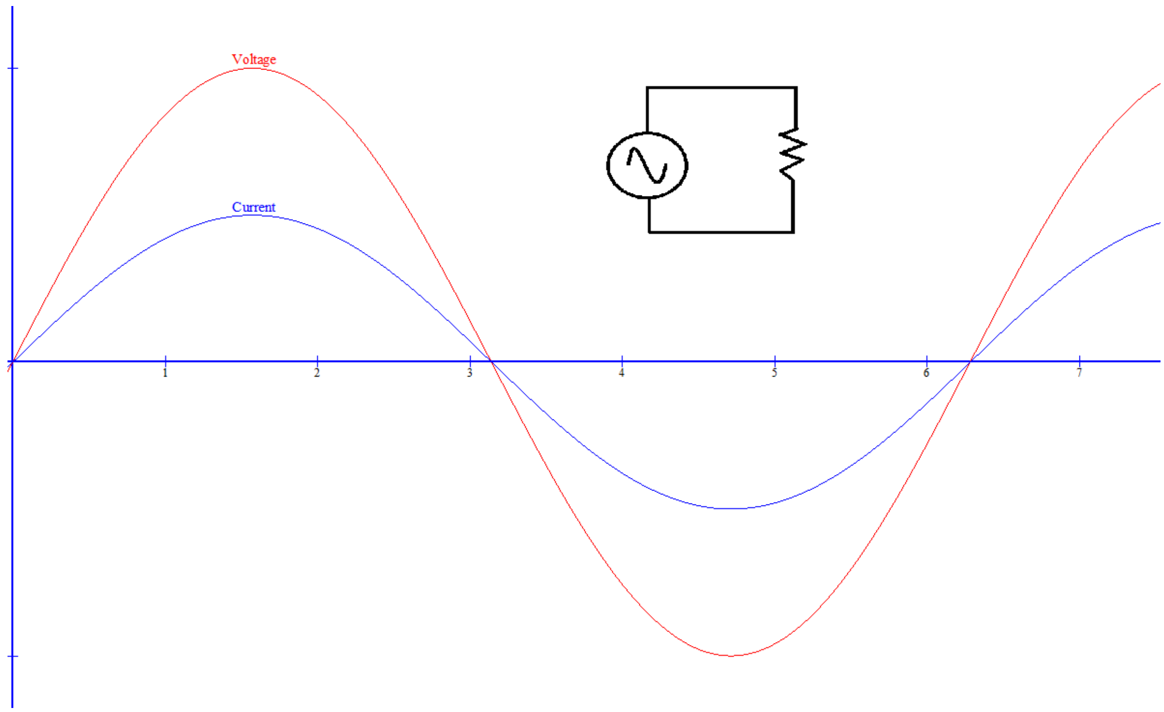
Uncomplicating Complex Numbers

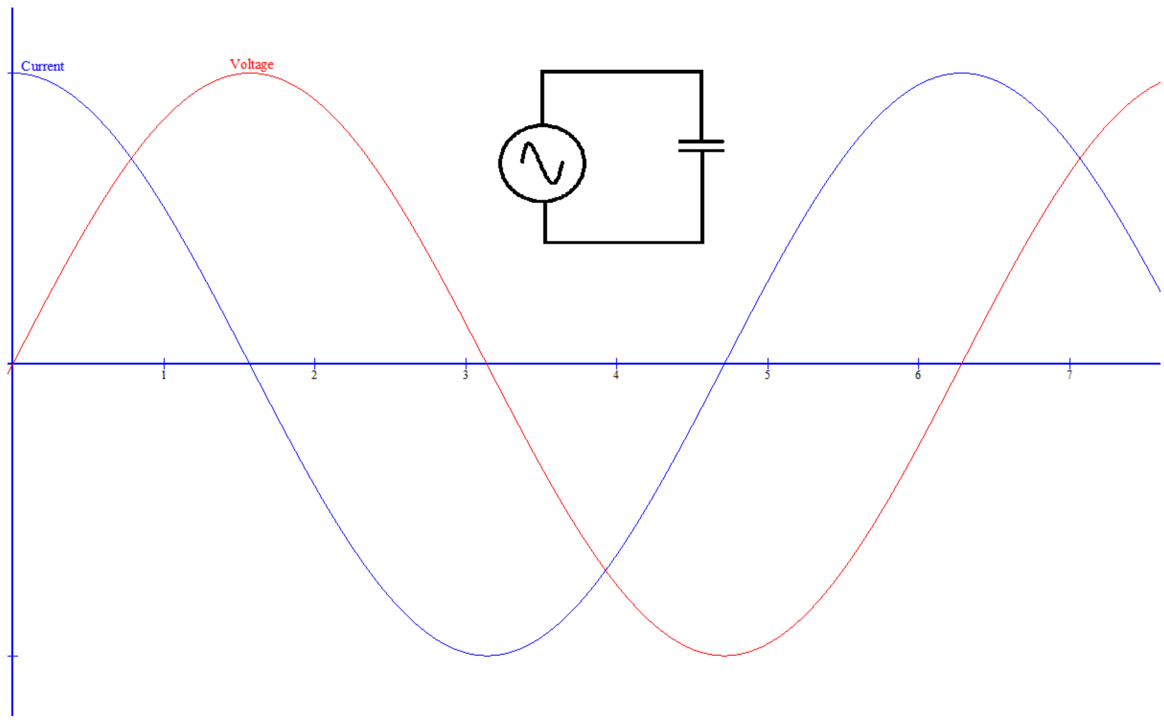
When most people teach themselves electronics, they initially limit themselves to the theories relating to DC and low frequency AC because that is what we were familiar with when we gave ourselves that electrician title. Therefore, when we think of amperage, the flow of current, we naturally come to assume that it is due to a voltage being applied to a load, and that if we give it more voltage, that whatever the current is, it will also go up.



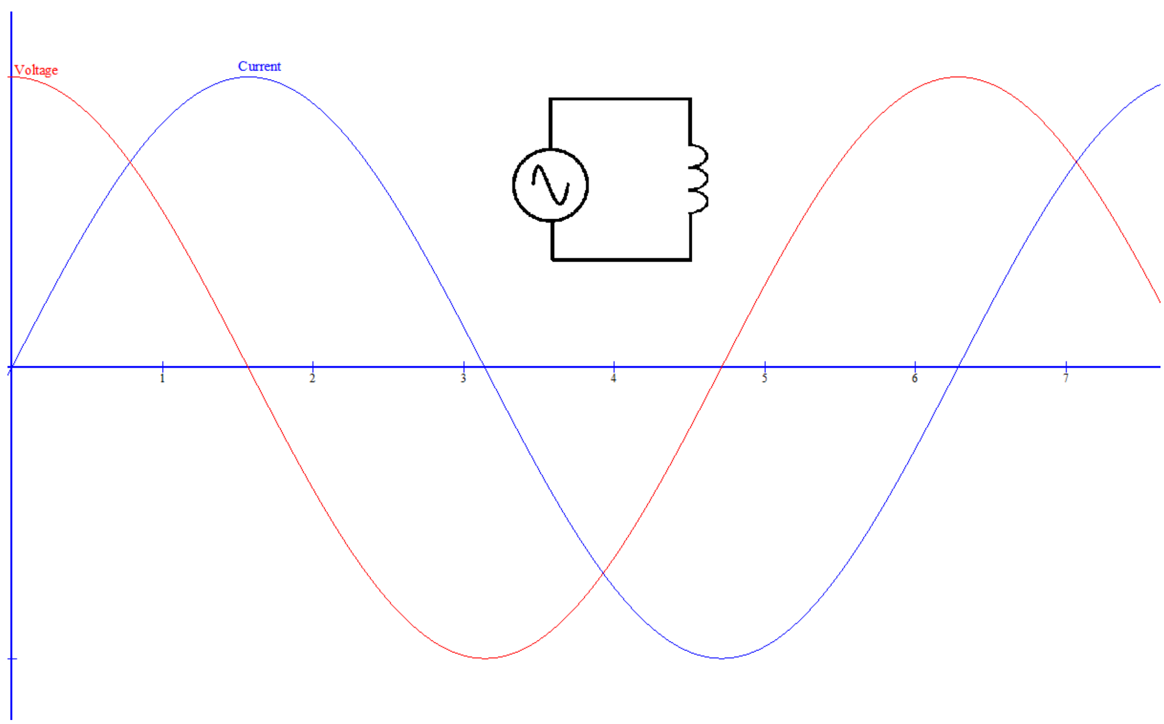
All that is fine and correct, but the problem is, a vast majority of us automatically make the assumption that when the voltage is at its highest, so is the current - and that is where the major problems start. They don't have to be highest at the same point in time!

Reactance

Here are a couple visual examples to illustrate this. Take a 9v battery and a fully discharged electrolytic capacitor and use the battery to charge the capacitor. What you might notice is that when the capacitor is dead, it takes a lot of current to charge. It may even make a tiny spark at the battery terminal when first connecting it. But as the capacitor charges and the voltage is getting closer to that of the battery, there is less demand on the battery to do work, in other words, it uses less and less current. When the capacitor reaches it's "full" charge, there is virtually no current flowing from the battery to the capacitor. Here we have a situation where the voltage is at its maximum, but there is no current flowing. In an AC system where this cycle repeats, the relationship between voltage and current looks like this:

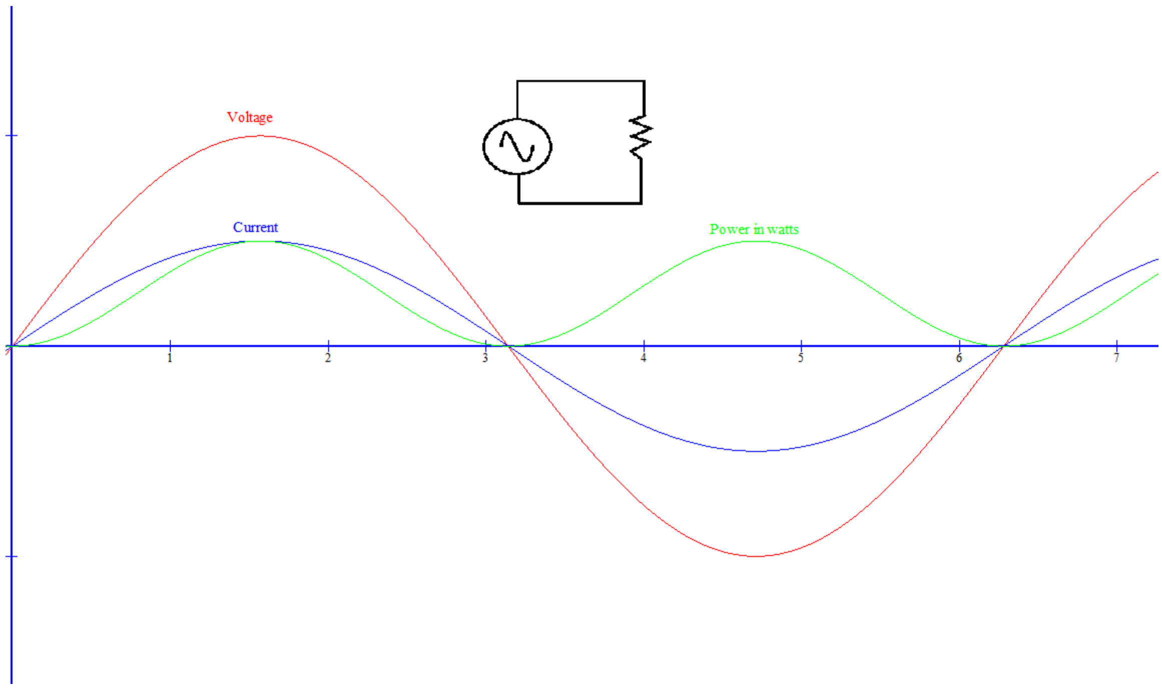


Similarly, if we use that 9v battery to activate an electromagnet (inductor), we see that as the magnetic field builds, the voltage is going down while the current is going up. When the coil is drawing as much current as can pass through it (the magnetic field is maximum), you will see that the voltage is at its lowest. If the coil had no resistance, the voltage would go to zero. Again looking at this behavior with an AC source, the relationship between the voltage and current looks like this:

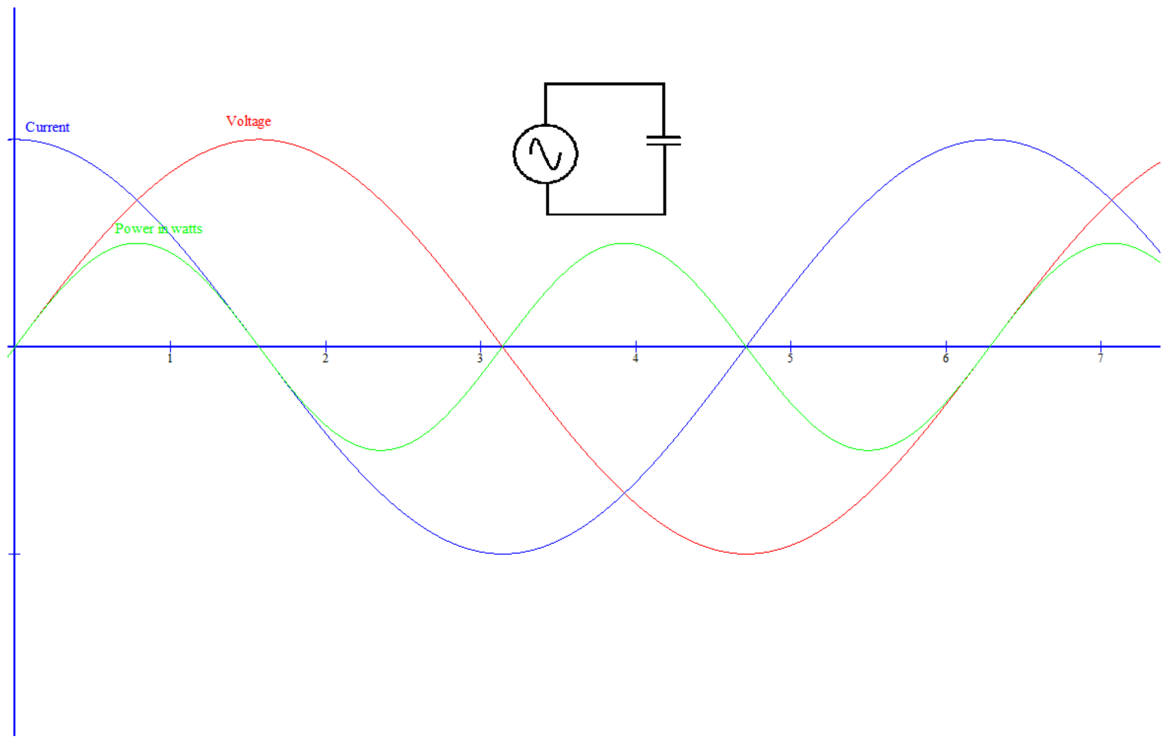


In other words, the voltage (E) on an inductor (L) leads the current (I) and, the current (I) in a capacitor (C) leads the voltage (E). "ELI the ICE man" will help you remember. For ideal capacitors, the phase shift is -90°, and for ideal inductors, it is 90°.

Lets look at the power used next. In a resistor, the voltage and current are always in phase, and since a positive times a positive always equals a positive value, and likewise a negative times a negative will also always equal a positive, we see that for just a resistor, the power used during an AC cycle is always positive, peaking every half cycle (I'm using an arbitrary current here so the waves aren't on top of each other and the x-axis values can be ignored right now). The important thing to take away from this is that the power is always positive.



But what if we did the same thing with a capacitor or inductor instead? Let's use a capacitor in our drawing. During half of the waveform, the voltage and current will have opposite signs, so that means the power put into the circuit during the first 90° will be put back into it during the next 90°. The power in watts is going positive *and* negative.



Said in a different way, to resist the change, it gives back some of what it has stored, or, takes more. This is called reactive power. Ohm's law says volts*amps=watts, right? With the resistor, our watts are always positive. With a capacitor or inductor, the watts go positive and negative, they put back what they take. Since it is putting power back every time it gets some, it does not generate heat like resistors do, and therefore, we *cannot* call it resistance. We call it ***reactance***. And if we know the frequency of the AC source and we know the capacitance or inductance, we can find out how much reactance (opposition to the flow of AC current) there will be.

The formulas for reactance are as follows. For capacitive reactance, we use:

Xc=1/(2*pi*Hz*F)

And for inductive reactance, we use:

Xl=2*pi*Hz*H

Where: Xc=capacitive reactance, Hz=Hertz frequency, F=farads capacitance and H=henry's inductance.

Impedance

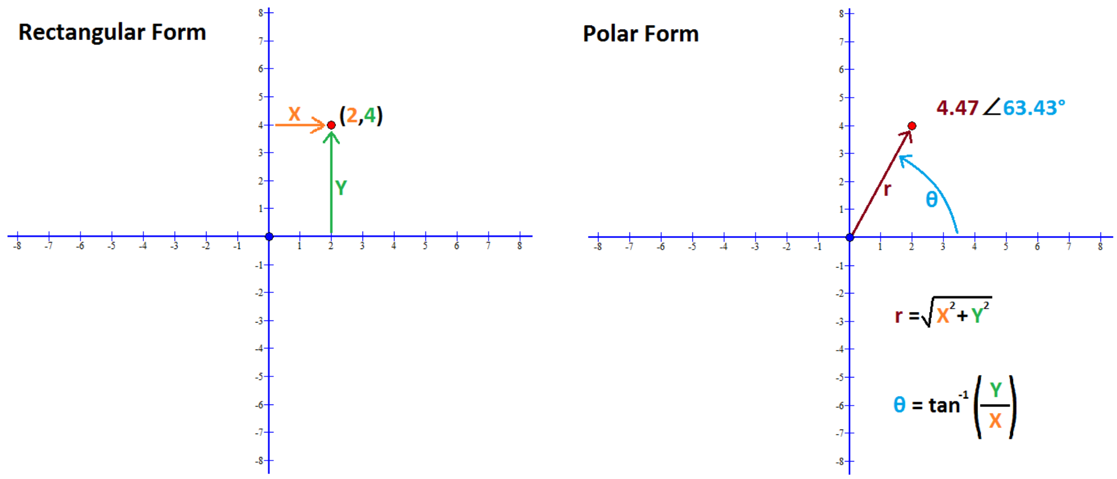
What if we had a load that had resistance *and* reactance? We call the combination of the two ***impedance***. Parts of the circuit want the current and voltage in phase, while other parts of the circuit do not. What we end up with is some amount of resistance and some amount of reactance and the phase between the voltage and current will be somewhere between -90° and 90°. In other words, the max power used will be at some point the voltage and current are not at their peaks. What we need is a way to represent resistance and reactance together, and there are two common ways to do so.

We can write R±jX (ignore that "j" for now, it's just there to scare you), or, we can write it as a magnitude and phase angle like r∠ θ. These are the rectangular form and the polar form, respectively. Lets take a look at that visually.

Represneting Impedances

Lets step outside the world of electrical impedance and talk about how rectangular and polar numbers are used in everyday math, completely forgetting about electronics for just a moment (there's a reason for this).

Do you remember the cartesian coordinate system from math class? And how we can represent any point on the coordinate system with an X and Y coordinate (X,Y)? ...and how that X,Y coordinate can also be represented as a distance from the center r and an angle θ?



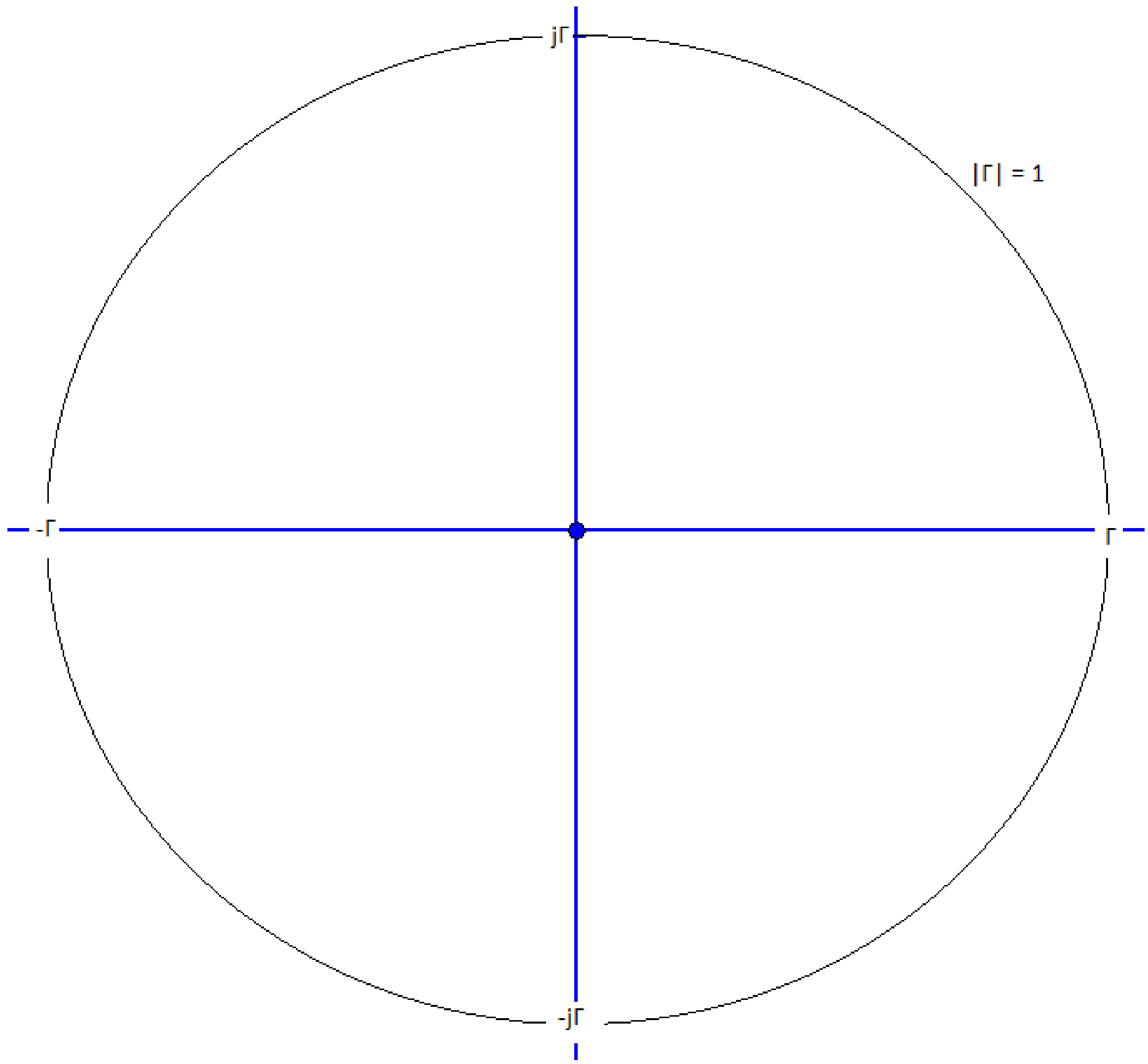
Ok, good, all that is fine and dandy, but we don't use charts like this for electrical impedance (well, not too often) for a few reasons. For one, we do not deal with negative values for R (unless you are building oscillators with negative resistance devices), and two, infinity is a long ways away on a linear graph, but not so far away when playing with RF circuits! Although the concepts of rectangular and polar forms are exactly the same in normal math as they are for radio, we need to draw them on a different kind of chart that can make it all work from 0Ω to infinity. The solution to this problem is using a Smith Chart.

Short Detour

....to making of our own smith charts.

If you do not care about this, you can just assume smith charts work and skip to the section on "Using Smith Charts". I'm gessing, however, that like me, unchecked curiosity is the number one way to hit the brakes on intellectual progress, and that one day you will want to understand the polar readings on your VNA. So before we talk about all the difficult problems a smith chart can *easily* and *visually* solve, lets talk about how we can go from that linear cartesian coordinate system found in everyday math (or the polar readings on a VNA) to drawing a (and plotting polar values on) a smith chart. You can dive deeper into the journey of Smith and Tosaku later, but for the time being, let us be satisfied with the ability to draw our own smith charts with a compass and simple math.

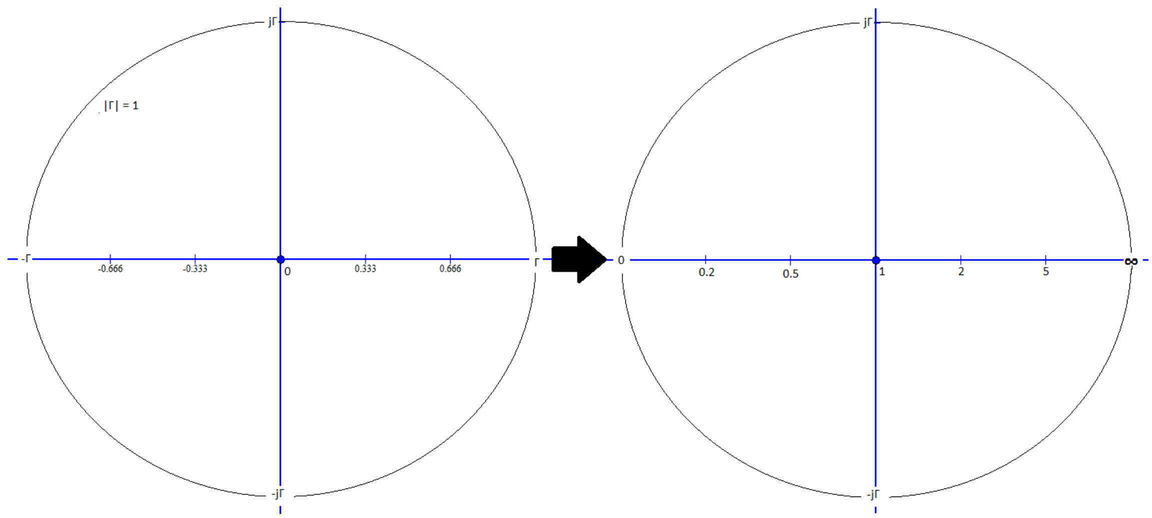
The first step in getting zero on the left and infinity on the right along the x-axis is to first put a negative 1 on the far left and positive 1 on the far right, with zero still in the middle. Then we make a rule that says we can go no further from the center than a unit of 1 and draw a circle around the graph.



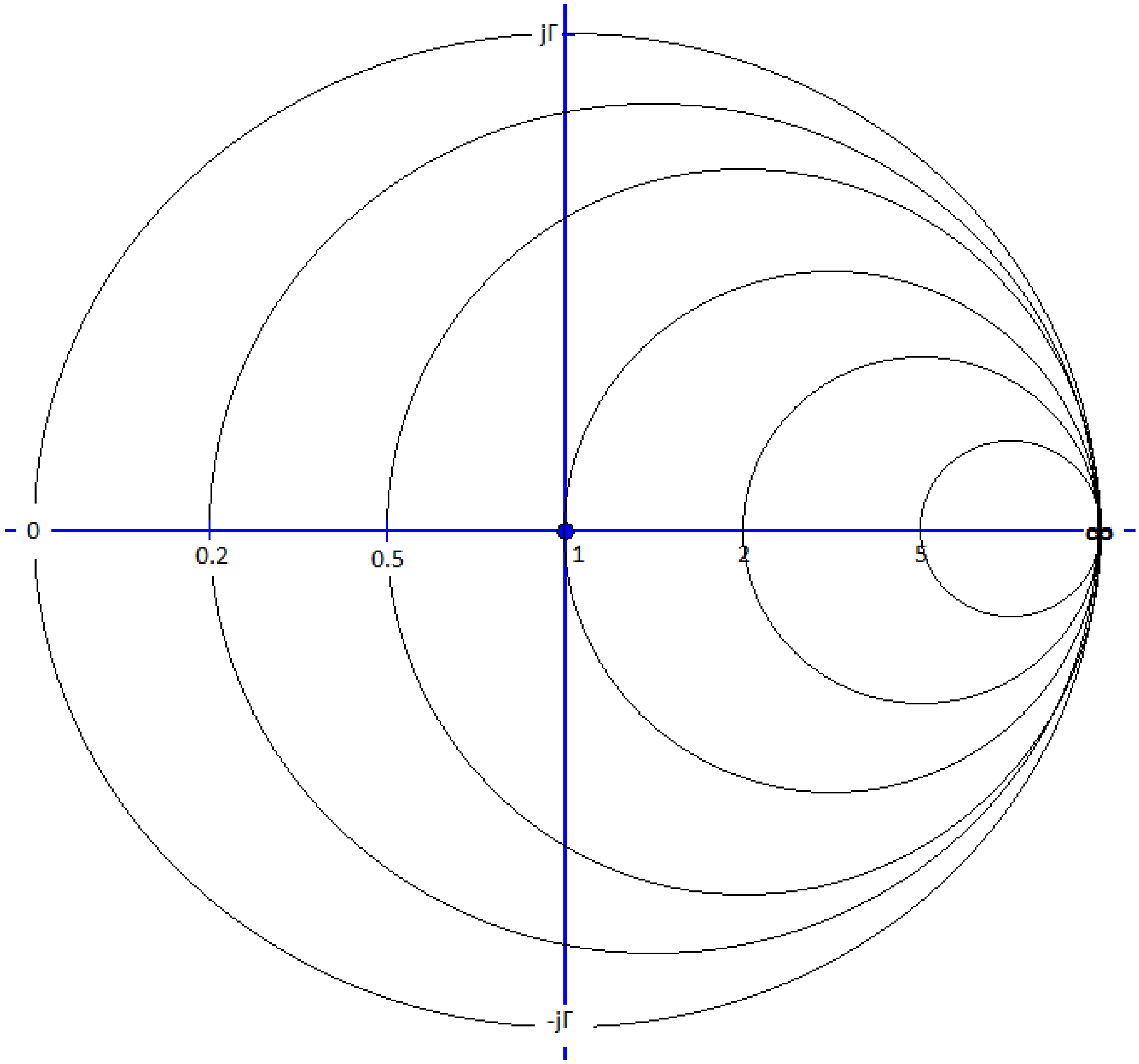
From here, we can draw the constant resistance circles. We will call all values along the x-axis gamma (Γ) and all values on the y-axis as imaginary gamma (jΓ). No need to be fluent in greek, it's just another letter arbitrarilly assigned to represent something (instead of saying x-axis and y-axis), and we call the absolute value (distance from center) the *reflection coefficient*. Remember, you do not need to remember any of this greek stuff or worry about that little imaginary letter j. Now, for every value along the Γ axis (-1 to 1), we do this simple equasion:

(Γ+1)/(Γ-1) = Resistance points on the chart

and replace the Γ value we used with the result from the formula above. For instance, if we have these points, {-1, -0.666, -0.333, 0, 0.333, 0.666, and 1}, we get out {0, 0.2, 0.5, 1, 2, 5, infinity (we will call a divide by zero error infinity here)}.



Now, lets say that those tick marks on the axis are tangent points of circles that have their centers coincident with the horizontal resistance line and midway between the tick mark and the infinity point. Grab a compass and draw them in. Then, lets get rid of that useless y-axis, we are done with it.

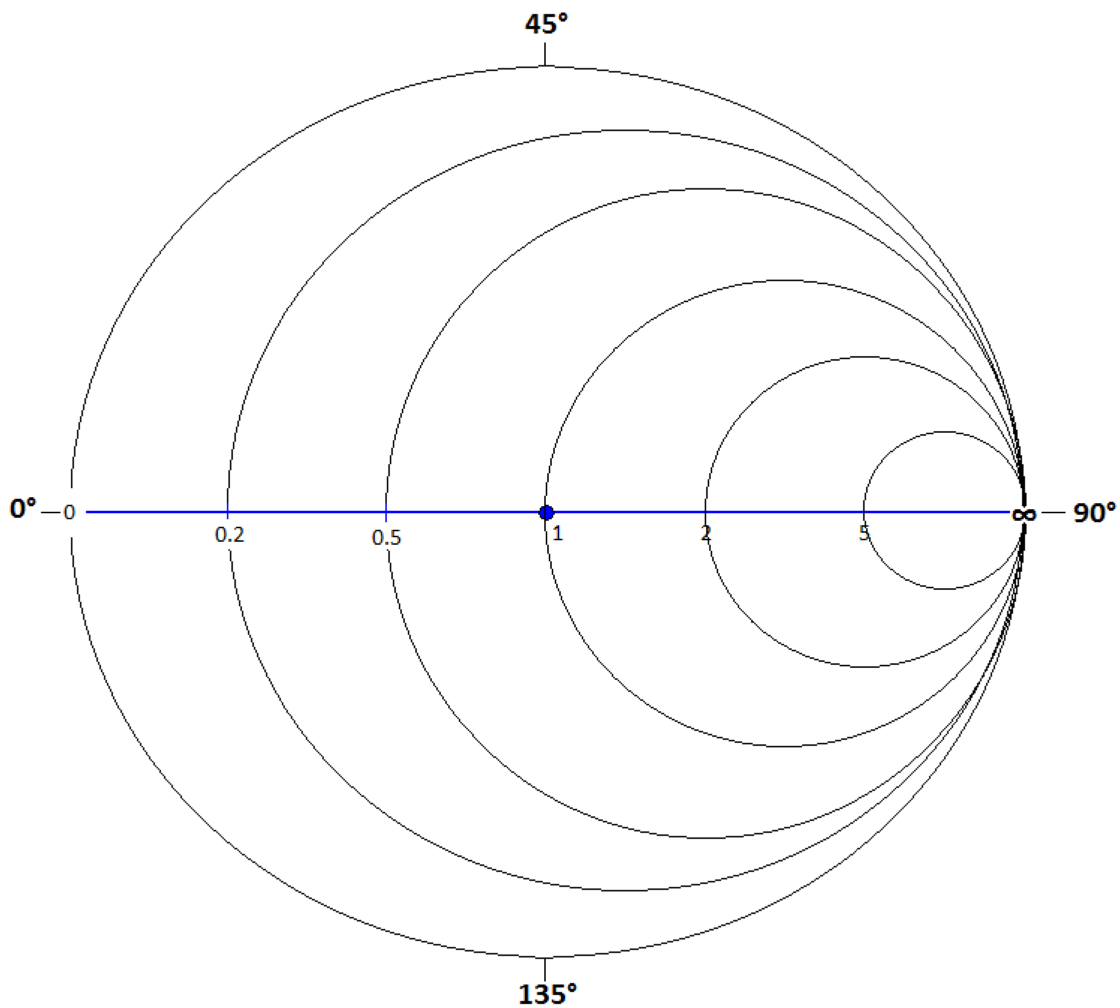


You just made your constant resistance circles, make as many as you want (but remember, the chart will get crowded fast!). Every point along those lines will have the same resistance.

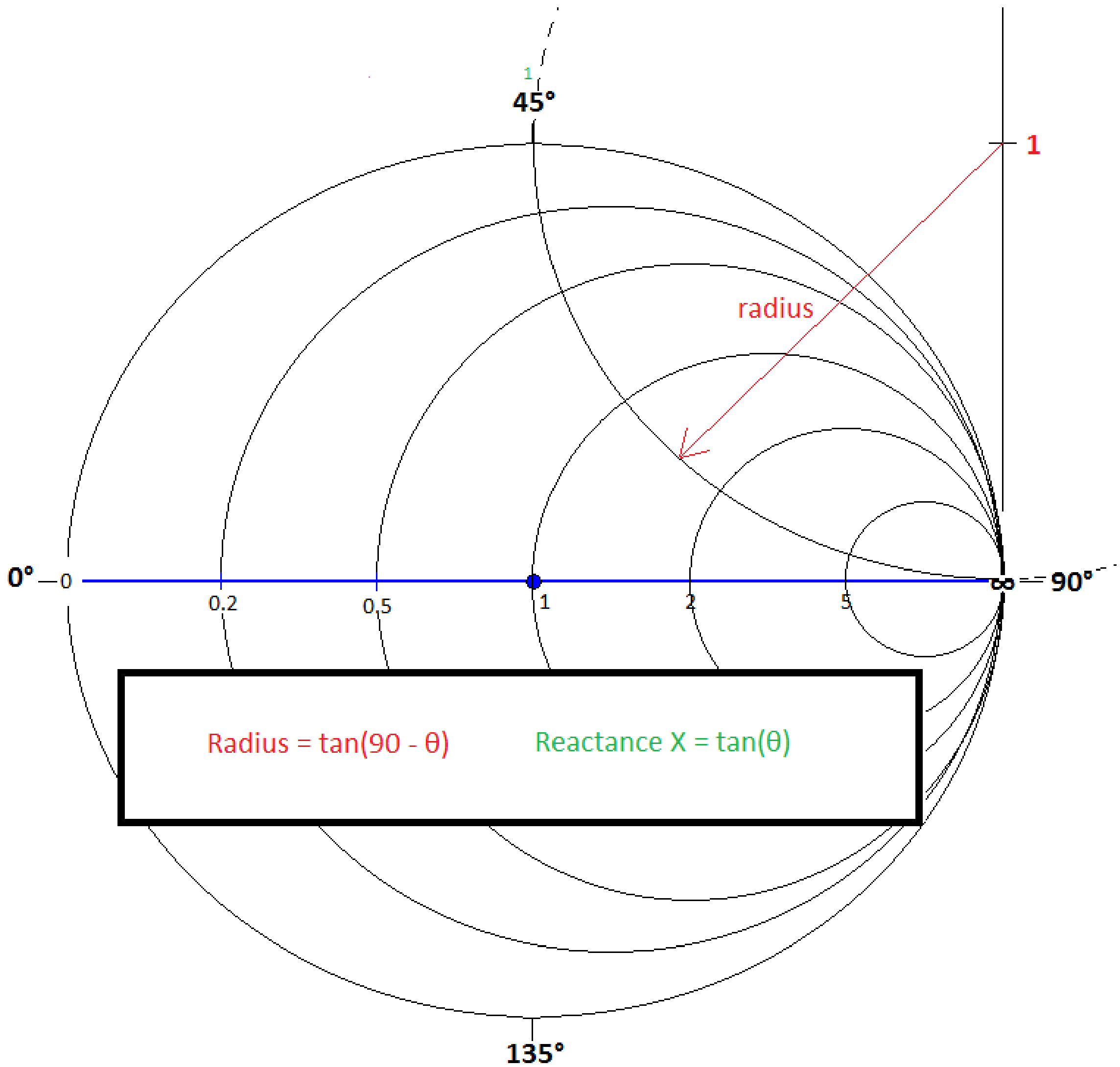
Now, before you say, "hey, they go to infinity, how can they be constant? Does 0Ω = infinity, does 50Ω, do they all?" The answer is no. Assume that all those circles, that for the sake of drawing, were made tangent to the infinity point, but are not actually touching eachother, merely infinitely close to one another. Luckily, for RF stuff, we call anything above a few kΩ's close enough to infinity and never need a microscope to use a Smith Chart!

We can now make our constant reactance curves, the lines on which points along it have the same reactance. But first, just like we chose points along the horizontal center line for resistance, we have to make marks where these will intersect the outer circle as the outer circle is now the "other axis". As it turns out, that axis also represents degrees or wavelengths. *Remember when we didn't want the side of the cartesian coordinate chart that had negative resistance values (and we essentially got rid of that half of the chart)?* Well, we got rid of half the degrees rotation too. On a smith chart, one entire revolution around it represents 180° of electrical travel, not 360°. You could say we took the half we wanted and stretched it into a circle.

We will start out marking the the outer circle with degrees or wavelengths (whichever you prefer). Lets start with 4 points around the chart: 0° (or 0λ) is on the left at 0Ω, 45° (or 1/8λ) is at the top, 90° (or 1/4λ) on the right at infinity, and 135° (or 3/8λ) at the bottom. We return at 0° when we reach 180° (or 1/2λ).



Now it's time to connect the dots like we did with the resistance circles, except this time, the circles get infinitely bigger instead of infinitely smaller. Bisecting the distance between two points and using a compass like before is not going to work here unless you have a very large table! We need to connect all the degree marks around the smith chart perimeter with the infinity point at the right using an arc of a circle, so, what can we do on a small table or computer? Lets imagine a temporary vertical line from infinity straight up. Somewhere along this vertical line will be the center of every arc's circle. We can get a distance to the center of every circle used (its radius) with some simple math based on the degrees we are trying to make an arc for, then measure up the vertical line that far and go nuts with the compass!



Now go to town doing the rest of the constant reactance curves. You just made your own smith chart! Every value of R and X should match up with the polar chart by using this formula:

$$R = \frac{1 - \Gamma_r^2 - \Gamma_j^2}{(1 - \Gamma_r)^2 + \Gamma_j^2}$$

$$X = \frac{2\Gamma_j}{(1 - \Gamma_r)^2 + \Gamma_j^2}$$

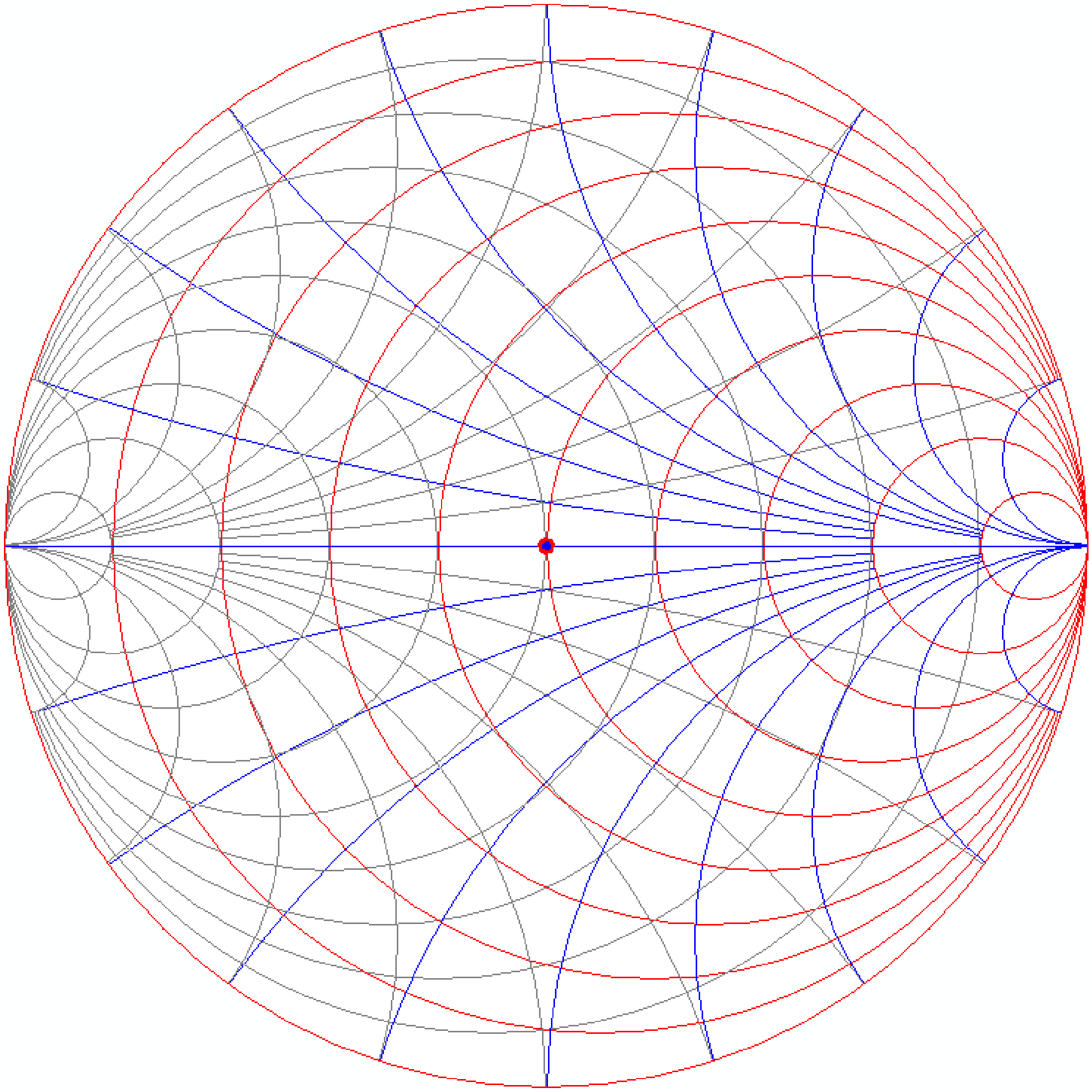
Γ_r = real part of Γ
Γ_j = imaginary part of Γ

Where's 50Ω?

So far, the numbers we have written on our chart have been ***normalized***, which means 1 is still the middle, but it doesn't have to stay that way. We can multiply all the resistance and reactance values by any number, and that number will be in the middle. Lets say we want the middle to represent 50ohm, so if we work with 50 coax, our transmission lines rotate everything about the chart center. All we need to do is multiply everything on this chart by 50!

Admittance

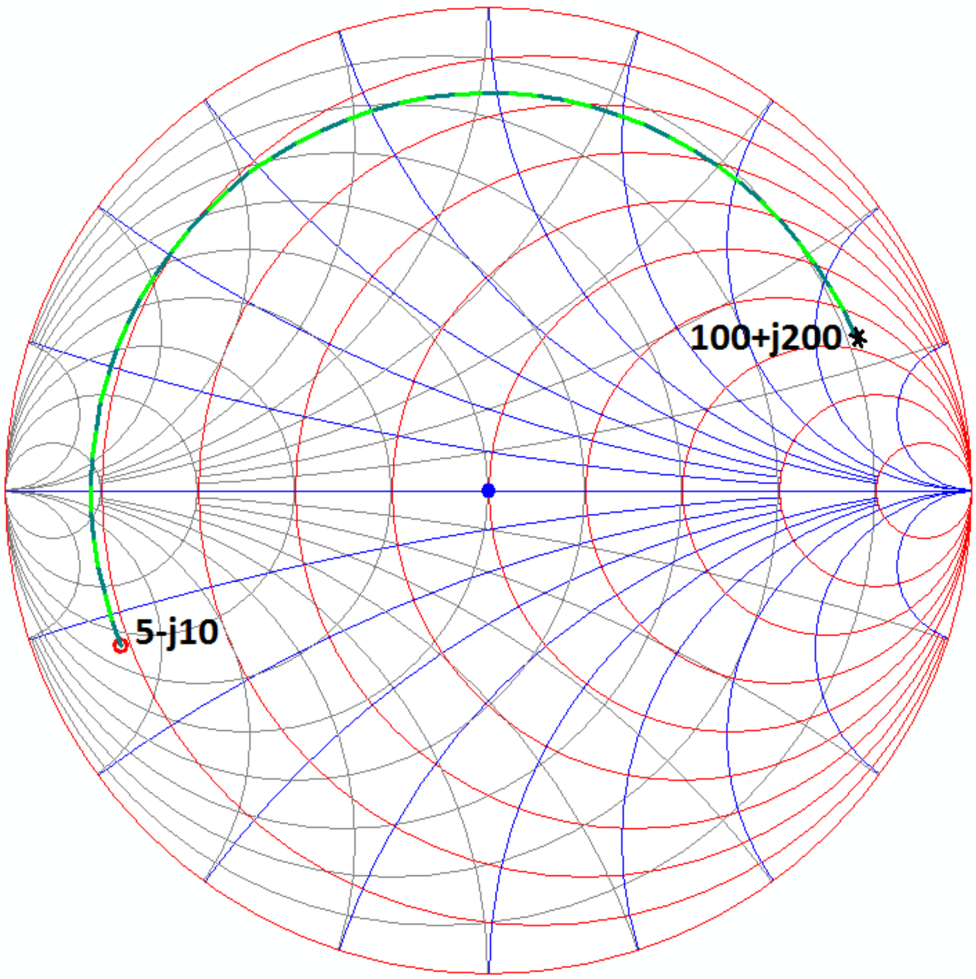
We need one more concept and a few more lines in order to use this chart, but instead of me drawing a complete smith chart, I am going to borrow a screenshot from my favorite free little program Iowa Hills Smith Chart. I read their terms of use, and it seems the only thing they mind is people ripping off their code, I hope they don't mind the screenshots.



Those grey lines are the admittance curves. You could draw them by turning your page upside down and starting the process over. They consist of constant conductance circles and constant susceptance arcs (arc, curve, I use them interchangeably depending on if I am talking about constructing them or using them, correct me if you must). Here's a big problem people run into here. In the DC way of doing things, we find conductance by taking the reciprocal of resistance. You cannot do that here, and I have a few more formulas to give you. Don't sweat it, we will cover all that after we find out how awesome it is to have those gray lines there. The top half of the chart represents inductive impedances, the lower half represents capacitive impedances. If you drew your own smith chart, you would already have discovered that given the - values around the outer perimeter of the bottom half.

Using Smith Charts

Lets dive right into the fun and see what smith charts can do. Say we have a load of 5-j10 and connect a quarter wavelength piece of 50Ω coax to it. What would the impedance be looking into the coax? Since we are dealing with 50Ω, we want to make the center of the chart 50Ω so the coax rotates the impedance around the center and not off on some other arc. We can either multiply everything on the normalized chart by 50 or just print one off from the internet (or divide what we are working with by 50 and plot those on the normalized chart that has 1 in the middle). Once we have our chart and we find 5-j10, we rotate about the center clockwise (because we are going away from the load toward the generator) 90°, or a quarter wavelength. Where the compass stops is the input impedance of the coax with the mismatched load connected to it (neglecting coax loss).



Thats how easy it is to calculate what a coax will do to an impedance value. Real coax has loss, and if you were to have a couple wavelengths of coax, you would see that circle spiraling inwards. On good smith charts, there is a dB loss scale that makes it easy to adjust for that on the chart. This is the first example of how a smith chart can solve a complicated math problem. The formula to do that, if you wish to try, is:

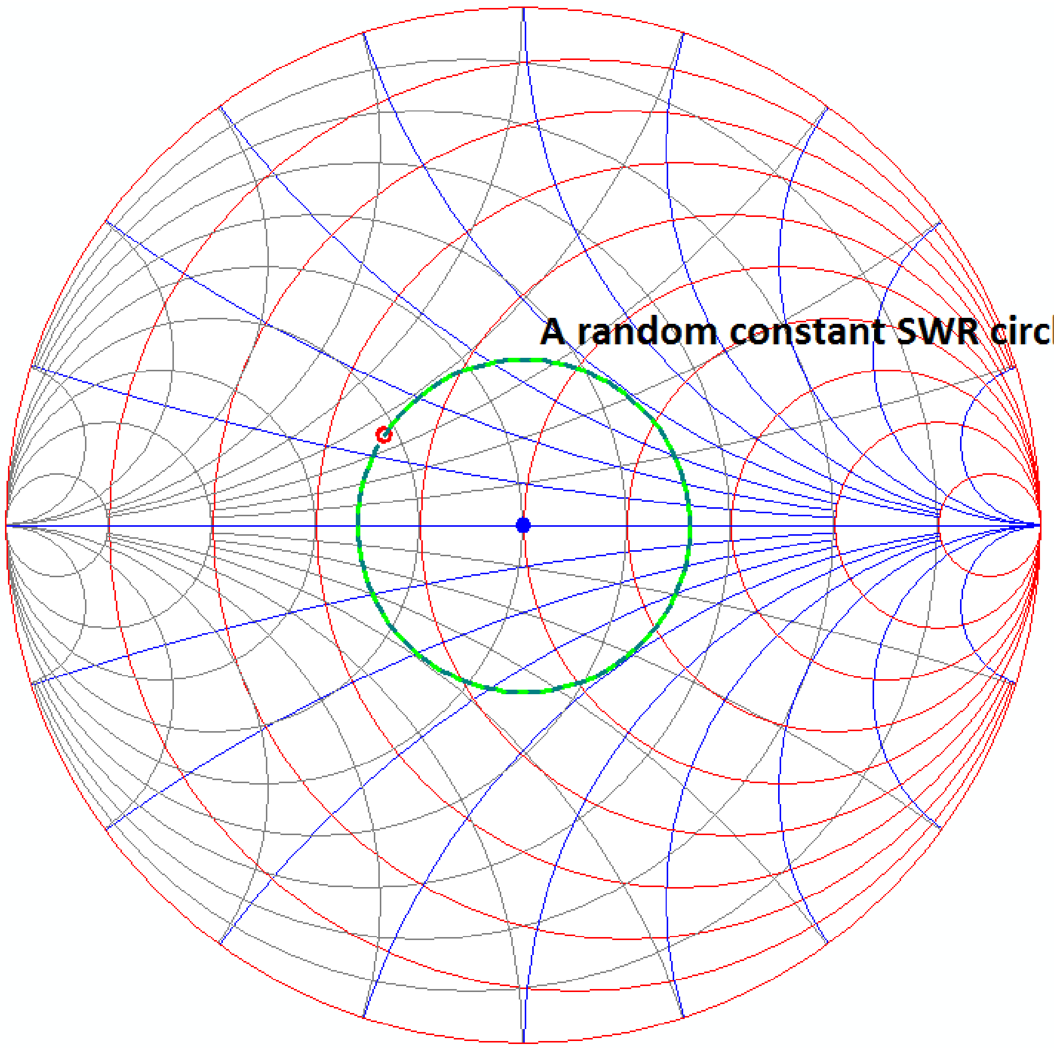
$$Z_{in} = Z_0 \cdot \frac{Z_L \cdot \cosh(nl) + Z_0 \cdot \sinh(nl)}{Z_L \cdot \sinh(nl) + Z_0 \cdot \cosh(nl)}$$

Z_{in} = complex impedance of the coax input
Z_0 = complex impedance of the load at the other end
Z_L = characteristic impedance of the coax
n = complex loss coefficient = a+jb
a = matched line loss attenuation in nepers/unit length (divide dB by 8.688)
b = phase constant of line in radians per unit length
l = electrical length of line in same units used for a

formula taken from the ARRL Handbook, and correcting for the typo that says to multiply rather than divide in the definition for b

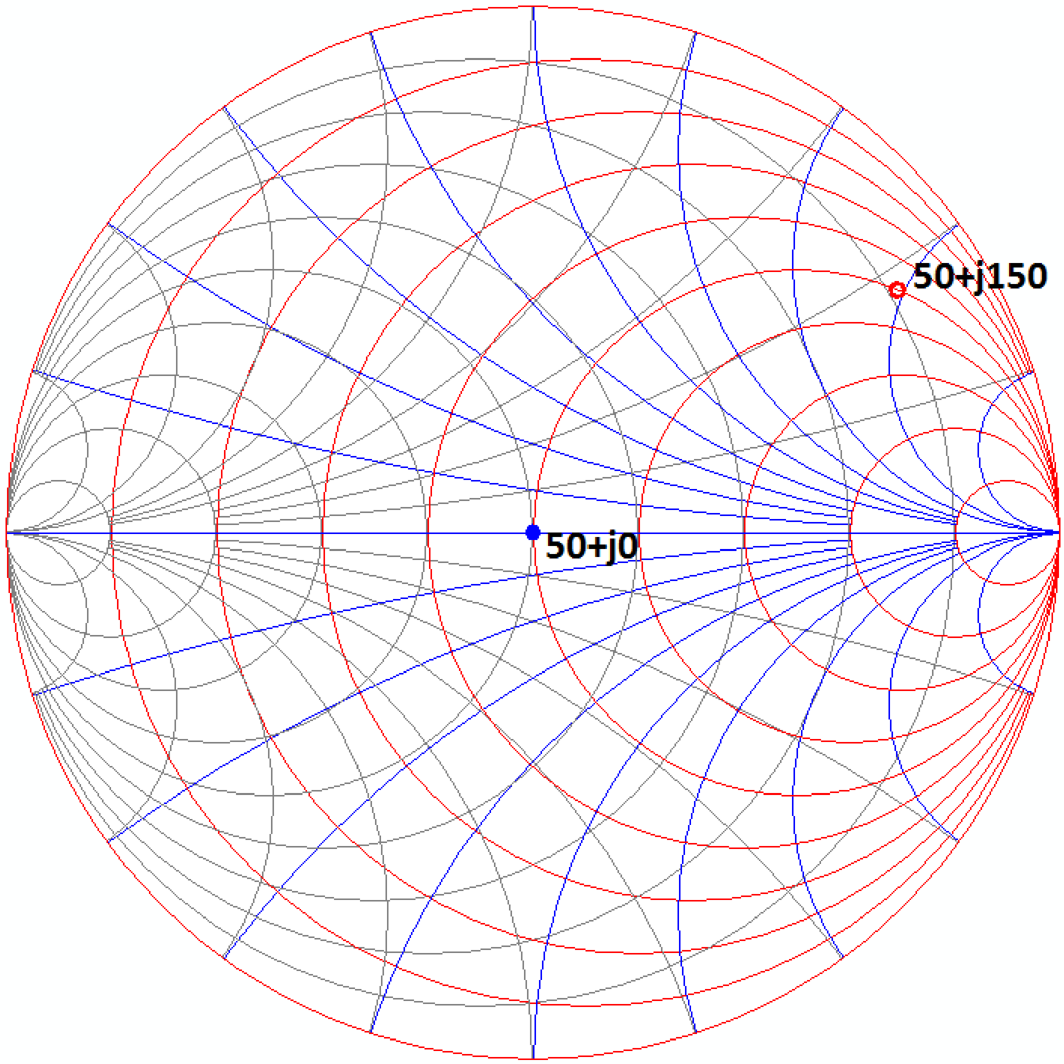
Constant SWR Circles

...appear on the graph just like a coax a half wavelength or longer with a mismatched load (otherwise it would be in the middle...) would. *Any* random radius circle centered on the chart represents the impedances that will all have the same SWR. In other words, if you find the absolute value of R and X, the absolute impedance |Z|, from anywhere on that line, the result will be the same. Remeber the polar graph when we did that for "r", how the radius is just the Pythagorean theorem. If it is the same distance from the center, the SWR will be the same. Always. So when you hear someone say coax transforms SWR, tell them it transforms impedance, not SWR, and that the only way it can affect SWR is if the coax is lossy. That "better SWR" due to coax is just a lie because your SWR meter cannot read a reflection that never makes it back past the loss!

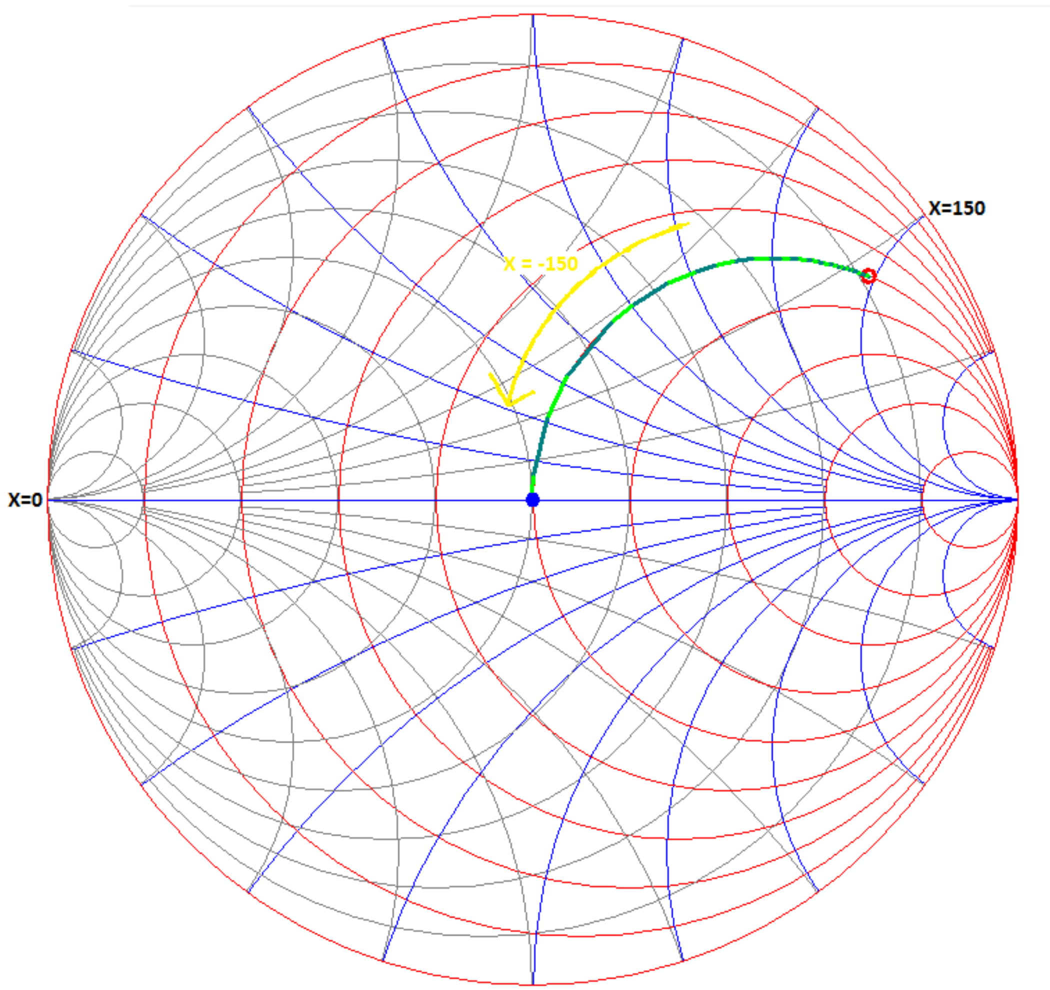


Capaci tors and Inductors

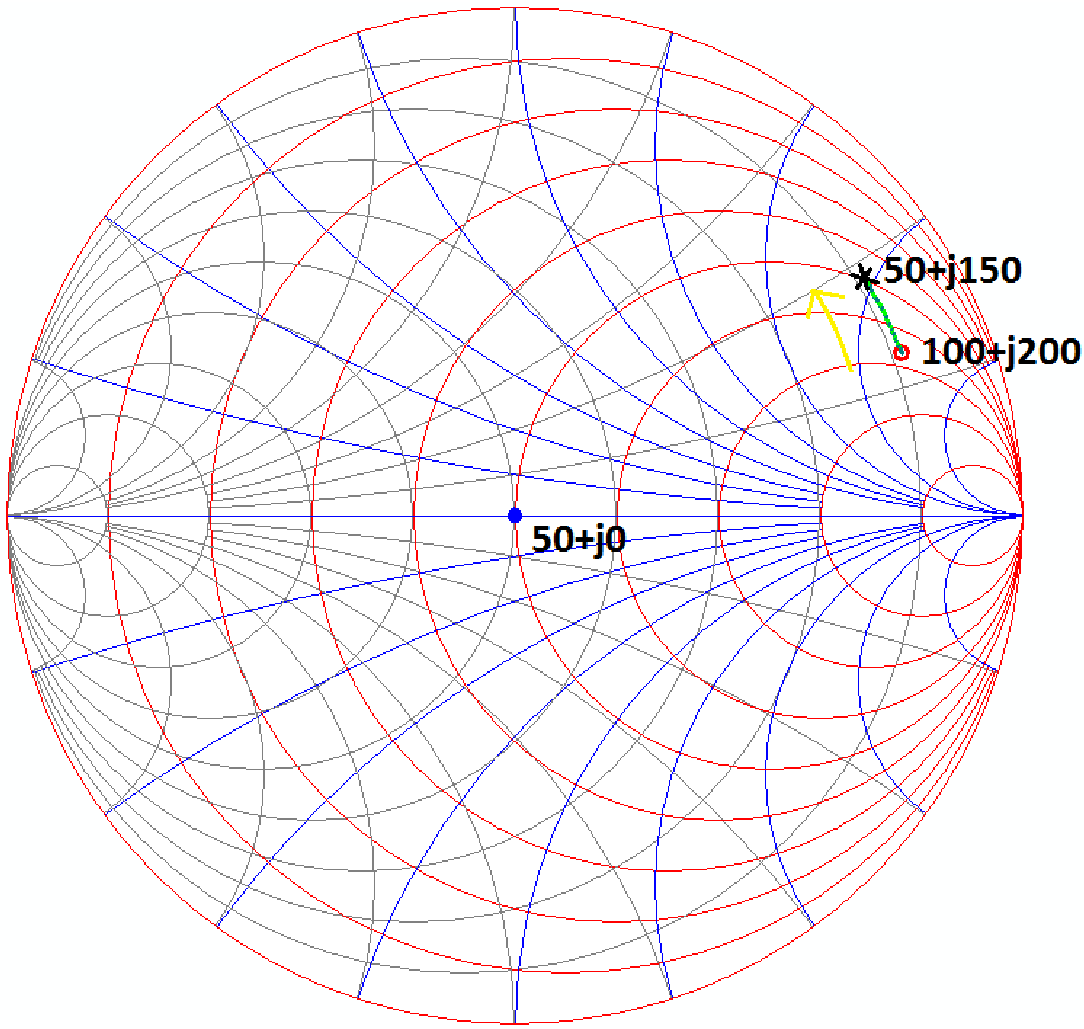
Let's pick a random spot on the chart. Lets say 50+j150. We need to get to the middle.



Do you see how the red point happens to be on the same red line (the same constant resistance circle) as the center of the chart? That is really handy because all series reactances will follow those red lines. Since we want to go down, toward the capacitive side, we know a series capacitor will follow that line down. Since we have 150Ω of inductive reactance, we know that (-)150Ω of capacitive reactance will cancel it.



But what if we didn't start out on the 50Ω constant resistance circle. What if we started at 100+j200? What we would do then is simply follow the *constant conductance circles* with parallel components! Again, going up, so we need an inductor. This time in parallel. We need to go up ***to*** that red line we used last time because, again, it gets us to the middle with series components. From there, a series capacitor finished the path to the center like above.



The issue here is that when using parallel components to move around the chart, we need to follow the *admittance* curves, and the difference in *admittance* between the two points will be the admittance of the inductor (or capacitor) we are using. However, remember earlier when I mentioned admittances *cannot* be found by taking the reciprocal like our DC days tell us? We can do that on the chart with a ruler and no effort, or, with a calculator. Here are the formulas to convert between impedance (R±jX) and admittance (G±jB):

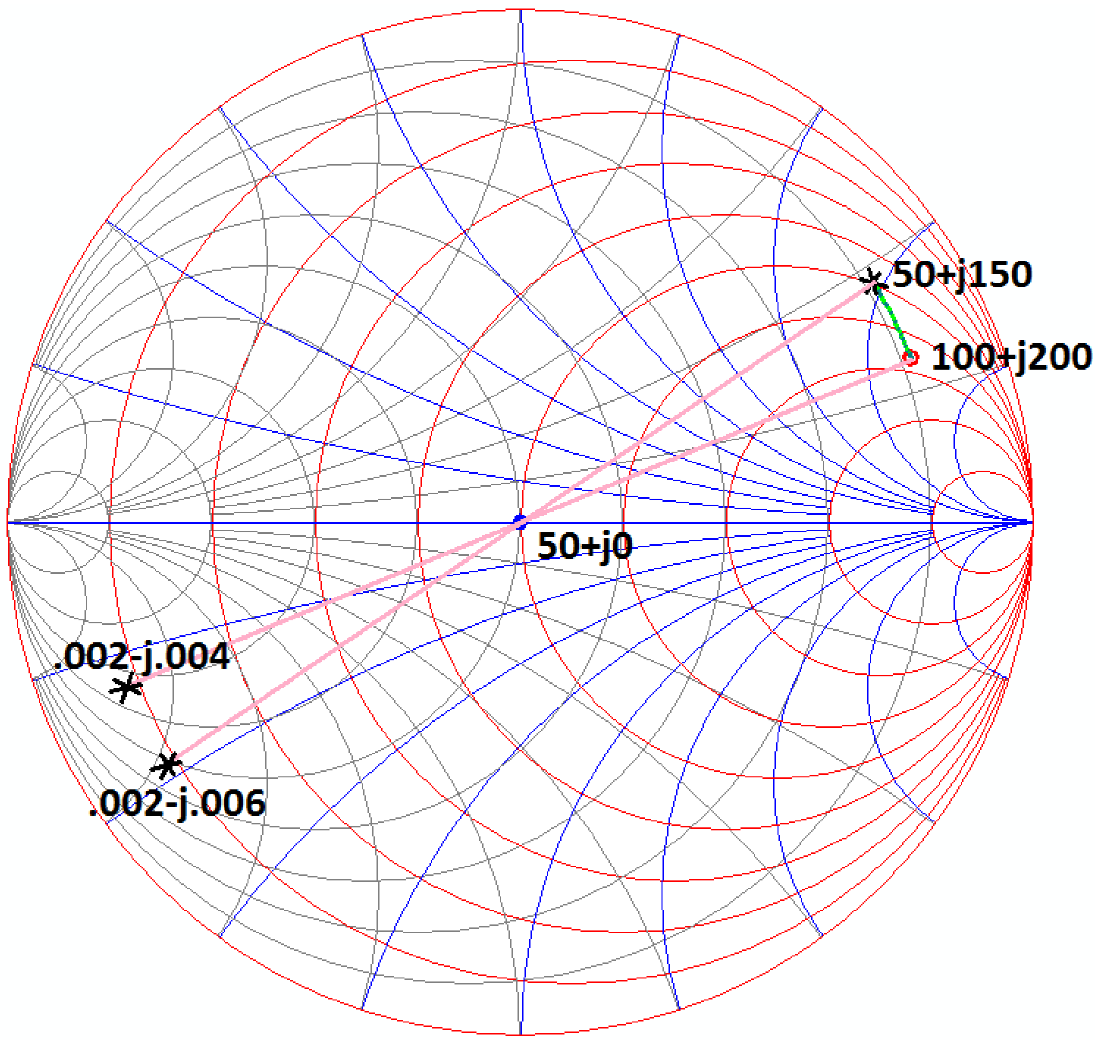
From impedance to admittance

$$G = \frac{R}{R^2 + X^2} \qquad B = \frac{-X}{R^2 + X^2}$$

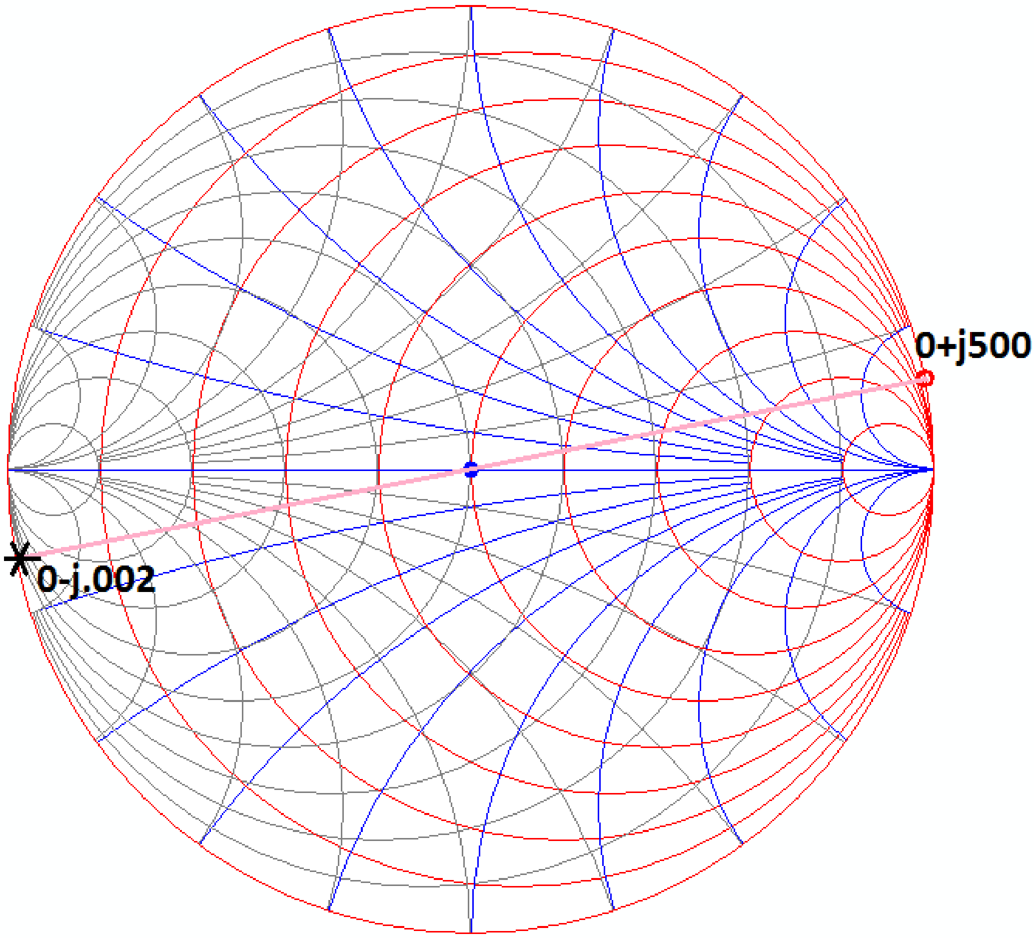
From admittance to impedance

$$R = \frac{G}{G^2 + B^2} \qquad X = \frac{-B}{G^2 + B^2}.$$

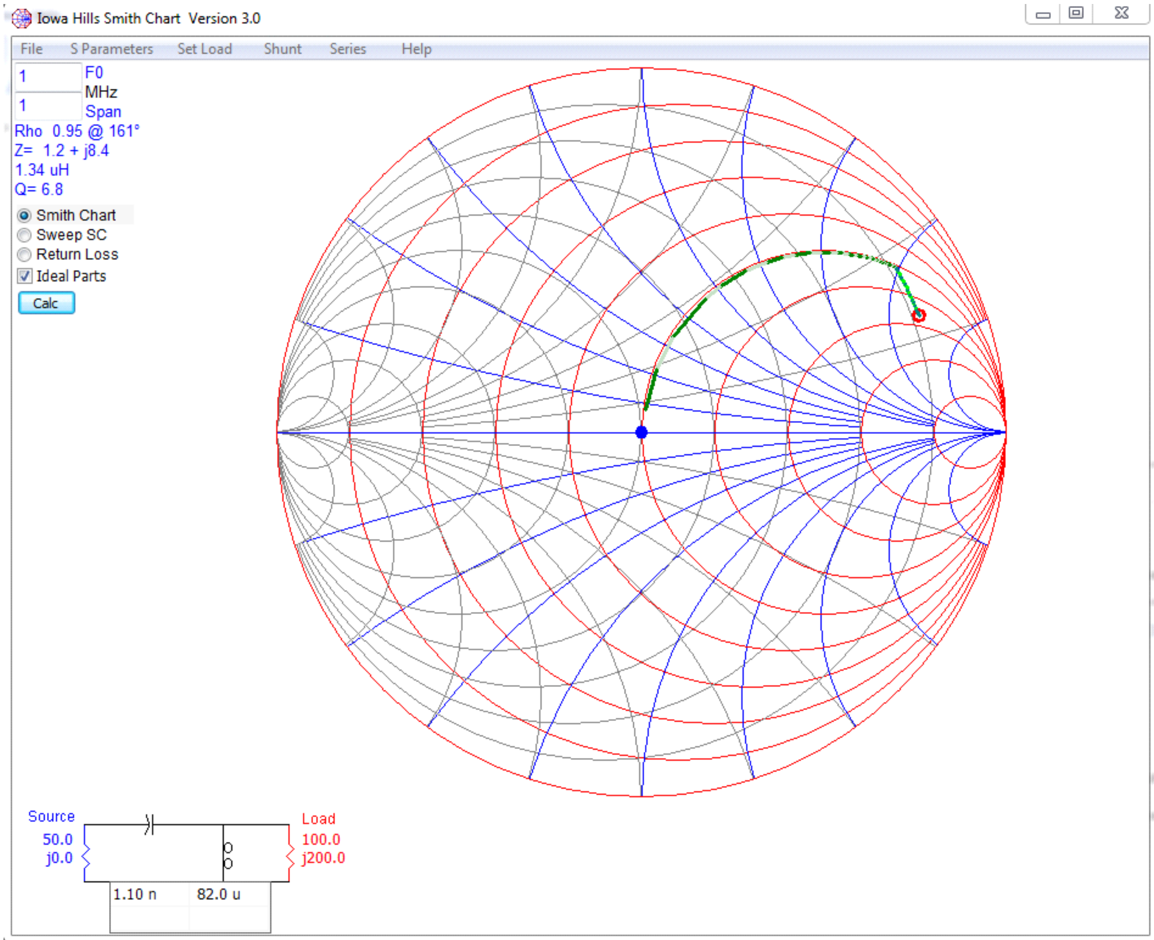
Just in case you hate math and want to let the smith chart solve another complicated issue, you can just draw a line from the point in question, through the center, and an equal distance the opposite way. Where we end up, half way around the chart, will be the values these formulas provide us. Of course, read from the admittance scales.



We see that there is a difference in susceptance of -0.002 between them, this is the susecptance of the inductor that gets us from 100+j200 to 50+j150. Again, using those formulas, or the smith chart, with the difference in suseptance, we get the reactance of the inductor we need. If we plot the that difference in admittance and shoot back across the chart as before, we see that the inductor needs to have a reactance of 500Ω.



So, to go from 100+j200, we know we need an inductor in parallel to that load with a reactance of 500Ω and a series capacitor between that and the source with a capacitive (negative) reactance of 150Ω. At 1MHz, that comes out to about 1.061nF and 79.6µH. We just made an L matching network!



Open and short coax stubs in parallel act as inductors or capacitors depending on wavelength (how long they are electrically) and transformers follow another kind of arc, but Ill let you download a smith chart app and discover those movements without me :)

Have fun!

de KE0KOY Brandon