Source Follower

Through the realization of the three topologies, we were able to get to our first approximations to an ideal buffer.

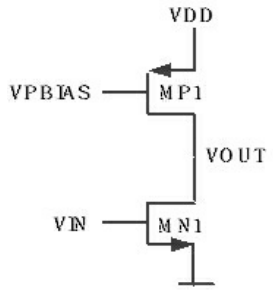
But many questions remain:

1. We have seen how ‘close’ the source follower topology is to an ideal buffer. But how do we move closer?
2. Is there at all a motivation to go back and try to realize a circuit that gives a high gain from its input to its output? Will that help us in our quest to realize a more ideal buffer than what we have done so far?
3. Does feedback always work as expected? When does it fail and what do we do when it fails?

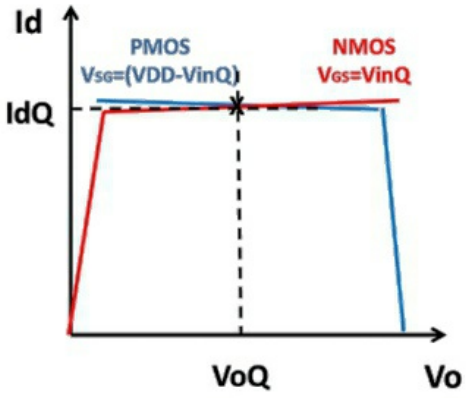
There are the many questions yet to be answered, and the next few chapters attempt to answer some of them.

The Differential amplifier

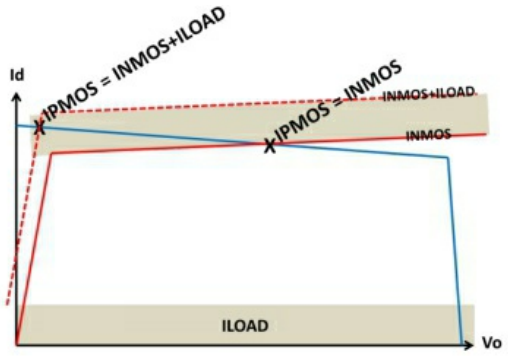
Let us now revisit the Common Source amplifier and whether we can make it take us further in our quest to find the Ideal buffer.



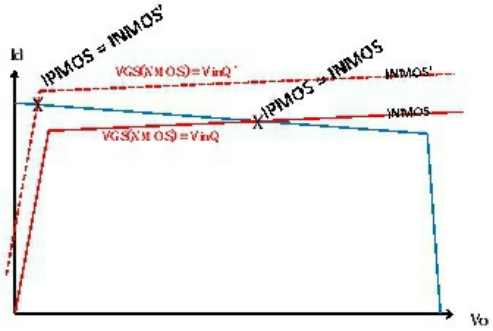
We saw that this circuit gave us a high gain from VIN to VOUT. To understand the complications in realizing this circuit practically, we just need to understand the simple fact that in this circuit, two current sources in series are trying to find a meeting point.



All it takes to disturb this meeting point is either of the current sources shifting a bit. **Case 1**: A loading of this circuit with a load of ILOAD requires IPMOS to now become (INMOS+ILOAD). As has been shown before, this causes a huge shift in the operating point as shown below.



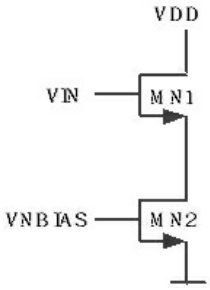
Case 2: A shift in the NMOS curve caused by the input slightly shifting from VinQ to VinQ'. This is illustrated below.



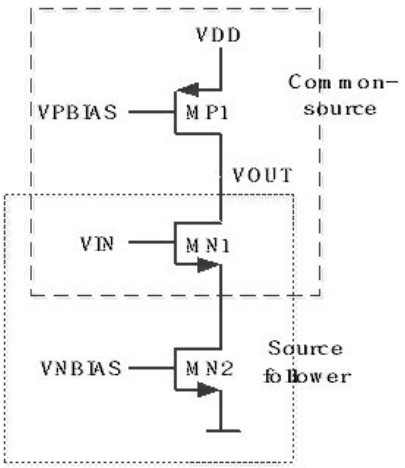
We shall first try and tackle Case 2 so as to remove the sensitivity of the circuit to its input bias point (VinQ).

Now among the circuits we have seen till now, can you think of one which is insensitive to VinQ?

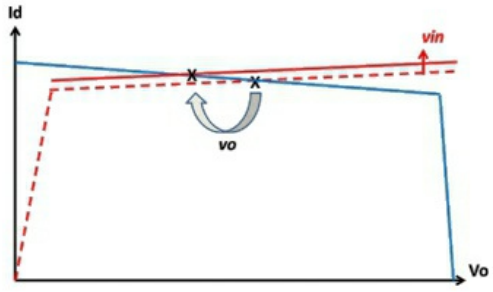
The Source follower!



It is not difficult to see the reason why. If the input bias point (the gate of the input NMOS) shifts, the source responds by shifting by almost the same extent, keeping the transistor's current the same. So will extending our Common source amplifier to something like the below circuit (by addition of MN1) solve the problem of sensitivity to the input bias point?

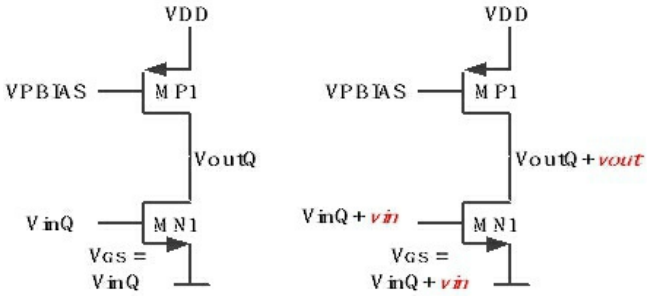


It is true that the circuit is now much less sensitive to VinQ. However, we have created a couple of new problems. One of them is easy to see – while MN1 no longer behaves like a current source, MN2 does behave like one. So we now have a potential conflict between two current-source like elements: MN1 and MP1! In fact, at first sight it seems like we have just transferred the conflict from MN2 and MN1. But keep in mind that MN2 is a constant current source unlike MN1, which is a current source dependent on the input signal. So if we can somehow match the MN2 current with the MP1 current (which is also a constant current source), we could possibly get this circuit to work. Nevertheless, we shall come back to this conflict soon. First we will deal with the other conflict – which is related to the ability (or rather the lack) of this circuit to respond to an input *signal*. What we define here as the signal is a change in Vin by *vin*. The common-source amplifier achieved a high gain between *vin* and *vo* by virtue of the current of MN1 changing. This change required the operating point of Vo to shift by a large amount, which meant a large *vo* for a small *vin*. This is shown below.

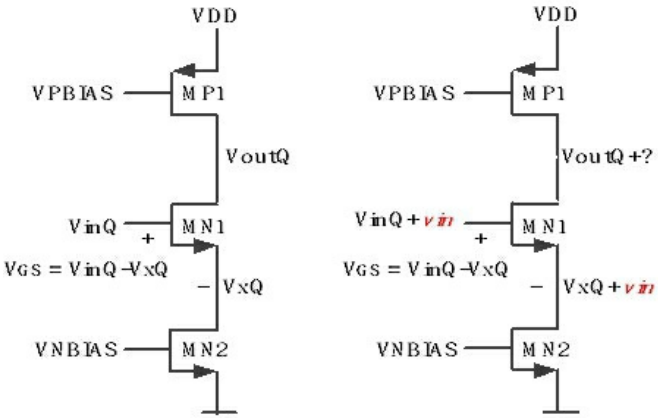


The property of the circuit responsible for its sensitivity to the input operating point is the same one that allowed it to achieve high gain! What made this possible in our original common-source amplifier was the fact that the entire

change in *vin* was impressed as a change in V_{GS} across MN1, causing its current to change.

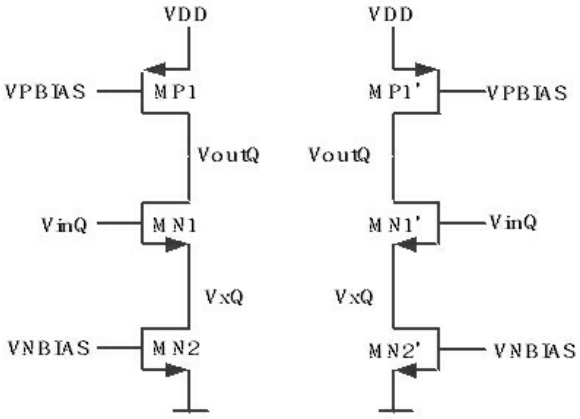


Now, with an addition of MN2 to the bottom side of MN1, the change in *vin* no longer appears across the V_{GS} of MN1. Since MN1 and MN2 now look like a source-follower, there is no change in the V_{GS} of MN1! This is illustrated below.



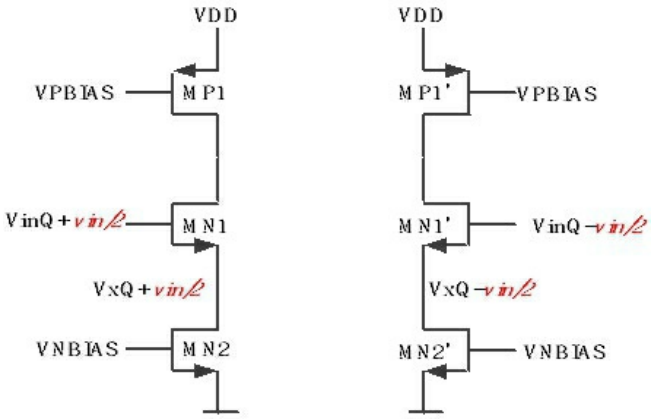
To state the problem we now have to solve - we want to have MN2 (because it desensitizes the circuit to VinQ); at the same time, we want *vin* to get impressed completely on to the V_{GS} of MN1. To break through and forge ahead requires a flash of genius.

Let's assume we have two such circuits – mirror images of each other as shown below.

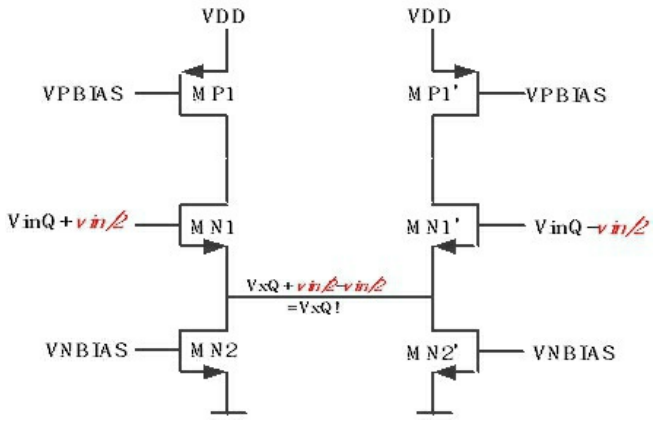


The two circuits are exactly identical, which is the reason why we have indicated both their operating points as VinQ and VoutQ.

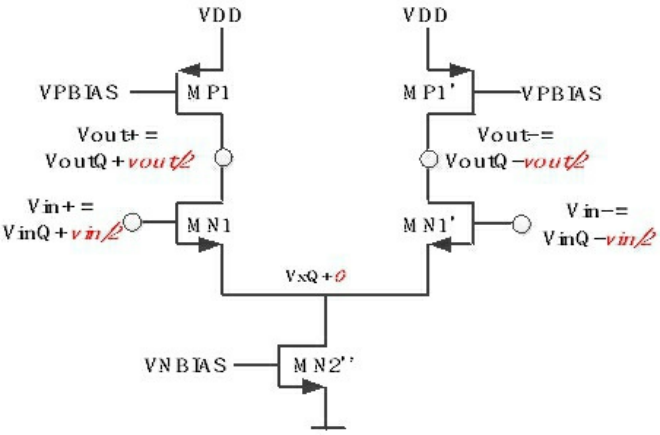
Now, let us impress a signal *vin/2* on the half circuit on the left side. At the same time, let us apply a signal *–vin/2* on the half circuit on the right side. The responses of the two sides are shown below.



As can be seen, the source of MN1 increases by *vin/2* whereas the source of MN1’ reduces by *vin/2*. Now for the moment of brilliance… What would happen if we connect the source nodes of MN1 and MN1’? With superposition of the opposing movement on the source nodes of the two sides, the now-shorted source node would not change with *vin*!



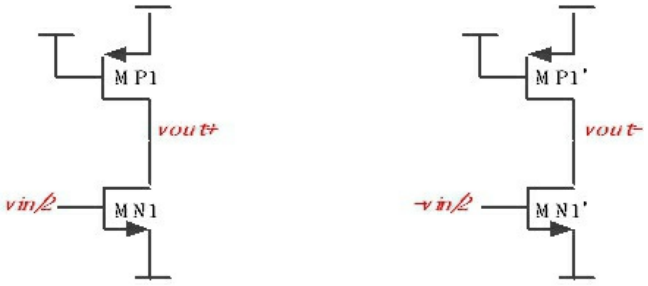
With the shorted source node not moving now with *vin/2*, the entire *vin/2* now gets impressed on the source of MN1, generating the signal current required to shift Vout. This is exactly what we wanted! As far as the response to the signal is concerned, the source of MN1 does not change, so the combination of MN1 and MP1 behaves and responds exactly like the Common source amplifier that we started with! The output of the left half of the circuit changes by *vout* (exactly the same extent of change the Common source amplifier would have given). Since the right half of the circuit has a signal of *–vin/2*, we can expect that the change to its output would be *–vout.* Also since MN2 and MN2' are like parallel current sources, we can combine them into a single current source MN2''. The resultant circuit looks like the below.



What we have just synthesized is a DIFFERENTIAL AMPLIFIER! The input signal of *+vin/-vin* applied to the input nodes on either side is what we refer to as a 'differential input' and the corresponding output signal *+vout/-vout* that appears on either side is referred to as the 'differential output'.

The reason why the differential amplifier circuit is a huge advancement over the simple Common source amplifier can be understood by seeing that this circuit decouples the establishment of bias point from the mechanism of signal gain.

The equivalent circuit for the response to a differential signal is as shown below.



Small signal equivalent circuit for a differential signal

What we have shown above is the 'small-signal' equivalent circuit, meaning that we have only indicated the relative changes to each node in the presence of a signal. The PMOS source and gates are constant voltage nodes, so are depicted as 'ground' (unchanging potential) when the signal is applied. As explained earlier, the source of MN1, MN1' do not change with application of a differential signal. Therefore, they are also shown as ground in the small-signal equivalent circuit. It is easy to see that the small-signal equivalent circuit of the Differential Amplifier exactly matches with that of the Common source amplifier. Hence we get the properties of high gain that we get from the Common source amplifier.

The input-output transfer function of the differential amplifier follows from that of the two constituent Common source amplifiers:

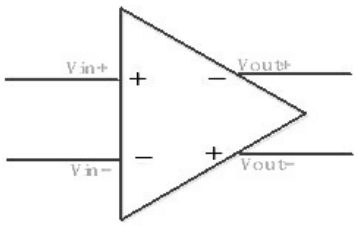
$vout+ = -gmN/(gdsN+gdsP)*(vin/2) = -G*(vin/2)$

$vout- = -gmN/(gdsN+gdsP)*(-vin/2) = G*(vin/2)$

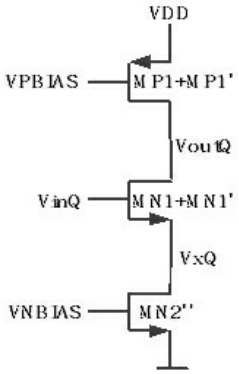
The negative sign in the transfer function means that as Vin+ increases, Vout+ reduces. The differential transfer function can be written as:

$vout = (vout+ - vout-) = -G.vin$

The symbolic representation of the Differential amplifier is the ubiquitous triangle symbol shown below:



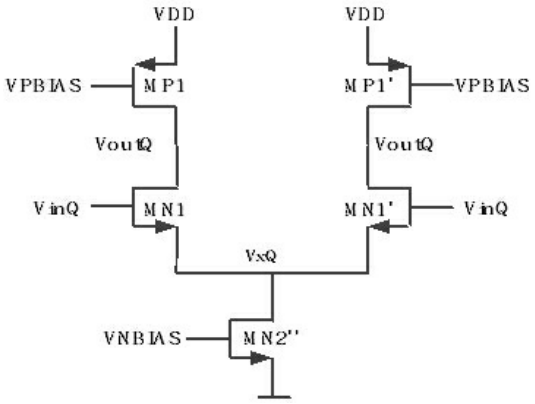
Next we look at the equivalent circuit for the establishment of the bias point.



As stated right at the beginning, we still have a potential conflict. The current of MN2'' should be exactly equal to the currents of MP1 and MP1'. If that is not satisfied then once again, we will end up with a circuit that can potentially work on paper but fail when we rig it up. Solving that problem will be lead us to our next shot at implementing an ideal buffer.

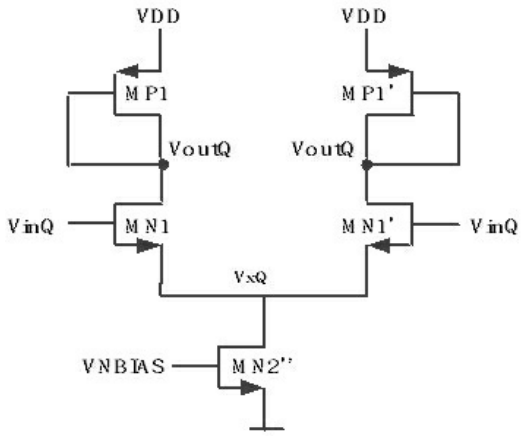
Our first Operational amplifier

Let us state the problem once more. How do we ensure that the current on MN2'' matches the sum of the currents of MP1 and MP1'?



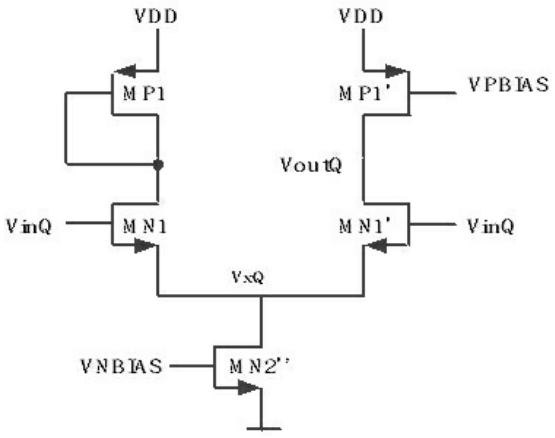
Let us first understand what gets impacted if the current of MN2'' is not matched with the PMOS side currents. Let's assume that the currents are perfectly matched to begin with, and VoutQ is at the desired value. VxQ will automatically adjust so that the V_{GS} of MN1 and MN1'' is at the right value required to take the current. If now, we slightly shift the current of MN2'', then MP1 and MP1' would need to adjust their currents so that their sum matches with the current of MN2''. The only parameter that is 'free' to change is VoutQ, which impacts the currents of MP1 and MP1' weakly through the change in V_{DS}. To adjust for a small change in the current of MN2'', VoQ will have to change by a large amount, potentially driving MP1 and MP1''. So in order to establish a stable VoQ, we need to figure out a way to match the N-side and P-side currents.

Clearly, we have three near-ideal current sources (in MN2'', MP1 and MP2). In order to allow the circuit to operate robustly, one of them has to stop being a current source. We had earlier encountered a similar situation and had solved it by the circuit modification shown below.



The circuit will work fine, and with a robust operating point since MP1 and MP1' no longer behave like current sources. Rather they adjust their V_{GS} so as to each take half the current of MN2''. For a differential input, each half circuit now behaves like a common source amplifier with a diode connected load. The only problem with this circuit though is that we have lost the high gain from input to output (which was our motivation to revisit the common source amplifier and modify it to get to the differential amplifier topology). The reason why we lost the gain was because we now have low output impedance on both the outputs because of the diode-connected PMOS transistors which behave like current-controlled voltage sources.

Let us look at another option. We only need one of MP1 or MP1' to stop being a current source. So how about the below circuit where we only do the diode connection (shorting the gate and source) on only the left side of the circuit (namely to MP1)?

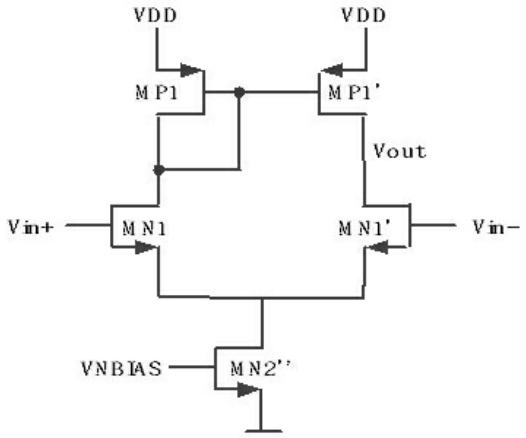


The way this circuit works to set the operating point is by allowing the current of MP1 to adjust itself to account for mismatches between the current

of MP1’ and half of MN2’’. In this circuit, we have retained the high output impedance on the right side (since MP1’ continues to behave like a current source) but we have converted the left side PMOS (MP1) into a diode-connected transistor. These modifications allow for a high gain from the input to the right-side output but kills the gain from the input to the left-side output. We have therefore marked the output as being only on the right side. Since we have lost the gain on the left side, the effective gain from *vin* to *vout* drops by a factor of 2.

While this circuit seems to have the extra knob needed to equalize the P-side and N-side currents, there is still no clear-cut mechanism to set VoutQ which is still set by the meeting point of two current sources. What is preferably to have is a mechanism where both MP1 and MP1’ are flexible to adjust themselves to establish the operating point, but MP1’ does not have a diode connection that kills the right side gain.

Both the above problems mentioned (the reduction in gain by a factor of 2, and the continuing conflict with the current of MP1’ being fixed) are solved through a single stroke of intuition – by not forcing the gate of MP1’ to PBIAS but simply by shorting it to the gate bias of MP1, which is already established through the diode connection of MP1.

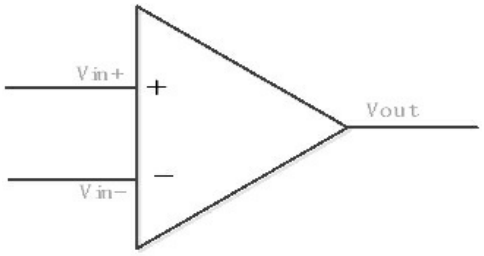


It is clear that the left side output still does not have high gain because of the diode connected MP1 causing the output impedance to be low. So there is high gain from the input only to the right side output. So the right side output (Vout) is referred to as the single-ended output and such an amplifier is referred to as an amplifier with a single-ended output.

A careful analysis of the above circuit shows that it checks off all the problems we were trying to solve:

1. MP1 and MP1’ are now both flexible to adjust their currents so that they match with the current of MN2’’.
2. The drain of MP1 (shorted to its gate) is set to a proper bias point because of the voltage-source nature of the diode connection of MP1. Essentially, the drain voltage of MP1 will come to a voltage equal to (VDD-V_{SG}) where V_{SG} is the source-gate voltage for MP1 to take half the current of MN2’’.
3. The symmetric nature of the circuit ensures that bias point of the right-side output (VoutQ), which is the drain of MP1’ comes to the same voltage as the drain of MP1. The robust mechanism setting the output bias on the left-side therefore gets transferred to setting VoutQ as well.
4. The final not-so-obvious fact is the restoration of the factor of 2 gain that was lost in the previous topology. Without much elaboration, we state that MP1 and MP1’ behave like a current mirror pair. So the signal current that gets generated on the left side of the circuit flows through MP1 and gets mirrored and added to the signal current generated in MP1’ due to the right side of the circuit. So though all the gain appears only on the right side, the signal gain from the differential input *vin* to the single-ended output *vout* is the same as what we had got in the case of our differential amplifier.

The above mentioned amplifier is referred to sometimes as a **Five-pack amplifier**: a stack of five transistors in a simple, elegant topology that bring out the essence of every concept we have learnt so far. Such an amplifier can be depicted as shown below:



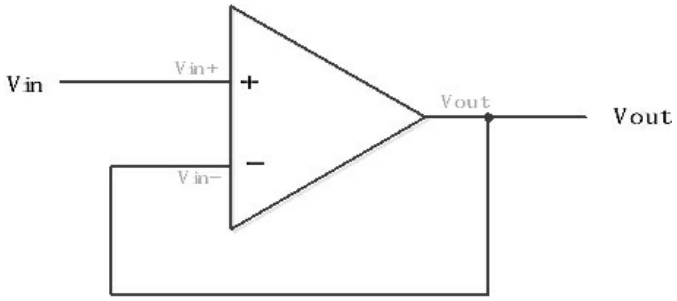
The gain of this amplifier from input to output is of the order of gm/gds (we leave out the details).

With the Five-pack amplifier, we have gotten to a robust topology that can potentially harness the high-gain properties of our original Common source

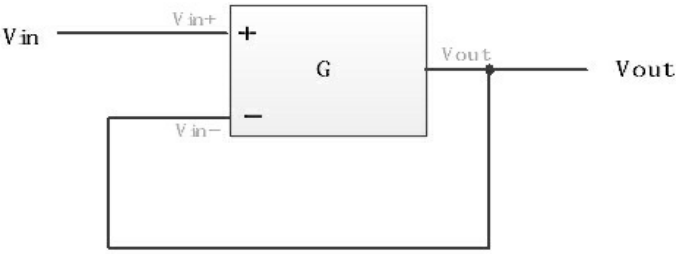
amplifier. But how we are going to use it to build a robust unity-gain buffer is a topic we will investigate next. Let us conclude this chapter by saying that our five pack amplifier is our first approximation to an **Operational Amplifier**. We will soon see the reason.

The buffer realized using the Operational amplifier

So we now have a high-gain single-ended amplifier built using five transistors. How do we build a unity gain buffer using it? Assuming the reader has a background in control systems, we just get to the configuration directly.



Here, we have used a negative feedback configuration to rig up a unity-gain buffer between terminals Vin and Vout. A control theory style depiction of the above would look like the below:



The gain block G (depicting the gain of the five-pack amplifier) would cause the following relation:

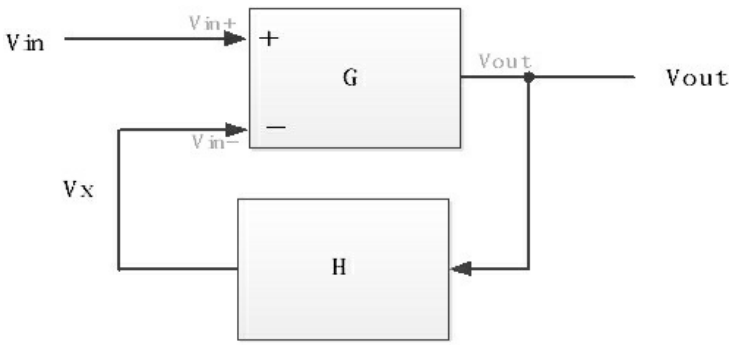
Vout = (Vin+ - Vin-)*G = (Vin-Vout)*G

Therefore:

Vout = Vin * G/(1+G)

If G were to be a number >> 1, this would result in Vout ~ Vin, which would result in our unity gain buffer.

A more generalized feedback topology would be the one shown below:



In this case, a fraction of Vout (denoted as H) is fed back to the negative terminal of the gain block. The resulting transfer function would be equal to:

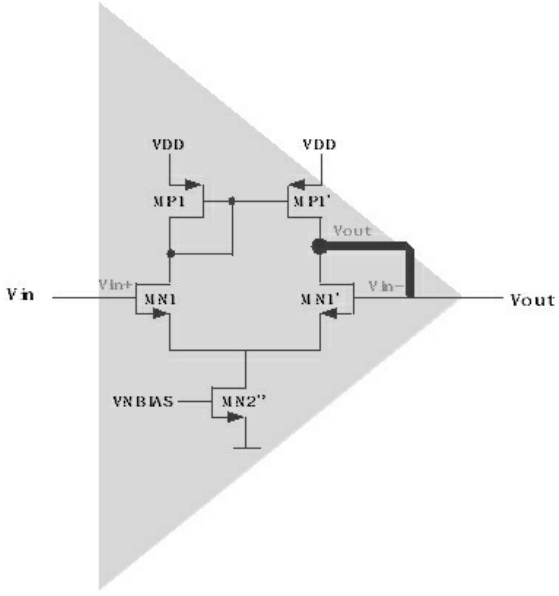
Vout = G/(1+GH)

For a case where G>>1, Vout ~ 1/H.

Assuming H is a fraction less than 1, this would essentially result in the realization of a buffer with a gain higher than 1, and equal to 1/H.

But let us get back to our unity gain buffer, where Vout is directly fed back (in other words H=1) to the negative terminal of the gain block.

The transistor level circuit would look like the one shown below.



The thickened line shows the feedback connection (from Vout to Vin-) that transforms the five-pack single-ended output amplifier into a unity gain buffer.

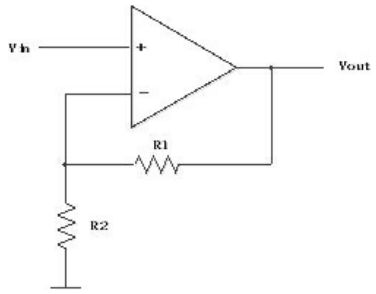
Assuming the gain of the five-pack amplifier to be gm/gds (here we ignore all

the pesky factors of 2 and the minor details about gmP, gdsN and the like!), the transfer function of the unity gain buffer would be:

Vout = Vin*gm/(gm+gds)

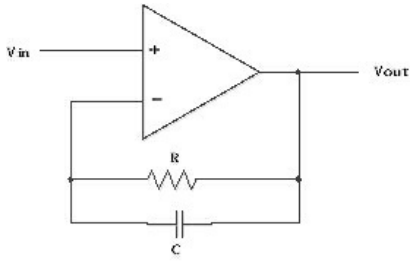
This bears a striking resemblance with the gain we got from the source follower. For example, if gm/gds of the transistors are on the order of 10, then the gain of our 'unity gain' buffer would be accurate to 90%. It seems like we have put in a lot of effort but have landed up with a circuit that is only as good as our simple source follower circuit. Pressing ahead along the same road will show us that this is not the case, and that the five-pack amplifier we have constructed can do a variety of useful functions that our humble source follower cannot. In fact, the five-pack amplifier is our first approximation of an **Operational amplifier**, which can help realize a slew of useful transfer functions. We look at a couple of examples below:

A buffer with a gain different from 1:



The circuit shown above is a buffer with a gain of (1+R1/R2).

A low-pass filter:

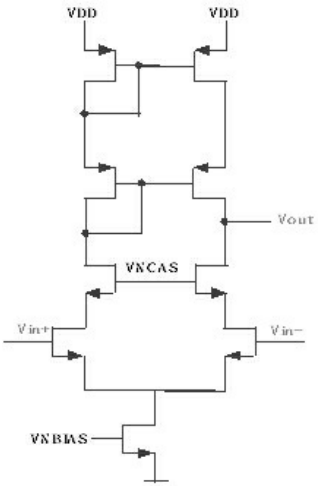


The circuit shown above is a low-pass filter with a 3-dB bandwidth given by 1/(2*pi*R*C).

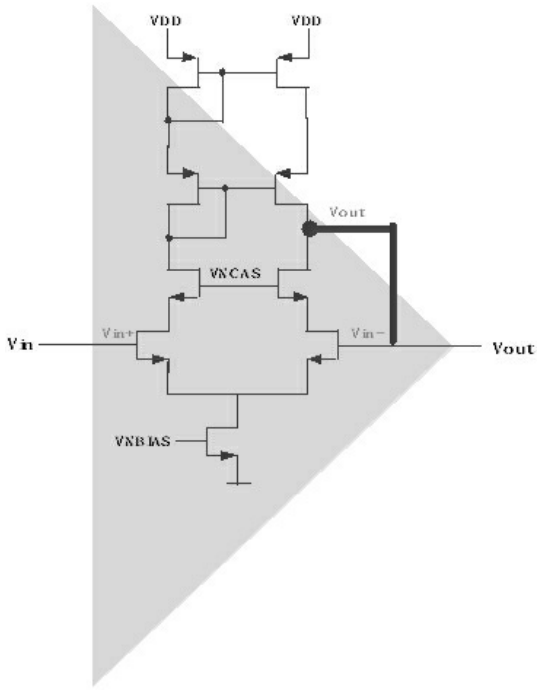
As we can see, all the efforts we put to get to the Operational amplifier are efforts well spent. But we had promised that we would get closer to the ideal buffer. What would it take to get to such a buffer?

One way to answer that question is to see whether we can modify our basic operational amplifier to increase its gain G. After all, how closely Vout tracks Vin is determined by how high the value of G is. For example, can we build on the five-pack amplifier to make its gain (gm/gds)^2? In that case, starting with transistors which have a (gm/gds) of ~10 would result in an inaccuracy of not 10% but just 1%.

Without much background, and just with a view to keep the reader's interest sustained for the continued pursuit of knowledge, we present below an extension of the five-pack amplifier that would achieve this. Such an amplifier is a telescopic cascode amplifier with single-ended output and has a gain on the order of (gm/gds)^2.



As can be expected, such an amplifier, when connected in feedback would lead to a buffer with an accuracy much better than the source follower.



We could continue to build on the above circuit, deriving for example a gain of the order of (gm/gds)^3. That would reduce the inaccuracy of our buffer even further. It seems like we are well on our way to finding the Holy Grail we had set out to at the beginning of this book. However, the laws of circuits have other plans. As we go further trying to rig up such circuits, we very soon find out that they do not work as expected. We start running into problems that we did not expect to. More specifically, feedback that has seemed to be our silver bullet till now, sometimes does not work as expected. The complexity that we are adding to our amplifier design (for example, in trying to realize a higher and higher gain) is seemingly the cause for these new and unexpected problems. We shall see more of this in the chapters to come.

On the relationships front - since the visit of Ang-Lao, Aman-Ra and Uman-Ra have managed to tighten up their relationship. Amidst the chaos of their life (Mil-Ra, Gamen-Ra and Lil-Ra), they have somehow managed to preserve their equanimity. They still have quarrels but then recall the principles of feedback to correct course and get back to their loving ways.

But is this new honeymoon phase going to last forever? The evil forces are plotting and are getting ready to throw some surprises. We shall see this in the next chapter.

When feedback goes bad

As the days roll by, our protagonist couple starts to find that feedback does not always work. In fact, they find there are times when it has an opposite effect to the intended one, and it seems to end up making the situation messed up even more! Here's an example - Aman-Ra has this tendency to yield to every demand of his mother, however crazy it might be. Uman-Ra is usually incensed when her husband comes home and tells her about it but in the interest of peace and harmony, usually chooses to keep quiet. One day, his mother asks him to find her a pet lizard. That evening, Aman-Ra goes into the forest in his quest for the pet lizard but as luck would have it, he gets bitten on his ear by a wild monkey. He reaches home, bleeding and in pain, and recounts his horror to his wife, hoping for comfort. The sight of her husband 's pain reminds her of the foolishness of her mother-in-law and his own lack of judgement while giving in to her unreasonable demands. In anger and frustration, she shouts at him '*What a fool you are. You have gotten what you deserved!*'

Needless to say, these incidents play a big role in widening once more the cracks in their relationship. To each of them, their own behaviour appears to be motivated by the best interest of the other. However, every such incident leaves a bad taste in the mouth and makes them drift apart. Now desperate to save his marriage, Aman-Ra is at his wit's end. All of a sudden, he remembers the shining light who had once given him a glimmer of hope – Ang-Lao! He sets out in search of the mystic master, traveling across the oceans to try and find him. Several months later, he tracks his whereabouts to a remote village in China.

When he reaches the village and gets to the cave where Ang-Lao now lives, he sees a large group of village folk in congregation. He learns that Ang-Lao has become a famous Guru and hundreds throng to see him every day. He has also taken a vow that he will not speak more than five words in a week. With great difficulty, Aman-Ra is able to meet Ang-Lao but the meeting is very brief. He describes his tale of woes, recounting how Ang-Lao's advice on feedback had initially helped his relationship with his wife, and how with time, feedback started to have the opposite effect. Ang-Lao listens carefully, occasionally nodding his head, sometimes closing his eyes. But he speaks not. After completing his narrative, Aman-Ra pleads to him '*Tell me the*

solution to my problem, O great master and all-knowing One.' Ang-Lao speaks slowly, measuring every word, making sure that he does not exceed his weekly quota of verbiage. Aman-Ra listens intently and absorbs like a sponge every one of those five words – '*The problem is the face.*'

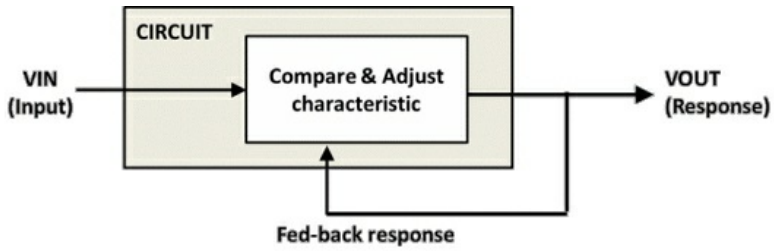
As Aman-Ra makes his long journey back to Egypt, he reflects continuously on Ang-Lao's words. Every once in a while, he stares into his pocket mirror, looking at his own face, hoping that it could give him the clue to unlocking the mystery. He sees how he has developed wrinkles and lines beneath his eyes. '*No wonder Uman-Ra has stopped loving me'*, he reflects, '*my face is not as dashing as what it used to be a few years ago.*' He makes a mental note to apply a mud pack from the banks of Nile every day to smoothen out the wrinkles on his face and to rejuvenate his skin once more. Confident that he has found the solution to his problem, he reaches the banks of Nile with renewed hope.

A whole month passes. Thirty days of mud pack seems to have done precious little to Uman-Ra's attitude. In fact, seeing him in the evening with his face covered in black mud seems to incense her even more. By now, the mud has worked wonders in smoothening the wrinkles on his face but have done nothing to remedy the cracks in their relationship.

Dejected but equally determined to decode the mystery of Ang-Lao's words, Aman-Ra decides to lock himself up in one of the pyramids he is building, and come out of it only once he has gotten the answer. In the dark interiors of the pyramid, he meditates for many days without food and water. He reflects on the various altercations he has recently had with his wife. The incident involving the monkey bite keeps coming back to him. He recalls the expression of anger in Uman-Ra's face when he first told her that he was going lizard-finding. How he wished she had stopped him at that point of time rather than deciding to vent her anger at the significantly more inopportune moment when he came back from the forest, with bleeding face and bruised ego. He realizes at that moment how the anger she vented at him was born out of her contempt for the unreasonable demands his mother made from him. In that moment, he sees once more the concern she feels for his well-being and he is filled once more with the warm glow of their love. And in that warm glow, he sees with clarity the full import of Ang-Lao's words – '*The problem is the face.*' And he realizes what Ang-Lao really meant to tell him… '*The problem is the* ***Phase****!*'

Feedback and Phase

Let us revisit the block diagram we have used to illustrate feedback to understand Aman-Ra's new-found understanding of Ang-Lao's words.

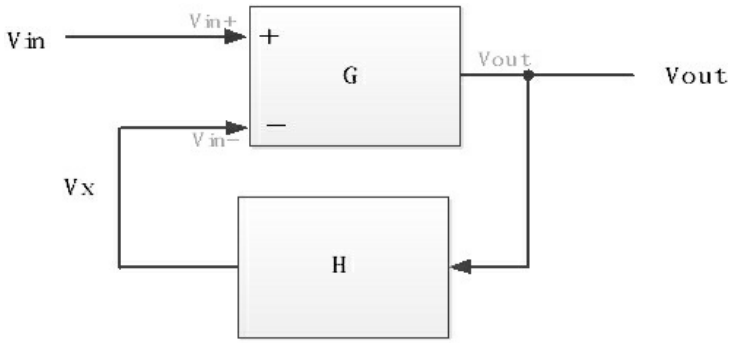


Feedback works like a charm when the fed-back response is consistent with the applied stimulus. However, if there is a delay in the fed-back response with respect to the stimulus, then it is possible that the feedback actually has an unintended effect. Let us look at the monkey-biting incident to illustrate this. The original stimulus presented to Uman-Ra was her husband coming home and telling her about his mother's demand for a pet lizard. Her first reaction was how her poor, hard-working husband was to go looking for a lizard. However, she kept quiet, not wanting to pick a quarrel. Had she expressed her concern at that point, it would have been likely that the matter would have gotten resolved. Aman-Ra himself was undecided whether to give in to his mother's demand or put his foot down and refuse. The voice of sanity from his wife would have tipped the scales decisively. However, the opportunity was lost because of the delay in her response to the stimulus. All the anger and frustration instead found ***delayed*** expression when she saw her husband come back home, his face covered in blood. The response to that sight turned out to be giving vent to her pent up anger from earlier in the day instead of the expression of sympathy he was so badly seeking. Yes, the problem was with the ***delay*** –in circuit parlance, we call this as ***phase***.

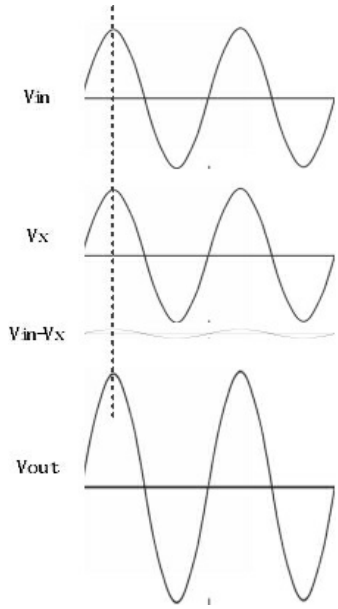
It is easy to see why delay and phase go hand in hand. If you had a sine wave input pass through a delay, you would get an output sine wave whose phase was shifted. There are many mechanisms that contribute to phase (or delay) in a circuit but the most fundamental one is the presence of capacitors. Think of a simple resistor (R) and capacitor (C) series combination – passing a sine wave of frequency 'f' through it will give you a phase shift of arctan(1/2 □ fCR). In fact every node of a transistor circuit has what we call 'parasitic capacitances' arising from just the terminals of the transistor. So

even if you do not have physical intentional capacitances in your circuit, you still end up with elements in your circuit that are equivalent of the simple R-C filter, each adding just a little bit of phase shift to the signal. Go around the entire feedback loop, adding the phase shifts introduced by each of these elements and you can end up with a scenario where you can have a large enough phase shift that causes the feedback mechanism to go kaput. We will explain all of this one step at a time.

Let us look at our generalized block diagram of feedback to see what could happen:

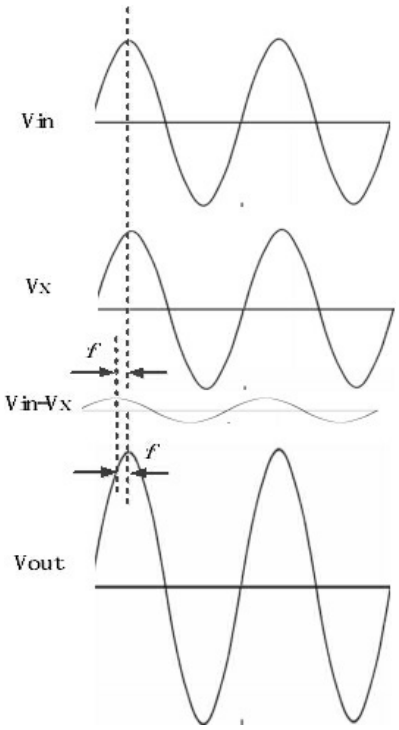


Let us first consider the case where there are no phase delays either in G or in H. So if we input a sine wave at Vin, what would happen at the various nodes? We know how steady state would look: Vout would be a sine wave very close to Vin/H and Vx would be a sine wave very similar to Vin. (Vin-Vx) is the input to the block G, and gets gained up by G to generate Vout. So there would have to be just that little amplitude difference between Vin and Vx so that their difference gained up by the factor G would result in Vout. Also since we have assumed there are no sources of delay (or phase) in either G or H, all three waveforms would have the same phase and would look below.



Now, what would happen if G introduced a phase delay of *phi* to the sine wave that was input to it? How would steady state operation look? We can start off by assuming that Vx would still be very close to Vin in both amplitude and phase. We will start off with this assumption and later check whether it is indeed a valid one or not. The reason for this assumption is because anything to the contrary would cause a very large difference voltage to go into the gain block G, which would result in a very high voltage at Vout, likely causing the gain block output to saturate and be unusable.

The assumption that Vin and Vx are close in amplitude and phase would require that Vout also be close to Vin in phase. After all, Vx is nothing but a fraction of Vout – and for now we have not assumed any phase resulting from the feedback block H. For Vout to be close in phase to Vin would therefore require that (Vin-Vx) would need to have an **advanced** phase of *phi* with respect to Vin (and Vx). With (Vin-Vx) have such an advanced phase, the additional phase delay of *phi* through the block G would cause Vout to have a phase nearly the same as Vin. We illustrate this below.



In the above diagram, we have depicted Vin, Vx and Vout as having roughly the same phase but (Vin-Vx) has a phase that is advanced (earlier) by *phi*. At first glance, this looks implausible. Let's take a value of *phi* to be 60 degrees. What in effect we are asking for is Vin and Vx to be very close to each other in amplitude and phase but for (Vin-Vx) to have a phase difference that is significantly (in this case 60 degrees) advanced with respect to Vin. When I started looking at it this way, it sounded pretty incredulous to me too. But then I did a simple analysis where I made Vx to be 99% of Vin and to lag the phase of Vin by a mere 1 degree. And when I subtracted Vx from Vin, what I got was something incredible – the difference of (Vin-Vx) turned out to have a phase that was advanced in phase by 60 degrees with respect to Vin! When I made Vx to be 99.9% of Vin (which could be a likely case when G was even larger than in the previous case), all I needed was about 0.1 degree of phase advancement of Vx with respect to Vin in order to create a phase advancement of 60 degree in (Vin-Vx) with respect to Vin. This bit of trigonometric marvel should convince you that indeed our negative feedback loop can do its job and achieve the (almost) same Vout independent of the phase introduced by G. Turns out that when you think of amplitude and phase in terms of the vector notation, the subtraction of two vectors that are close to each other in amplitude and phase can actually result in a third vector that has a large phase difference with respect to the two vectors that were subtracted.

So even with a simple revisit of Electronics 101, the plausibility of this should be apparent.

Let us take a step back and look at the import of what we just discussed. The delay in G, at first glance, makes it appear as though our feedback circuit is having a delayed response to the stimulus presented to it. However, from the above analysis, we have seen that Vx, the fed back voltage results in having the (almost) same phase as Vin. In other words, though there is a delay contained within the feedback loop (in the previous case, a phase delay from G), the feedback loop itself is able to provide an almost perfect response to the stimulus, *in phase* with the stimulus. It appears that the feedback loop is indeed tolerant to delays within the loop. This is true –almost. If you keep increasing the delay (the value of *phi* which we assumed to be 60 degrees in our illustration). Something totally unexpected happens as you get to a *phi* of 180 degrees. And in it is contained the flux of what can go wrong with feedback.

Before we are able to understand what can happen when *phi* goes to 180 degrees, let us revisit the case of *phi* equal to 60 degrees. Let's say we did not give any input. In other words, we set (Vin=0). Let us assume that despite the lack of an input, some mysterious hand was able to impress a sine wave of non-zero amplitude at Vout – do not ask me yet how this is possible. The question we will first try to answer is whether a steady state operation can be re-established with Vin continuing to be zero and Vout continuing to be non-zero. To analyse this, let us start at Vout and go 'around the loop' to see what would happen. A fraction of the sine wave would appear at Vx with the same phase as Vout. Since Vin is equal to zero, (Vin-Vx) would have a phase opposite to the phase of Vout. So, if we started with Vout having a phase of 0 degree, then (Vin-Vx) would have a phase of 180 degrees. This would require Vout to change its phase to 240 degrees, since the gain block G introduces a phase shift of 60 degrees. If you keep going around the loop again and again, you will see that there is no steady state possible. In other words, the only state state possible when there is no input, Vin is a Vout equal to zero. Such a system is a 'stable' system in the sense that Vout responds nicely and as expected when there is a Vin, but stays at zero when Vin=0.

Now, let us analyse the case when *phi* is set exactly to 180 degrees. Let us again consider the case where Vin=0 and by the same mysterious hand, a

sinusoidal waveform with a non-zero amplitude has gotten impressed at Vout. Let us go around the loop once again. If we assume that the starting phase of Vout is 0, then (Vin-Vx) would have a phase of 180 degrees. Since the block G also adds a phase of 180 degrees, completing one full circle of the feedback loop would result in the phase of Vout to be 360 degrees, which is … 0 degrees! In other words, the phase that Vout needs to be in order to maintain steady state is equal to the phase we started with! In other words, when *phi* is exactly equal to 180 degrees, a sinusoid of non-zero amplitude can miraculously sustain at the output even in the absence of an input! Also, if we started off with a miniscule amplitude on Vout, you can see that when every time we go around the loop, the amplitude has gotten gained up by a factor G*H! It is like this little monster that appears out of nowhere and progressively becomes bigger and bigger. In fact, the amplitude is boundless – or rather it is bounded only by the power supply rails of the circuit. Our feedback loop has developed a mind of its own. It has gone bonkers! We call such a feedback loop as a **Positive feedback** loop. *What it does at its output has no connection to what the stimulus is!* I hope you can start seeing the relationship analogy!

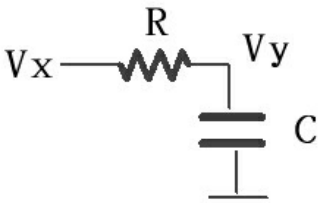
OK, it appears that I am pulling a fast one here. You may ask me – '*but how can something miraculously appear on Vout in the first place'*. The simple answer is that every circuit has components that generate noise and noise is something that superimposes on top of 'signal'. So well, the seed of our little monster waveform at Vout could have just been 'noise'! The next question may be – '*but is it not too much of a coincidence that the noise is introduced at the precise frequency where phi is equal to 180 degrees?*' The answer to that question is that the sources of circuit noise are usually 'white' – what that means is that they have some small but non-zero energy at all frequencies. So as long as there is one frequency for which the amplifier G has a phase *phi* equal to 180 degrees, the circuit will find a way to trigger positive feedback and will generate an output at that frequency even in the absence of an input. In fact, what we have described is a case of an 'amplifier' turn into an 'oscillator'! The precise frequency at which the phase *phi* goes to 180 degrees will be the frequency at which the circuit oscillates (producing a sine wave output even in the absence of any input).

We saw that the origin of the whole problem of positive feedback comes from the phase introduced within the loop – most commonly inside the op

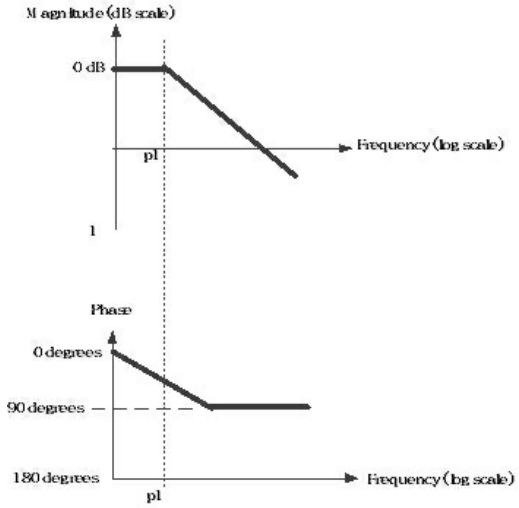
amp. We will next look at what are the sources of this phase and what we can do to counter its effects. We will look to take these insights back into the relationship domain and come up with a possible way to fix the complications of the kind that Aman-Ra and his wife have been grappling with.

The sources of phase

Let us first look at the concept of phase in the simplest circuit we can think of – the R-C circuit as shown below:



If you had a sine wave Vx=Vo.cos(wt), the output of the filter would be Vy=*m*.Vo.cos(wt-*phi*) where *m* and *phi* are the magnitude and phase of the R-C filter at the frequency of w. *m* starts off at 1 (or 0 dB in the dB scale) at low frequency and keeps dropping lower and lower as the frequency goes towards infinity. *phi* starts off at zero at low frequency and reaches a maximum of 90 degrees at infinite frequency. These are depicted by the magnitude and phase plots shown below.



In the plots, the frequency is shown in the log scale in these plots and the plots have been approximated to highlight out some key points of interest:

1. The magnitude is roughly constant till a frequency marked as p1 and then starts to fall with a constant slope of 20 dB/ decade. What this means is that for every 10x (decade) of frequency change, the magnitude drops by a factor 10 (which is 20 dB)
2. We have shown an abrupt change in the slope of the magnitude plot at p1. However, the change is gradual. In fact, the magnitude

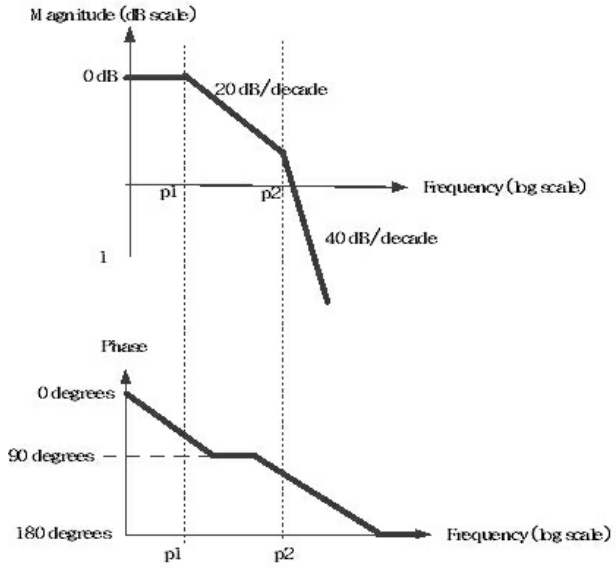
is -3 dB at p1. The frequency p1 is equal to 1/(2.pi.R.C).

3. At p1, the phase is equal to 45 degrees. As the frequency goes much higher than p1, the phase eventually reaches 90 degrees and does not increase any further.

The frequency p1 described above is referred to as a ‘pole’ of the R-C circuit. In fact any node in the circuit that is connected to a resistor (or an element that mimics a resistor) and a capacitor has an associated pole. With such a pole comes the magnitude and phase effects as described for the R-C circuit.

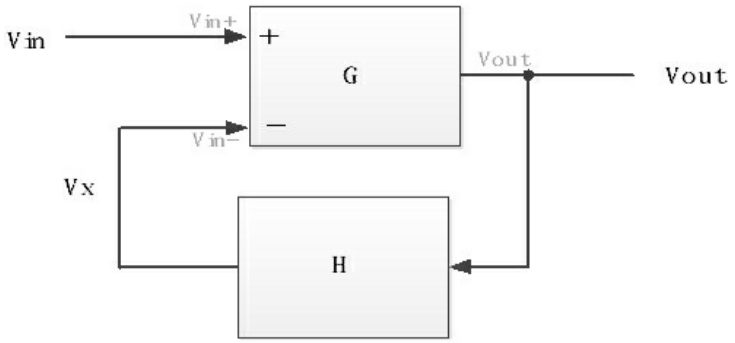
Without much elaboration, let me state that the source of phase delays in our amplifiers are R-C type of circuit elements. You may point out that we have been using transistors and not resistors in the amplifiers and buffers we have built till now. But recall for example our analysis of output impedance of our buffer. We saw that the amplifier we built with transistors has the equivalent circuit of a voltage controlled voltage source (VCVS) with a resistor at its output. The value of this resistor was determined by the small signal parameters (gm and/or gds) of the constituent transistors. The capacitor at the output node could be a combination of the capacitance of the load driven by the buffer as well as the parasitic capacitances of the transistor terminals connected to the output. So we automatically have a ‘pole’ at the output node. In fact, every node in our circuit can potentially give rise to a pole. Given that you can have many such nodes in even a slightly complicated amplifier circuit, you have a scenario where the circuit has multiple poles and therefore, multiple sources of phase. If you recall from an earlier chapter, I had mentioned that the origin of these new complications was the increasing complexity we were looking to create. For example, to idealize our buffer, we were looking to get to amplifier topologies that resulted in much higher gains - for example, of the order of (gm/gds)^3. These implied building circuits that had many more nodes (than for example, the simple source follower) and thereby many more poles.

We have already seen the effect of having one pole in a circuit (the simple R-C filter). What is the consequence of having more than one pole? Note firstly that the resultant magnitude plot from multiple poles is a product of the individual magnitude plots from each pole. The phase plot is an addition of the phase from each pole. The effect of two poles p1 and p2 on the magnitude and phase plots is shown below:



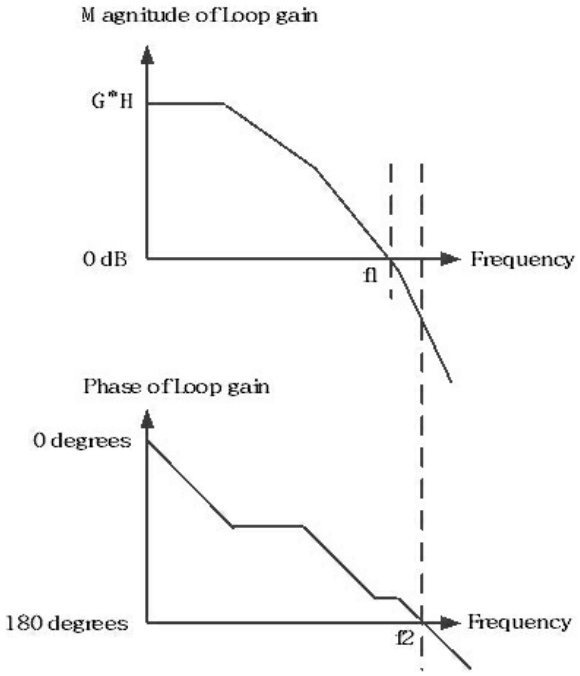
As can be seen, the first pole causes the magnitude to start falling at 20 dB/ decade and at the onset of the second pole, the magnitude starts to fall at 40 dB/ decade. Each pole contributes to a phase of 90 degrees. So at infinite frequency, the total phase from both the poles is 180 degrees.

So if you had two such R-C type elements in your amplifier , you could attain a total phase of 180 degrees, but not at any finite frequency. So we are still safe from the threat of positive feedback. But it would appear that if you had one additional (a third) R-C type of elements in your amplifier, you could end up with a maximum phase of 270 degrees – which means that you could attain a phase of 180 degrees at some finite frequency. Given that you have many nodes in a circuit and that each node can potentially introduce at least one such R-C type of element, it would appear that every amplifier is doomed to oscillate at some frequency. Thankfully, this is not the case. To understand the reason why, let us go around the loop once more with the phase of G (*phi*) set to 180 degrees.



Now let us take a closer look at the feedback loop in the context of the

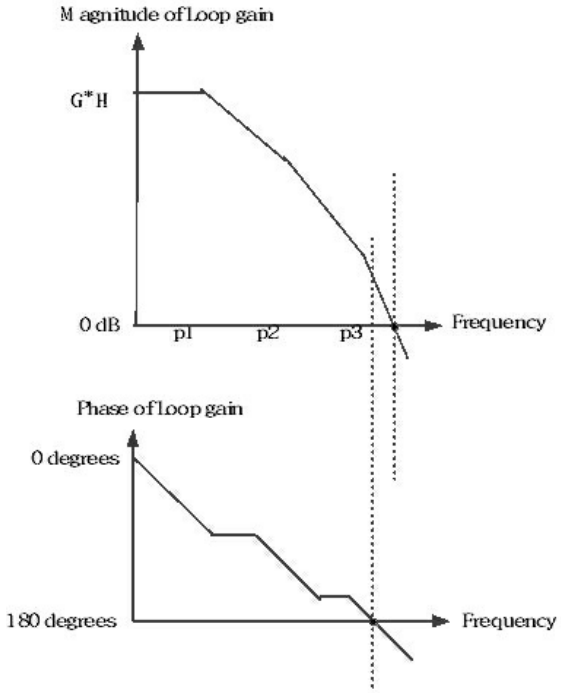
magnitude and phase of G. Let us start with an amplitude of Vout set to 1. Vx would have an amplitude of H. Now, let us assume that the gain G of the amplifier drops with frequency (this should not be surprising if you think of the magnitude response of an R-C filter). At the frequency where the phase of G is exactly 180 degrees, let us assume that the magnitude of G is equal to G'. So going around the loop once completely would require Vout to now have an amplitude equal to H*G'. The case we considered earlier was one where this amplitude was larger than the starting amplitude (which we have assumed to be 1) and therefore, Vout grows bigger and bigger. But what if H*G' is less than 1? This would cause Vout to reduce, and you can easily infer that in steady state, Vout would be back to 0. So there are two criteria required for our feedback loop to become unstable – the first of course is that at some frequency, the phase should go to 180 degrees. The second criteria is that at that frequency where the phase goes to 180 degrees, the magnitude of G*H' should be larger than 1. In fact, the product of G and H, which is the gain going around the loop, is referred to as the 'Loop gain' and is the most important parameter of interest while analysing the stability of our feedback loop. Analysing the frequency response of the magnitude and phase of the loop gain tells us directly whether the feedback loop is going to be stable or not. We call such plots as Bode plots and they look like below.



You can see in the above plot that by the time you get to the frequency (f2)

where the phase reaches 180 degrees, the magnitude has fallen to well below 0 dB (‘1’ in the linear scale). So a feedback loop with such a Bode plot would be indicative of a stable system.

Contrast this with a Bode plot like the one shown below.



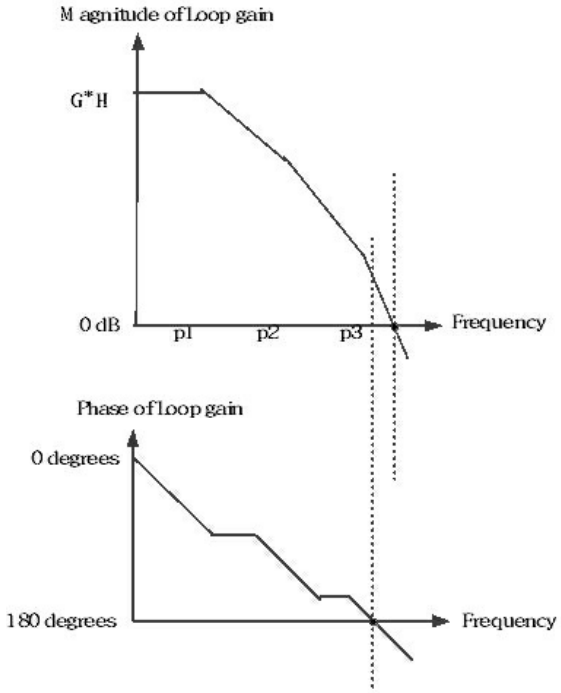
In the above Bode plot, at the frequency where the Phase of the loop gain has gone to 180 degrees, the Magnitude of the loop gain is still above 1. So such a Bode plot would be indicative of an unstable system.

The frequencies marked as p1, p2 and p3 come from characteristics of the circuit, specifically from the R-C type elements we talked about. Similar to the notation we used for the simple R-C filter, each of them is referred to as a ‘Pole’ of the circuit. Each such pole causes the magnitude to drop faster and also the phase to go towards 180 degrees.

So we have understood the mechanism of how feedback can go totally awry. It is, as Ang-Lao says, all in the ***phase*** – or more accurately, the phase in relation to the magnitude. But with the insight that Aman-Ra has been able to get, can he get the mechanism of feedback to work all over again? Can he use his newfound understanding to remedy his relationship?

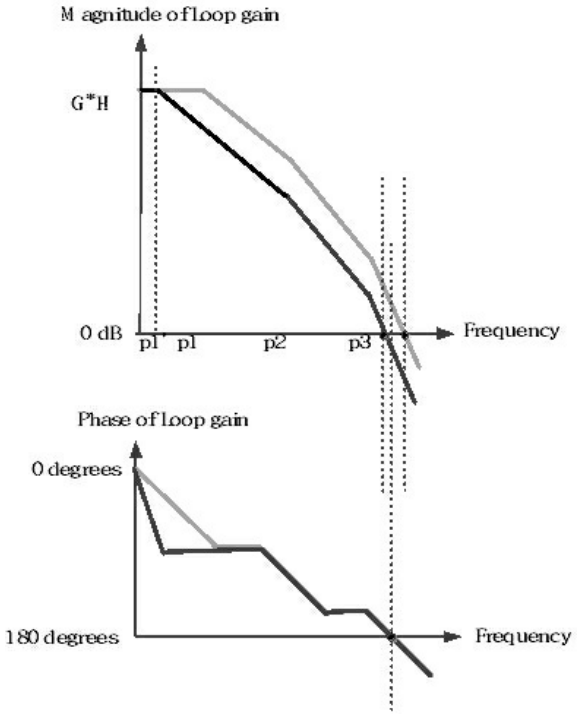
A simple idea for making a feedback loop work

So let's assume we designed an amplifier and ended up with a Bode plot as below. As mentioned earlier, the phase becomes 180 degrees at a frequency where the amplitude is still higher than 0 dB; and so such a feedback loop would be unstable and prone to oscillations.



Can we do a simple tweak to it to make it stable?

As we have seen before, a feedback loop is unstable if the magnitude of the loop gain is 1 or greater at the frequency where the phase of the loop gain goes to 180 degrees. The key to getting the system to become stable is therefore to get the magnitude to drop to below 1 *before* the phase reaches 180 degrees. If you look at the above Bode plot, you can probably think up many ways to do that. But here's a commonly used technique. Let me tell you the idea – what do you think would happen if you kept everything else the same but you pulled in the first pole closer to zero frequency – from p1 to p1' as shown below. The other two poles, p2 and p3 continued to remain at their original location. The original magnitude and phase plots are shown in grey lines and the new plots (on account of pulling in the first pole) are shown in black lines.



You can see from the above Bode plot that pulling in only the first pole closer to zero frequency (keeping the location of p2 and p3 the same) causes the magnitude to start falling earlier. In any case, the contribution of the first pole to the overall phase is only a maximum of 90 degrees. So pulling p1 to p1' does not have much effect on the total phase close to the frequency where the phase goes to 180 degrees – this is because the phase contribution to the first pole has already maxed out at 90 degrees in both cases. However, by pulling in p1 to p1', we have gotten the magnitude to start falling from an earlier frequency. The magnitude has fallen that much more so that by the time we get to the frequency where the phase reaches 180 degrees, we have succeeded in pulling the magnitude below 0 dB. We have stabilized the amplifier by making p1' the 'dominant pole'. Such a technique is one of the popular techniques of stabilizing amplifies. But how does one make p1' the dominant pole – simplistically by identifying the node that is giving rise to p1 and adding extra capacitance at that node so as to pull it to a lower frequency p1'. While the poles were the mechanism threatening to introduce phase and cause the amplifier to become unstable, we have used the same mechanism (by creating a dominant pole) to get back the amplifier to behave in a stable manner.

While we have not gotten into the exact circuit topologies related to the above concepts, we have enough insight now to offer a solution to Aman-Ra.

In fact, without knowing the A-B-Cs of circuit design or feedback theory, he has figured out a lot of these concepts himself on the relationship front. So let us look at what his insights are and compare them against the ones we have gotten on the circuit front.

Aman-Ra has figured out the equivalent of the 'poles' in his life – they are quite simply put, the factors that have conditioned his mind, and that cause resistance to him responding to the actual stimulus on hand. An example is the social conditioning that it is out of place for a woman (read wife) to express her opinion forcefully. There is also the constant emotional conditioning by his mother and her constant reminder that his first duty is towards her. We have already compared the constant pressure from his manager, Gamen-Ra to be like a resistive load. Think of it as contributing to an additional capacitive load, and there, you have one more 'pole'. We have already identified three such poles, each playing its role in the 'stability' of his relationships, and causing him occasionally to react in a less-than-ideal manner.

Having identified the poles and having realized the disastrous effect they can cause, what Aman-Ra comes up with next as the solution is pure genius. He discusses at length with his wife about all of his insights. And he tells her '*the next time there is any stimulus that threatens our equilibrium, we just need to come up with a simple, dominant routine that will precede all else.*' And they think long and hard about what that routine is going to be. And finally they come up with what that routine will be – it is their equivalent of '*five deep breaths*' or '*counting till ten*'; simple techniques that modern-day relationship counsellors advocate.

And this is what they come up with as simple rules every time something big or small threatens to rock their world.

> There will be no shouting or blaming each other…
> Changing into our swimming attire …
> We will go to the River Nile …
> And half-submerged in water, we will hold in our minds the image of Ang-Lao …

And this simple ritual is what will become the guiding practice in their relationship for as long as they live … their ‘dominant pole’!

The Sun sets that evening on a very happy Nile.

Copyrighted material

Copyright © 2015, by Anand Udupa. All rights reserved.