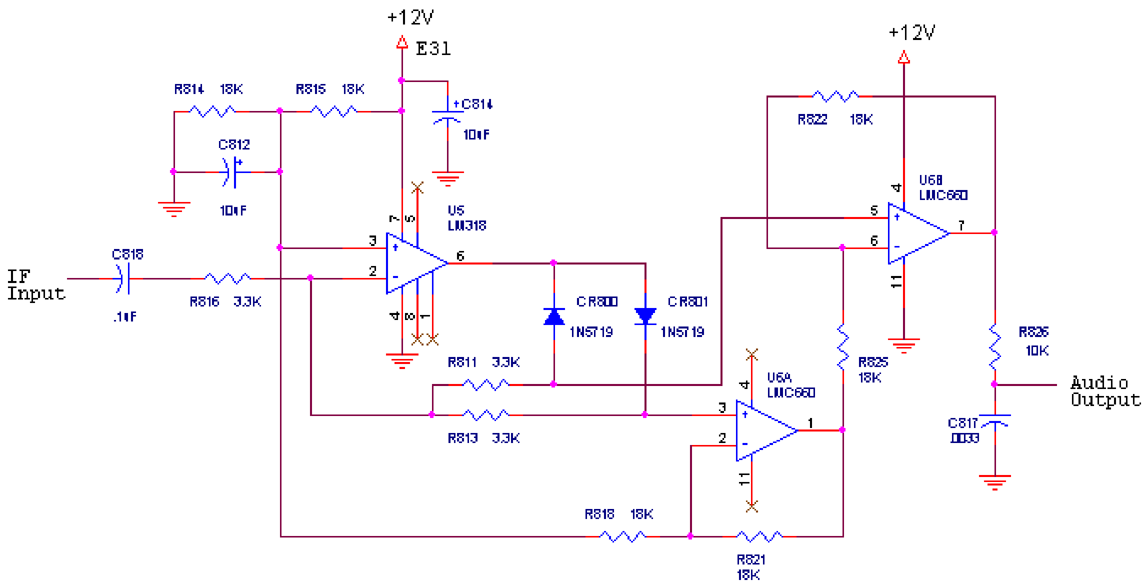
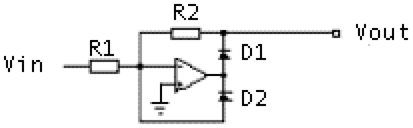
An improved, precision, full wave AM detector. By Rob Schenck, K2CU K2CU@arrl.net

Detector Schematic



Theory of Operation

This circuit is a variation of the classic "Precision Half Wave Rectifier" as described in many op-amp application notes as shown below:



Its operation is based upon basic principles of operational amplifier circuits. An ideal op-amp has extremely high input impedance, extremely low output impedance, and extremely high gain. Note the operative word "extremely". The non inverting input (+) of the op-amp is connected to ground. Feedback is provided by the two paths R2+D1 and D2. The op-amp will drive its output pin to a voltage such that the inverting input (-) is at ground voltage. There will be a slight variation (a few millivolts) from 0 volts as a result of what is known as input offset voltage, a property of real op-amps. The inverting input is often called the summing junction, as the sum of the currents in that node will be zero. Since virtually no current flows into the inverting input (-) of the op-amp itself, any current coming in through the input resistor R1 will be met by an equal and opposite current coming through either of the two paths of R2+D1 or D2.

The I.F. input signal will produce an input current of:

Iin = Vin/R1

During the negative half of the input AC waveform, the op-amp output will drive positive such that the current through R2+D1 will be equal and opposite the input current, or:

$Iin = Vin/R1 = -I_{D1} = -I_{R2}$

The voltage developed across R2, and hence the output voltage as the left side of R2 is at ground potential, will be:

$V = Vout = -I_{R2} * R2 = -Vin/R1 * R2$

By setting R1 = R2, the output voltage will be;

$Vout = Vin$

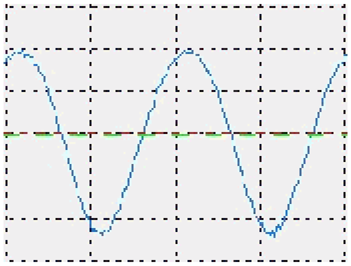
Similarly, the positive going input cycle produces a feedback current through D2. Any non linear voltages developed across the diodes is not relevant as the output is derived from the feedback resistor only.

A full wave detector was desired for a 50 KHz IF application so that the ripple frequency out of the detector would be 100 KHz, and more easily filtered. In the full wave detector version of the circuit, a second feedback resistor is added to the D2 diode path to produce a voltage of the other half of the input waveform. A virtual ground at +6 Volts was created with R814/R815/C812, so that the circuit could operate from a single supply. Again, any non linear voltages developed across the diodes are not relevant. The two rectified half waves are combined in the next two op-amp stages. The two outputs of the rectifier on R811 and R813, are connected directly to the non-inverting inputs of the op-amp stages so that there will be minimal loading to these outputs. U6A provides a gain of two to the signal voltage on R813. U6B provides a gain of two to the voltage on R811, and inverts and adds the output of U6A. The result is a full wave rectified signal at the output of U6B. Simple low pass filtering will remove the carrier energy which is at twice the ripple frequency.

The ability of this circuit to accurately track the input waveform is dependant on the speed of the op-amp and the switching speed of the diodes. For the circuit to track the input, the op-amp must be considerably faster than the input frequency. The LM318 with its 15 MHz bandwidth was selected for this reason and because it is readily available and stable. This circuit was built up and set up for comparison with the stock envelope detector in a Drake R4-C with a 50 KHz I.F. A HP8640 signal generator was used to provide a 1Khz modulated carrier at 3.885 MHz which was input to the antenna connector. The input level was set to S9 (30 uV). The detectors were compared at modulation levels of 80% and 99+ %. A Link Inc. , PC based oscilloscope was used to record and display the resulting waveforms.

R4-C Stock Detector at 99+%

R4-C Stock Detector at 80%



The distortion is visible at the crest of the wave, where the input signal is at the modulation trough in this inverting detector. The distortion at 80% is less visible than at 99+%, but still audible.

Applying the 99+% modulated signal and using the precision rectifier results in the following waveform:

Precision Detector 99+% Modulation

This waveform displays virtually no distortion not only visually, but also audibly. A pure tone is heard coming from the speaker. Actual on air comparative listening tests of 75 meter AM signals at levels from just over the noise to well over S9 were quite conclusive. Low level signals could be more easily heard through the noise, and high level signals we noticeably clearer.

Comments: Purists may want to use LM318 op-amps for all three stages, but I found no significant improvement when doing same at 50 KHz. I wanted to pick parts that were readily available and in particular in DIP packages so that any builder would be able to build up the circuit on a perf board, preferably the type with a ground plane on one side. This circuit is so stable at 50 KHz, that it was actually built and tested using a prototyping board with no ground plane at all. Operation at 455 KHz will definitely require a ground plane in the construction. At 455 KHz IF frequency and higher, the combining circuitry requires the use of three LM318's or equivalent.

The Math Behind The Detection Process.

There have been several articles written discussing the production of harmonic distortion by diode envelope detectors. Almost all of these start with the exponential diode equation and then examine the cross modulation products derived from the series expansion of the exponential function.

$e^X = 1 + X + ½ X^2 + 1/6X^3 + 1/24 X^4 + \ldots\ldots$

The simplistic AM modulated signal is:

$V(t) = \cos(Wc)\, t + m/2 \cos(Wc - Wm)\, t + m/2 \cos(Wc + Wm)\, t$, where m is modulation index

= Carrier + LSB + USB

The term of significance in the e^X expansion series is the square term as it provides the cross product of the carrier with each sideband, as in synchronous detection. Unfortunately, it also provides the cross product of the two sidebands which produces second harmonic distortion of the modulating signal.

Consider : $(a + b + c)^2 = a^2 + ab + ac + b^2 + ba + bc + c^2 + ca + cb = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$

Where;

$\cos(Wc)\, t = a$

$m/2 \cos(Wc - Wm)\, t = b$

$m/2 \cos(Wc + Wm)\, t = c$

The three squared terms produce signals at twice the carrier frequency by the trigonometric double angle formulas:

$\cos^2(W) = ½ (\cos (2W) + 1)$

The 2ab and 2ac terms multiply each sideband with the carrirer:

$2 (\cos(Wc)t) (m/2\cos(Wc-Wm)t) = m/2 [\cos (2Wc –Wm)t + \cos (Wm)t]$

$2 (\cos(Wc)t) (m/2 \cos(Wc+Wm)t) = m/2 [\cos (2Wc +Wm)t + \cos (Wm)t]$

Which results in double sideband AM at twice the carrier (2Wc) plus the desired, demodulated output $m \cos (Wm)t$.

These detectors are often referred to as square law detectors for this reason.

The problem is the 2bc term which is the cross modulation of the two sidebands themselves. It results in:

$2bc = 2 (m/2 \cos(Wc-Wm)t) (m/2\cos(Wc+Wm)t)$

$= m/4 [\cos(2Wc)t + \cos (2Wm)t]$

The $m/4 \cos (2Wm)t$ term represents some 25% second harmonic distortion, which is much more than what is usually experienced in the real world.

In the real world, diode detectors work into a load resistor which often has a parallel "filter" capacitor. If the diode were an ideal device, it would only allow current to flow in the load resistor during precisely one half of the input waveform. The resultant voltage developed across the resistor would then be precisely the half wave rectified voltage of the input signal. The reality is that the current/voltage relation in a diode is a non linear, exponential function as has been discussed. This means that the output voltage on that resistor would be the half waveform of the input signal minus the voltage across the diode.

Vout = Vin - Vdiode

It is the voltage across the diode that is the non-linear portion of the output signal, which becomes dominant at low input levels, such as the modulation trough of an AM signal. There are techniques that have been employed to reduce the diode component of the output waveform. One is to drive the detector into a high impedance load resistor. With the current very low, the voltage produced on the diode will be low. This approach is limited by the generally ignored reverse current of the diode. At high impedance levels, the rectifying function of the diode becomes compromised by the reverse leakage current, resulting in other performance problems.

The precision rectifier uses the features of op-amp feedback design to eliminate the diode error from the rectification process, resulting in a "perfect" rectifier. Well, why does a perfect rectifier work as a detector anyway? Most diode detector discussions talk about the square law function as discussed above to explain the demodulation process. The received carrier multiplies against the received sidebands. This is where the error comes in the thinking that it also should produce an objectionable level of second harmonic distortion. Let’s go back to the pure half wave rectified signal that the precision detector produces and the diode with the load resistor tries to emulate. What is the process that takes an input signal at some frequency, Wc, and outputs only the positive half of the waveform. It is just as if the signal were multiplied by one during the positive half cycle, and zero during the negative half cycle. The type of waveform to do this would be a square wave, with a values of one or zero, and at the same frequency and phase as the incoming carrier. The mathematical representation of a square wave is what is known as a Fourier Series expansion. For our sequence of one, zero, one, zero... at a frequency of Wc the function is:

$f(X) = ½ + 2/Pi \cos Wc\, t + 2/3Pi \cos 3Wc\, t + 2/5Pi \cos 5Wc\, t + ...$

When this is multiplied times the incoming signal, we get a result that is the half wave rectified waveform. Our interest is in the fundamental term, $\cos Wc\, t$. The higher frequency harmonic terms and the DC term will produce results that are all at or above the incoming IF frequency. So the product term of interest becomes:

$$\text{Vout}(t) = \{A \cos(Wc)\, t + \tfrac{1}{2} AB \cos(Wc - Wm)\, t + \tfrac{1}{2} AB \cos(Wc + Wm)\, t\} * 2/\text{Pi} \cos Wc\, t$$

The demodulated sideband terms are:

$$\text{Vout}(ty) = AB/2\text{Pi} * \cos(Wm)\, t + AB/2\text{Pi} * \cos(Wm)\, t \text{ , or}$$

$$\text{Vout}(ty) = AB/\text{Pi} * \cos(Wm)\, t$$

There are no intermodulation terms to contend with. In fact, functionaly, this is identical to synchronous detection. The negative half wave signal is derived in a similar fashion. The two signals, when combined in the summing circuit produce an output with twice the ripple frequency to be filtered by an output low pass filter. This is of value when the IF frequency is only 50 KHz, as in the Drake R4-C and others. For receivers with a 455 KHz or 500 KHz IF, replace the two LCMC660 op amps with a pair of LM318's.