Chapter 8: Noise and Modulation Techniques

One of the unfortunate facts about electronics is that there is always electrical noise to interfere with out measurements. If the signals we are trying to measure are large, noise is not usually a problem. However, if the signals are small or we are trying to measure them under "difficult" circumstances, noise is often a problem. The noise interferes with our measurements and we have to separate the signal from the noise.

A: Interference – External Noise

The noise may come from many sources. One source is interference. This is usually the term for special external sources of electrical signals like the power lines or other outside sources of electrical signals. They are external to your measurement system. The 60Hz electrical signals from power lines are almost ubiquitous in our society. Just grab an oscilloscope probe with the hot lead in your right hand and the ground lead in your left hand and look at the trace on the scope. If you want to measure someone's EKG, you have to get rid of this interference. Transformer "leakage" will produce 180Hz signals. If you are trying to tune a radio station under high voltage power lines, the corona discharge may interfere with your station's signal.

There are several techniques used to try to eliminate interference.

1. If the interference is at one frequency, you may be able to remove most of it with a notch filter, providing the filter does not affect your signal too much.
2. If your signal is at one frequency, or in a narrow range of frequencies, you can use a narrow band pass filter to amplify your signal and not the interference. This is how a radio receiver eliminates the signals from other stations.
3. Another technique is shielding or screening. If the signal you are measuring and your measurement system can be physically isolated from the interference, you can surround your system with conducting material and the conductor will block external electric fields from your system, i.e. it shields your system from the interference or screens out the interference. It is common to use wire mesh, screen, as the conducting material, but a solid conductor can be better. It is not unusual for an entire room to have its walls lined with copper mesh to shield the measurement system in the room. (It is a little harder to screen out low frequency magnetic fields, because they can penetrate some distance into a conductor.)

I should mention that the power mains can also carry other types of interference into your system via your power supply. As a result most instruments now filter out high frequency noise from the power leads coming into their power supply. (Computers also have filters to prevent interference from the computer from getting onto the power lines.) Motors turning on and off produce noise. Switches can produce noise if there is arcing. Motors using brushes can also produce noise.

B: Internal Noise

Unfortunately some noise sources come from within your system. One common source is thermal noise. A resistor generates noise over a broad range of frequencies. This is called Johnson noise. The expression for the rms voltage noise in a given frequency range is

$$V_{rms} = \sqrt{4k_B TR\Delta f} \qquad 8.1$$

where T is the absolute temperature in K, k_B is Boltzmann's constant, $1.38\text{x}10^{-23}$J/K, R is the resistance and Δf is the frequency range in Hz.

If the bandwidth is 1000Hz, T = 300K and R = $10^3\Omega$, then the noise voltage is V_{rms} = 0.13μV. If you are measuring voltages in the mV range, it isn't bad, but if the signal is a couple of μV's it may become noticeable. However, if you increase the resistance to $10^6\Omega$ and the bandwidth to 10^5Hz, the noise voltage becomes 41μV. This could be a problem if you are measuring a signal that is only a few tenths of a mV.

There are a three of ways to try to reduce this problem.

1. Reduce the resistance. Decreasing R by 100 reduces the noise voltage by 10.
2. Reduce the bandwidth. Reducing the bandwidth by 100 through filtering will reduce the noise voltage by 10 again.
3. Reduce the temperature of the device that contains the critical resistance.

This is only a problem for small signals and comes from the real part of the impedance, or the resistance. Also, this is the minimum noise voltage due to thermal fluctuations in the resistor; most resistors exhibit more noise than this, especially carbon film resistors. Metal film resistors are better and metal wire-wound resistors are usually best.

A second source of internal noise comes from shot noise. This results because the current consists of discrete charges, typically electrons. This produces small current fluctuations and therefore in current noise. The rms noise current is given by

$$i_n = \sqrt{2eI_o\Delta f} \qquad 8.2$$

where e is the fundamental charge of 1.6×10^{-19}C, I_o is the average current and Δf is the bandwidth in Hz.. (This assumes the motion of the charge carriers is uncorrelated. In a metal wire, the motion of the carriers may have some correlation so this expression (eqn 8.2) does not always apply. However, the current flowing across the junction in a diode would exhibit this effect.) This is only likely to be a problem if the currents are very small.

If the average current is 1mA and Δf = 1000Hz, then $I_n = 5.7\times10^{-10}$A, or 0.57nA. If this current flows through a 1kΩ resistor, the resulting noise voltage will be 0.57μV.

Another source of internal noise is flicker or 1/f noise. This noise gets worse as you lower the frequency. You usually try to eliminate this by shifting the measurement to higher frequencies.

C: Modulation

As an example, consider the problem of trying to measure the intensity of an LED in a room where you cannot eliminate outside light. How do you separate the light from the LED from the other sources of light? One common technique is to modulate the light from the LED. You drive the LED with a function generator at 1000Hz and only look for light intensities at 1000Hz, using some sort of filtering. Another technique is to chop the light from the LED with a chopper, a fan-like device that alternately blocks and passes the light as the blades rotate at a fixed rate. Basically you move the measurement frequency to get away from interference and 1/f noise problems.

Modulation is a little more general than this. Typically you have some slowly varying function of time, f(t). To modulate it you **MULTIPLY** it by a sinusoidal function of time, say $\cos(\omega_m t)$, where ω_m is the modulation frequency, which is typically a high frequency. The modulated signal looks like

$$f(t)\bullet\cos(\omega_m t) \qquad 8.3$$

Here you want any frequency components in f(t) to be much less than ω_m. Then the frequency components in the modulated signal, f(t)•cos(ω_mt) will be close to ω_m. If the highest frequency component in f(t) is ω_s, the modulated signal will lie in a frequency range of $\omega_m \pm \omega_s$, so $\Delta\omega = \pm \omega_s$ around ω_m. One then amplifies the modulated signal with some sort of band pass circuit centered on ω_m with a Q such that signals within ω_s of ω_m will not be attenuated very much. If $\omega_s = 0.1\omega_m$, then you would want $Q \le 5$. If $\omega_s = 0.01\omega_m$, then you would want $Q \le 50$.

Now the problem is to demodulate the signal and recover the original f(t). To do that I would multiply the modulated signal by cos(ω_mt) again. Now the signal becomes

$$f(t)\times\cos(\omega_m t)\times\cos(\omega_m t)=f(t)\times\frac{1+\cos(2\omega_m t)}{2} \qquad 8.4$$

Since $\omega_m >> \omega_s$, one runs this through a low pass filter just above ω_s which greatly reduces the term f(t)cos($2\omega_m$t) which has frequency components around $2\omega_m$. This leaves f(t)/2 as the remaining signal. This is basically how a lock-in amplifier works. The only problem is that many lock-in amplifiers multiply by a square wave instead of a sine wave. It is easier to multiply by ±1 than by a sine wave. Some newer and more expensive lock-in's multiply by sine waves.

Another problem can arise if the signal is phase shifted in the amplification process, then when you multiply by cos(ω_mt) it really is shifted in phase from the incoming signal and looks like cos($\omega_m t + \theta$). This means that after low pass filtering your result is

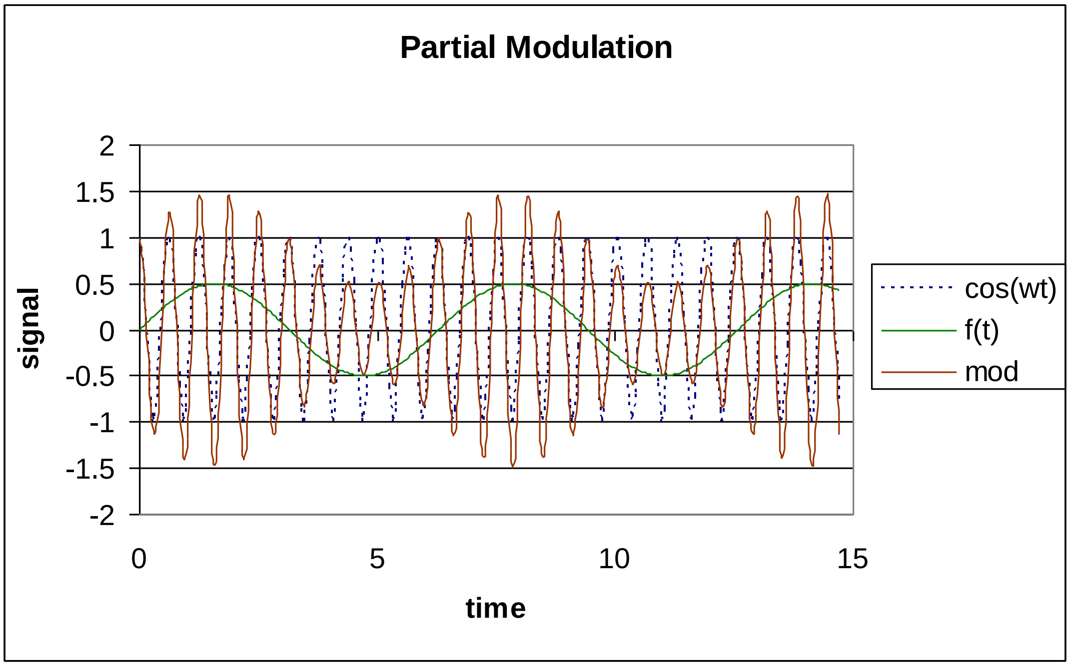
$$\frac{f(t)\times\cos(\theta)}{2} \qquad 8.5$$

Lock-in amplifiers allow you to vary the phase of the multiplying signal to make $\theta=0$ in the above expression to get the maximum signal out, and measure how much phase shift has occurred. (Sometimes that is useful information, because the phase shift can also be due to the experiment and be an important part of the measurement.) Again newer and nicer lock-in amplifiers will give you the amplitude and phase shift at the same time. (This is essentially the transfer function of the experiment and equipment together.)

AM radio uses the idea of modulation this way, but it uses a cleaver trick to make demodulation easier. In the above modulation scheme, you can think of f(t) as modulating cos(ω_mt). Then the cos(ω_mt) is often called the carrier wave. Instead of multiplying f(t) times cos(ω_mt), one does the following. If |f(t)| < 1 then one multiplies

$$(1+f(t))\times\cos(\omega_m t) \qquad 8.6$$

This is called partial modulation and the term (1 + f(t)) >0 all the time. The modulated signal looks like that below. Here I've used f(t) = 0.5 sin(ωt) at 1/10 the frequency of the modulation frequency. Normally I would try to make the frequencies farther apart, but this shows the effect.



The interesting thing is that if one takes the positive part of the modulated signal one recovers the shape of f(t). To actually recover f(t) one must low pass filter the positive of the modulated signal with a filter whose cut off frequency is greater than the maximum frequency in f(t) but much smaller than ω_m. This will reproduce f(t). To take the positive part one can use a diode, even though it has that annoying 0.6V offset. The positive part of the modulated signal is shown below along with the low pass filtered version. (I've assumed an "ideal" diode, i.e. no 0.6V drop.) Because the two frequencies are close together, it is hard to remove all of the carrier at 10 times the frequency of f(t) without attenuating some of the signal, nevertheless you can see that this gives back the form of f(t).