Frequency response

Log plots

If I try to plot $f(x) = \frac{1}{1+x}$ on a linear scale for values of x between 0 and 10, I get a fairly reasonable plot. If I try to extend the range to include 100, 1000, or one million, the plot quickly stops conveying any useful information (try it sometime).
So what if I plot the log of f(x) vs. the log of x? Now I find that I can actually draw a curve over many orders of magnitude of values of x and f(x) and still get useful information about it. Better yet, all I really need to draw is straight lines.

The Bell - a unit of relative power (that's a joke - get it?)

One Bell is one order of magnitude of power gain. The Bell is a relative unit, in the sense that you need to specify a reference power (or a ratio of powers) in order to calculate how many Bells something is. For example, for sound pressure, the unit of power commonly used is the minimum power audible by the human ear, which corresponds to a pressure wave of roughly 0.00002 N/m^2. Conversational speech, at roughly 5 Bells (or 50 deci-Bells) SPL (sound pressure level) has then 5 orders of magnitude more power than the minimum audible sound. A loud concert might be somewhere around 10 Bells, or 10 orders of magnitude more sound power than the minimum audible sound! 10 orders of magnitude = 10,000,000,000. No wonder your ears ring a little after the concert!

To calculate how many Bells you have, take the log (base 10) of the ratio of two powers: $\log_{10}\left(\frac{P}{P_{ref}}\right)$. If you find that this unit is inconveniently large, you might want to multiply by 10 and talk about tenths of Bells, or deci-Bells, which is usually contracted to decibels or dB.

Probably the most common use of Bells or decibels is to represent the power gain of a linear system, in which case the ratio of powers is the output power over the input power. For a truly linear system, this ratio is independent of the magnitude of the input power.

Another common reference power level is the milli-Watt. Often power in RF circuits is written as some number of dBm, or deci-Bells relative to a milli-Watt.

Sometimes the reference power will be some mid-band power level in a given circuit, as in "the power is 3 dB down from it's DC value".

Because we usually work with voltages and currents rather than power levels, we can replace the powers in the equation above with the expression for power as a function of voltage to get

$$\log_{10}\left(\frac{P}{P_{ref}}\right) = \log_{10}\left(\frac{V^2/R}{V_{ref}^2/R}\right) = \log_{10}\left(\frac{V}{V_{ref}}\right)^2 = 2\log_{10}\left(\frac{V}{V_{ref}}\right)$$

Which is the formula to use when calculating Bells based on voltage. To calculate deci-Bells based on voltage, you need to multiply by 10 of course, which is where we get the familiar formula $20\log\frac{V}{V_{ref}}$.

Who cares about transfer functions?

Given a linear system with a transfer function $H(s)$, if the input of the system is driven with a sine wave, the output of the system will be a sinusoid with some different magnitude and phase. The wonderful thing about the transfer function is that it tells you what the magnitude and phase of that output sinusoid will be.
In particular, if $v_{in} = \sin(\omega t)$, then $v_{out} = |H(j\omega)|\sin(\omega t + \angle H(j\omega))$. This is such a simple property to use, and such a powerful thing to be able to do, that we often plot $|H(j\omega)|$ and $\angle H(j\omega)$ as a function of ω so that we can quickly look up what's going to happen to a sine wave at a given frequency. If you plot the *log* of the magnitude vs log frequency, we call these plots Bode plots. Why do we want to plot the log of magnitude and frequency, but plot phase on a linear scale?

Voltage dividers

These are easy: $\frac{R_1}{R_1 + R_2}$. The thing to remember is that this works with one or more capacitors too.

Complex numbers - magnitude, phase, and graphical calculation

Given a transfer function

$$H(s) = a_m\frac{(s-Z_1)(s-Z_2)\dots(s-Z_m)}{(s-P_1)(s-P_2)\dots(s-P_n)} = a_m\frac{\prod(s-Z_i)}{\prod(s-P_j)}$$

the magnitude and phase can be conveniently broken down in terms of the magnitude and phase of the individual terms of the form $s-x$, which has a nice physical meaning in the complex plane. $s-x$ is the vector from x to s. If we need to know the vector's magnitude, we measure its length. If we need to know its angle, we measure that relative to the positive horizontal axis.

To get the magnitude of $H(s)$, we see that it is just the product/quotient of a bunch of lengths:

$$|H(s)| = |a_m|\frac{\prod|s-Z_i|}{\prod|s-P_i|}$$

So we just need to measure all of the vectors from zeros to 's', and divide that product by the product of all of the lengths of the vectors from all of the poles to 's'.

Similarly for angles, if you represent the complex numbers as exponentials, you can quickly convince yourself (see below if you need a little help in the argument) that the angle of H(s) is just the sum/difference of a bunch of angles:

$$\angle H(s) = \sum \angle (s - Z_i) - \sum \angle (s - P_i)$$

The long and short of it is that you can draw your poles and zeros, and think about how all of these lengths and angles change as you move around in the complex plane. Usually you will be particularly interested in sliding 's' up and down the positive complex axis. (If you can't figure out why, go back up and look at 'Who cares about transfer functions?')

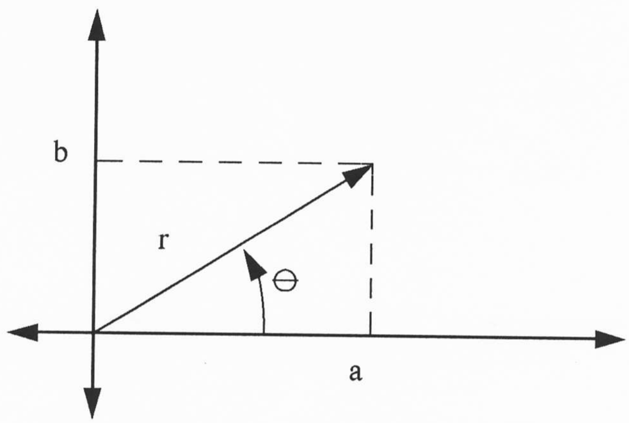
The capacitor - a frequency dependent resistor?

One very useful way to think about capacitors is just as frequency dependent resistors. If you see an RC circuit and want to know how it will perform at a given frequency, just replace the capacitor (mentally!) with a resistor of value $\frac{1}{\omega C}$. A 1pF capacitor gives you $1M\Omega$ at $\omega = 10^6$ rad/sec, and $1k\Omega$ at $\omega = 10^9$ rad/sec. (Quick quiz - what are those frequencies in Hertz?) If the capacitor is in parallel or series with a 10 kΩ resistor, something interesting is going to happen between those two frequencies.

Is this the *right* way to think about capacitors? In some cases yes - but don't forget that you're neglecting the phase shift (that's really supposed to be a $j\omega$ not an ω in the formula for the impedance of a capacitor).

Complex numbers

A complex number of the form $a + bj$ can also be represented as an exponential of the form $re^{j\theta}$. In the first form, a and b specify the position of a point in a two dimensional space in a familiar Cartesian way. In the second form, the position of the point is specified as a radius r and an angle from the horizontal axis, θ. You can use geometry to figure out what the relationship is between the pair a, b and the pair r, θ.



One advantage of the exponential form of a complex number is that the magnitude and phase can be read directly from the representation. The magnitude is r and the phase is θ.
This feature makes it convenient to work with transfer functions using exponential notation. If we represent a transfer function as

$$H(s) = a_m \frac{(s - Z_1)(s - Z_2)\dots(s - Z_m)}{(s - P_1)(s - P_2)\dots(s - P_n)} = a_m \frac{\prod (s - Z_i)}{\prod (s - P_j)}$$

and then substitute in an exponential representation for each of the $(s - Z_i)$ and $(s - P_i)$ terms, we get

$$H(s) = a_m \frac{r_{Z_1} e^{j\theta_{Z_1}} rZ_2 e^{j\theta_{Z_2}} \dots r_{Z_m} e^{j\theta_{Z_m}}}{r_{P_1} e^{j\theta_{P_1}} r_{P_2} e^{j\theta_{P_2}} \dots r_{P_n} e^{j\theta_{P_n}}} = a_m \frac{\prod r_{Z_i} e^{j\theta_{Z_i}}}{\prod r_{P_i} e^{j\theta_{P_i}}}$$

now remember that products of exponentials turn into an exponential of a sum, and we see that

$$H(s) = a_m \frac{\prod r_{Z_i}}{\prod r_{P_i}} e^{j\left(\sum \theta_{Z_i} - \sum \theta_{P_i}\right)}$$

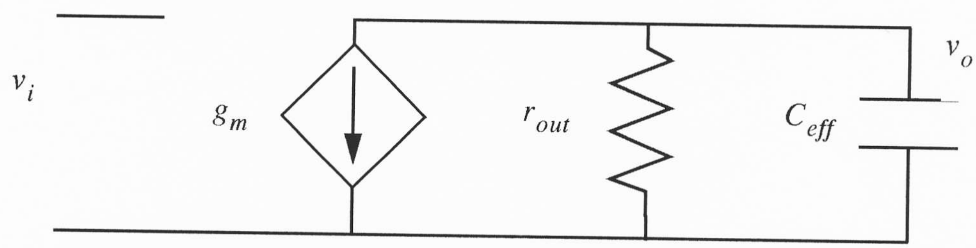
The magnitude of the transfer function at a given point s is just the product of the distances from the zeros to s divided by the distances to all of the poles (and don't forget about a_m). The phase is the sum of all of the phases of vectors drawn from zeros to s minus the phases of vectors drawn from poles to s.

Frequency response, part II

Most amplifiers can be broken down into a series of stages. Each stage has its own relationship between gain and frequency, which we call the frequency response of the stage. The overall gain of the amplifier as a function of frequency is just the product of the frequency dependent gain of each stage. In many cases, the stages can be modeled independent of each other. We will see that in some cases (notably Miller compensation) the gain of one stage has a strong effect on the fequency response of another stage.

Single stage gain and frequency response

Most gain stages can be modeled by a voltage controlled current source or transconductance (g_m), an output resistance (r_{out}), and an output capacitance (C_{eff}).



The low frequency gain of the amplifier is independent of the capacitor value, and is just $-g_m r_{out}$. At some higher frequency, the capacitor will begin to "steal" a signifcant fraction of the current coming out of the transconductor. This will mean that less current is flowing in the resistor, hence the output voltage will decrease, and the gain of the system will drop accordingly. To find the frequency at which the capacitor has the same *magnitude* impedance as the resistor:

$$\left|\frac{1}{j\omega C_{eff}}\right| = \frac{1}{\omega C_{eff}} = r_{out}$$

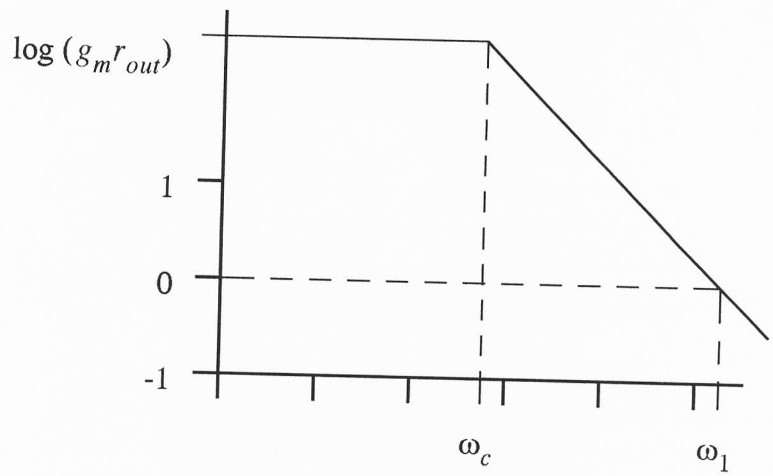
Now solve for ω to get $\omega_c = \dfrac{1}{r_{out} C_{eff}}$.

If we were to replace the capacitor with a resistor of equal magnitude impedance, the voltage gain of the amplifier would decrease by a factor of two. Because the capacitor has a complex impedance, the actual decrease in gain at frequency ω_c is only $\sqrt{2}$. (Why? You should be able to show this both graphically and algebraically).

For frequencies higher than ω_c, most of the current from the transconductor goes into the capacitor, and the resistor plays an increasingly small part in the circuit, the load impedance looks very much like $\dfrac{1}{j\omega C_{eff}}$, and the output voltage becomes $v_o = \dfrac{-g_m}{j\omega C_{eff}} v_i$

Another frequency of interest is the frequency at which the amplifier gives a gain of 1. The unity gain frequency ω_1 is the frequency at which amplifier stops giving voltage and *gain*, and begins to attenuate the signal. This occurs when $\omega = \frac{g_m}{C_{eff}}$. These three expressions for the DC gain, corner frequency, and unity gain frequency are very general - they apply to MOS amplifiers, bipolar and vacuum tube amplifiers, even hydraulic systems. They are based on the assumption that the stage is linear, which is always true for "small" signals - where "small" essentially means signals that are small enough to make the amplifier look linear (so we've got some circular logic here).

The relationships can be summarized graphically using a plot of the log of the magnitude of the gain versus the log of the input frequency.

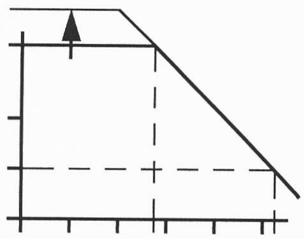


Let's plug in some numbers: if we have a g_m of 1 mA/V^2, an output resistance of one Megaohm, and a 1 pF capacitor, we get a DC gain of -1000, a corner frequency of one million radians per second, or around 160kHz, and a unity gain frequency of one billion rad/sec, or around 160 MHz.

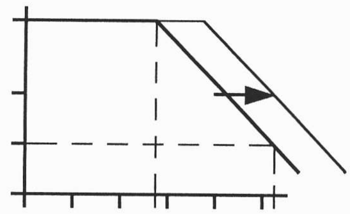
It is worth remembering how the three components affect the shape of the transfer function as well. Increasing the output resistance r_{out} increases the low frequency gain, but decreases the corner frequency, so the high frequency performance remains unchanged. Increasing the effective capacitance C_{eff} decreases both the corner frequency and the unity gain frequency, but has no effect on low frequency gain. Increasing the transconductance g_m increases both the low frequency gain, and the unity gain frequency, but does not affect the corner frequency. Notice that the unity gain frequency is equal to the corner frequency times the DC gain, $\omega_1 = (g_m r_{out})\,\omega_c$. In general, the gain at frequencies above the corner frequency is given by

$$A(\omega) = \frac{g_m r_{out}}{1 + \omega/\omega_c} \approx \frac{g_m r_{out} \omega_c}{\omega} = \frac{A_o \omega_c}{\omega} = \frac{GB}{\omega}$$

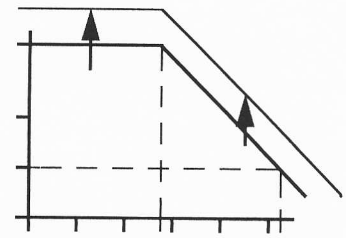
where GB is the gain-bandwidth product of the amplifier.



Increasing r_{out}



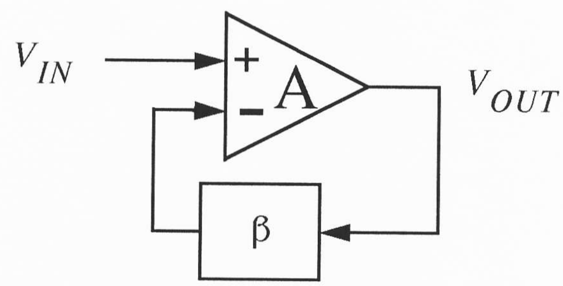
Decreasing C_{eff}



Increasing g_m

Feedback and stability

Feedback improves linearity, reduces output impedance, reduces input offset, increases bandwidth (sort of), allows gain to be set accurately, and all sorts of other wonderful stuff. Unfortunately, it also introduces the possibility of instability



With a little algebra you can show that the gain from input to output in this figure is

$$\frac{V_{OUT}}{V_{IN}} = \frac{A}{1 + A\beta}$$

All of the quantities in this equation can be frequency dependent, and the relationship still holds, and is often written $G(s) = \frac{A(s)}{1 + A(s)\beta}$. If A is very large (or A(s) has a large magnitude), you can approximate the gain as $\frac{1}{\beta}$. To get a large gain, you need a small β, which may seem counterintuitive, but one way to look at it is that the amplifier needs to set V_{OUT} so that V_+ and V_- are (approximately) equal. If β is small, it takes more of V_{OUT} to produce the desired effect at V_-, resulting in large gain.

From the equation for the gain, it is clear that something interesting will happen if $A\beta = -1$. The equation indicates that the amplifier will provide infinite gain in this case. The unfortunate reality of this is that the amplifier will oscillate, become unstable, ring, buzz, pop, and in some cases, melt.. It will certainly not provide you with the nice clean signals that you would like to see.

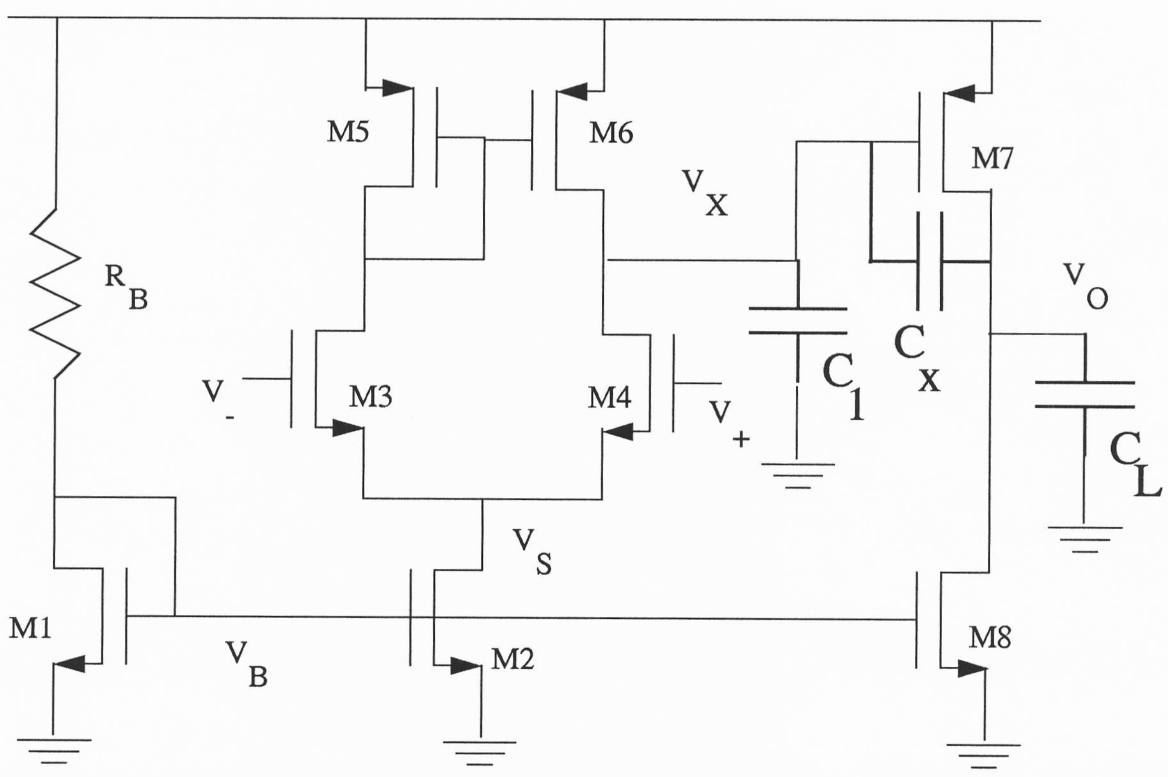
To avoid this unstable state of affairs, let's first see how it can happen.

A(s) equal to negative one implies that the magnitude of A is one, and the phase of A is 180 degrees. Is this the only combination to worry about? What if the magnitude of A is 2 when the phase is 180 degrees? The answer, as you might expect, is that if the magnitude A is greater than or equal to 1 when the phase of A is 180 degrees, the system will be unstable. Let's call the fequency at which the amplifier gives a phase shift of 180 degrees ω_{180}. From the figure we can see that a signal arriving at the positive input of the amplifier with frequency ω_{180} will be amplified in magnitude by $|A(j\omega_{180})|$, and effectively multiplied by -1 (due to the 180 degrees of phase shift) and sent into the negative input of the op-amp. The two minus signs will cancel, and the process will repeat. Each time around the loop the signal will be multiplied by $|A(j\omega_{180})|$, and inverted twice. Clearly, if $|A(j\omega_{180})|$ is greater than 1, the output signal will increase rapidly with time until something non-linear happens (like hitting the supply rails of the amplifier).

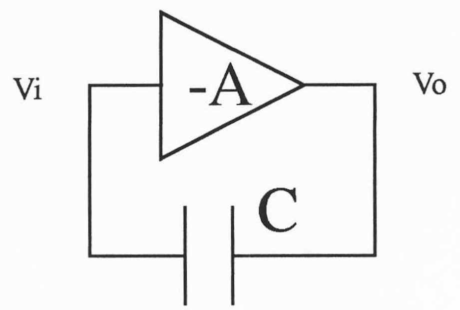
In summary, to design a stable amplifier, you need to make sure that you design the frequency response so that the magnitude of the transfer function is less than one by the time the phase has reached 180.

Miller compensation

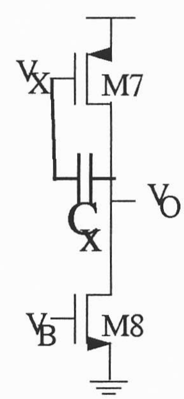
The two stage CMOS op-amp below has a primary pole associated with each stage. Both poles



are of the form, $\omega_p = \frac{1}{r_{out}C_{eff}}$, but in both cases something interesting happens. The capacitance seen at the output of the differential amplifier is much larger than the apparent values of the capacitors $C_x + C_1$. This is due to the Miller multiplication of the compensation capacitor C_x. When a capacitor is placed across a gain stage it's effective capacitance changes. Consider the figure below, in which $V_o = -AV_i$. The charge on the capacitor is given by



$Q = C(V_i - V_o)$, but since we know what V_o is once we know V_i, we can just write $Q = C(V_i - -AV_i) = C(1+A)V_i = C_{eff}V_i$ where $C_{eff} = (1+A)C$. So the effective capacitance is equal to the actual capacitor value times the gain of the stage it's across.
In the figure below, this means that $C_{eff1} = g_{m7}r_{out2}C_x$ and that the pole of the differential



amplifier is at $\omega_{p1} = \frac{1}{r_{out1}C_{eff1}} = \frac{1}{r_{out1}(g_{m7}r_{out2})C_x}$.
If this were the only pole of the op-amp, the corresponding unity gain would be

$\omega_1 = A_o\omega_{p1} = \frac{(g_{m3}r_{out1})(g_{m7}r_{out2})}{r_{out1}(g_{m7}r_{out2})C_x} = \frac{g_{m3}}{C_x}$. As we will see, it turns out that this generally is the unity gain frequency of a well designed Miller compensated op-amp.
The pole frequency of the gain stage is also affected by the compensation capacitor, but not because of Miller multiplication. There is no Miller multiplication since the gain looking from the output of the stage back toward the input is small. At higher frequencies, however, the compensation capacitor does effectively couple the output signal back to the gate of transistor M7,

resulting in an effective output resistance of $r_{out2} = \frac{1}{g_{m7}}$ for frequencies above ω_{p1}.

The capacitance seen at the output node is the sum of the load capacitance C_L and the series combination of C_x and C_1. If $C_x \gg C_1$, then the output capacitance is just $C_{out2} = C_L + C_1$. Generally the load capacitance will be higher than the internal capacitance C_1, so the second pole frequency is approximately $\omega_{p2} = \frac{g_{m7}}{C_L}$.

There is one additional effect of the compensation capacitor. In addition to supplying a high-frequency coupling from the output to the gate of M7, it also allows signals to travel from V_x to V_o without going through the inverting path through M7. This results in a *right* half plane zero located at $\omega_z = \frac{g_{m7}}{C_x}$.

Noise

There are many sources of noise in analog circuits. There is external noise, which may come from electrostatic or magnetic interference (you've all seen 60 Hz on an oscilloscope). This type of noise can be minimized by using differential signals, running signals in twisted pair wires, or co-ax cable. Unfortunately for the analog circuit designer, there is a more fundamental type of noise in electrical circuits - thermal noise.

All resistors generate noise. Electrons in a resistor are always bouncing around - they have thermal energy. In the absence of an externally applied electric field, their motions are random. At any given time, some of them will be poking their heads out one end of the resistor or the other. This generates a voltage which is random, and this voltage is what we call the thermal noise voltage of the resistor. Nyquist (I think) showed that the amount of power in this noise voltage was 4kT for every Hz of bandwidth. Since the amount of noise power is the same all across the spectrum (4kT in each Hz) this is often called "white noise" (Q: what does "white" have to do with evenly distributed power over frequency?). kT is a fairly small amount of energy (about 4×10^{-21} Joules at room temperature), so we're not talking about much power here. If we look at all of the thermal energy from DC to a gigaHertz, we only get about four picoWatts. Equating the power that Nyquist tells us about with the formula for power in a resistor, we come up with $\frac{V^2}{R} = 4kT\Delta f$ and if we want to know what the noise voltage will be, we can solve for V to get

$$V = \sqrt{4kTR\Delta f}$$

For a $1\text{k}\Omega$ resistor, this works out to about 4 nanoVolts times the square root of the bandwidth of interest, or 4 $nV/\sqrt{Hz}$. So in a 1 MHz bandwidth, we could expect to see on average about 4 microVolts of noise across the resistor.

Active devices (like MOSFETs and BJTs) generate noise because they have internal resistances. Ideal passive components such as inductors and capacitors do not generate noise, but real devices will generally have some parasitic resistance which will generate noise.

Questions

1. Since the gain of a multistage amplifier is the *product* of the gains of the individual stages, why is it that I can talk about *adding* the magnitude and phase of the transfer functions? (you may want to use both graphical or analytical explanations to answer this question)
2. Derive the formula for the output impedance of a cascode stage ($r_{out} \approx g_m r_o^2$).
3. Derive the formula for the impedance at the source of the cascode transistor ($r_{out} \approx \frac{1}{g_m}$) in a cascode amplifier.
4. If I invent a new type of MOS transistor with a drain current given by $I_D(V_{DS}, V_{GS}) = K(V_{GS}^2 + \beta V_{GS}^3)\left(1 + \lambda\sqrt{V_{DS}}\right)$, derive expressions for the transconductance and output impedance (g_m and r_o).
5. Design a folded cascode CMOS amplifier with a DC gain of 10,000 and a unity gain frequency of 200 MHz. Assume that the load will be capacitive, with no more than 10pF of capacitance. How would your design change if you were told to minimize power dissipation at all costs? What would the minimum power dissipation be? What if you were told to minimize die area? (Assume that die area is directly related to the sum of W times L for all devices). What if you were told to design for a power supply of less than 5 volts. How low could you go?
6. Design a two stage Miller compensated CMOS amplifier which will be able to track input signals with better than 0.1% accuracy when used in unity gain feedback with a 1k resistive load. How would your design change under the same constraints of power, die area, and supply voltage above?
7. What is the fastest op-amp you can design which operates from a 5 volt supply, allows input and output voltage swings to within 1 volt of each rail, and dissipates less than $100\mu W$ of power?
8. Compare the performance of a CMOS two stage Miller compensated op-amp drawing 1mA in the input stage and 5 mA in the output stage with the performance of a similar bipolar operating with the same bias currents. In particular, if we assume a V_{GT} of 250mV for the CMOS devices, how do the transconductances and output resistances of the individual devices compare? What size capacitor is needed in each case to make the amplifiers unity gain stable? What is the DC gain? Unity gain bandwidth? How do these answers change if we increase of decrease V_{GT} for the CMOS devices?
9. Explain why unity gain feedback is the worst form of passive feedback ("worst" from a stability perspective). For an amplifier with $g_{m3} = g_{m7}$ and $C_x = 10C_L$, what is the minimum feedback gain (or maximum feedback factor β) for which the system will have 90 degrees of phase margin?