Circuit analysis of the diode ring modulator

July 22, 2015

J.L.

A ring modulator is just a distinct name given for the modulator constructed out of four diodes in a circular formation. The circuit diagram of the ring modulator is given in Figure 1. In official RF theory terms, the ring modulator

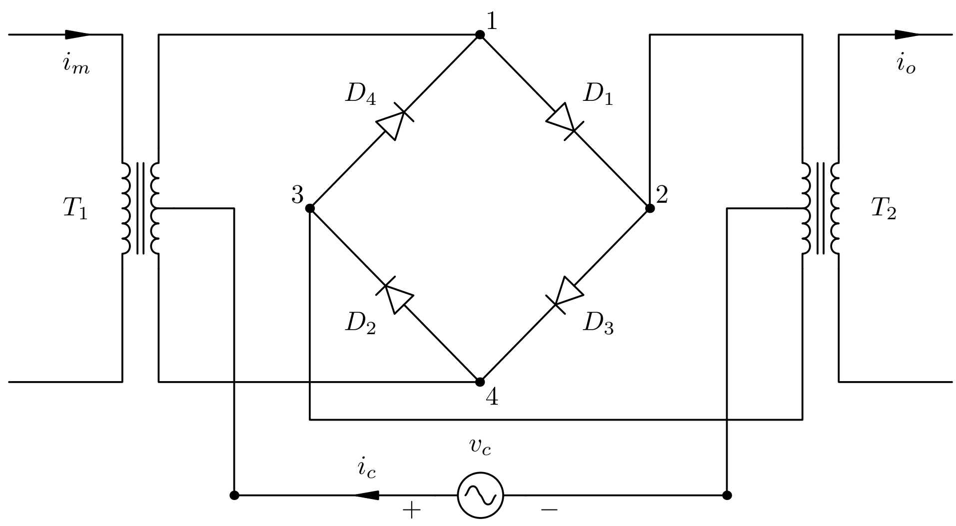

Figure 1: A diode ring modulator circuit diagram

implements a balanced modulator, which approximately generates a double side-band modulated signal with a suppressed carrier.

The circuit uses centre-tapped transformers, which might frighten away the casual DIY builder, like myself for instance. The transformers are there to perform a push-pull type of action, which helps to cancel away most of the excess frequency products. The main actuator of the circuit, the diode, is a nonlinear component, which affects current flow in exponential manner when forward-biased using a proper control voltage. On the other hand, any nonlinear function can be represented as a power series, where the nonlinear mapping of a quantity is approximated by a polynomial function. The first few terms of the power series expansion reveal the main frequency components obtained as the result of the diode modulation. The following analysis demonstrates what frequency components are present in the modulated signal.

The diode ring mixer has two input ports and one output port. Basically any port can be selected as the output port and the end-result should remain the same. However, the analysis method differs slightly for different configurations of the inputs and outputs. As a hopefully sufficient example, we'll analyse two different forms of the modulator in the following sections.

Also, the official terms used in literature are a bit confusing, since some use the notations LO (local oscillator), RF and IF ports. Basically it would have been better to use those terms here as well, but in this text the 'carrier' basically corresponds to the LO, and 'modulating signal' matches to RF, which is again even more confusing. The 'output signal' refers to IF.

1 Diode ring mixer with carrier input to centre-taps

The first form of the diode ring modulator is a variation, where the carrier signal is fed into the centre-taps of both transformers. To ease out the analysis, the circuit of Figure 1 is redrawn in Figure 2 to show more clearly the diode directions in relation to the two transformers. The labels $1, 2, 3, 4$ are

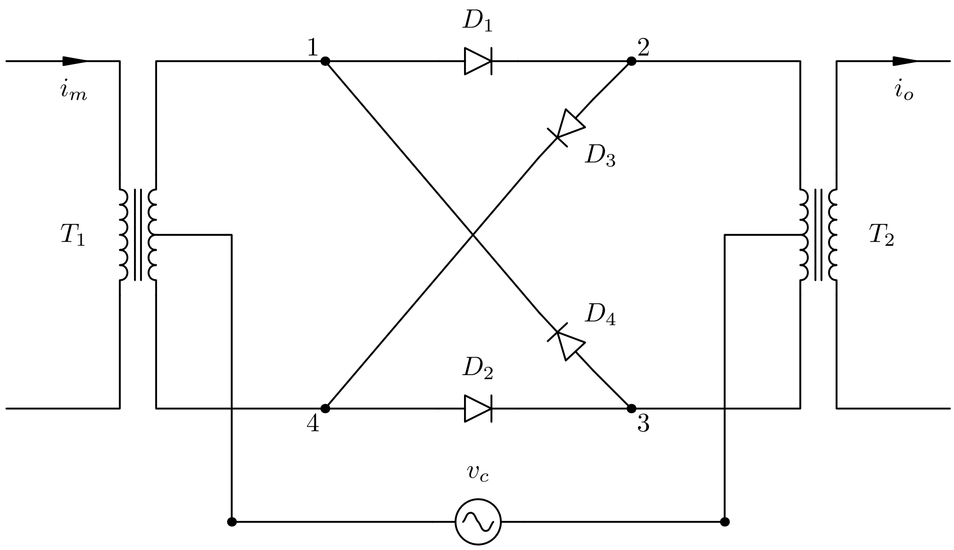

Figure 2: Simplified ring modulator circuit diagram

circuit nodes where the carrier voltage v_c and modulation voltage v_m will produce the control voltage v_d over the diodes. If the voltage difference between the anode and the cathode of the diode is large enough to make the diode forward-biased, then the current through the diodes theoretically obeys the exponential law

$$i_d = I_S \exp(xv_d), \tag{1}$$

where I_S is the diode saturation current, x is a scaling factor consisting of diode internal parameters and v_d is the voltage over the diode, measured from anode to cathode. Hence, v_d should be positive to make the diode conduct. Here we can neglect the details of the scaling factor x, because we are interested only on the fact that the current depends on the voltage via an exponential relation.

In a typical use case of the diode ring modulator, the carrier signal has a much higher frequency and high enough voltage to switch the diodes on and off in both directions. The modulator frequency is typically an audio signal of lower frequencies and lower amplitude, but is of course allowed to have a higher frequency as the carrier (local oscillator) signal. So the carrier controls the diode conductivity and the modulating signal is summed on top of the carrier in the secondary of transformer T_1.

Let's analyse the situation in the two cases where the carrier voltage (typically a simple sine wave) is seen either positive or negative in the direction from left to right of the diagram. Figure 3 illustrates this situation.

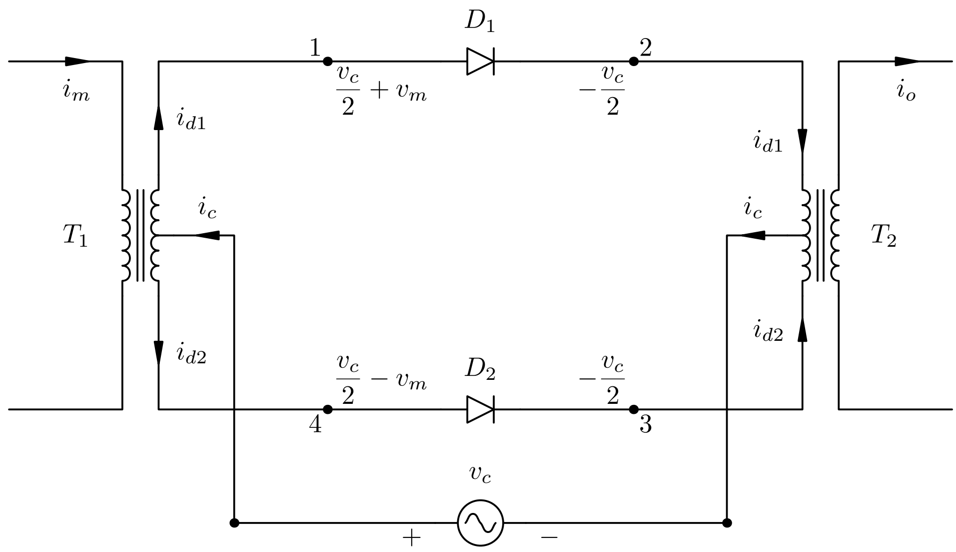

Figure 3: Case of positive carrier control voltage

Since the coil of the transformer looks like a resistor for the carrier frequency, the centre-tapped coil behaves like a current divider, which splits the carrier current i_c into two equal halves with voltage amplitude proportional to $\frac{v_c}{2}$. Transformer T_1 sees the positive side of the carrier and when looking from the T_2 point of view, the carrier looks like a negative signal.

The modulating voltage v_m is composed in the transformer T_1. In this context we neglect the true magnitude of v_m in the primary of T_1 and just concentrate on the signal at the secondary side. Let's assume that the rate of change of the modulating signal is much less than the carrier, so that the analysis assumes a modulating signal having a reasonably steady positive voltage at the top half of the T_1 secondary coil. Since the secondary is centre-tapped, the modulating signal is induced as negative in the bottom half of the secondary coil. Therefore we have $+v_m$ at the top and $-v_m$ a the bottom of T_1 secondary. These voltages are superimposed (summed) on top of the carrier signal at T_1 secondary.

If the amplitude of the carrier voltage is large enough and in this case at positive potential, the diodes D_1 and D_2 are forward-biased. For the situation where diodes D_1 and D_2 are conducting, the currents through the diodes are proportional to the following expressions:

$$i_{d1} = I_S \exp\left(\frac{v_c}{2} + v_m - \frac{-v_c}{2}\right) = I_S \exp\left(v_c + v_m\right)$$
$$i_{d2} = I_S \exp\left(\frac{v_c}{2} + (-v_m) - \frac{-v_c}{2}\right) = I_S \exp\left(v_c - v_m\right)$$

Then the push-pull feature of transformer T_2 comes into play. It means that oppositely travelling currents in a centre-tapped configuration induce a difference signal in the other side of the transformer. If it is now assumed that downwards going current in T_2 secondary will induce an upwards going current in T_2 primary side, then i_{d2} is subtracted from i_{d1}. When simplifying the currents to be proportional only to the voltages inside the exponent function, the power series expansion gives the following difference

$$\begin{array}{rl} & 1 + v_c + v_m + \frac{v_c^2}{2!} + \frac{2v_cv_m}{2!} + \frac{v_m^2}{2!} + \frac{v_c^3}{3!} + \frac{3v_c^2v_m}{3!} + \frac{3v_cv_m^2}{3!} + \frac{v_m{}^3}{3!} + \frac{v_c^3v_m}{4!} + \ldots \\ - & 1 + v_c - v_m + \frac{v_c^2}{2!} - \frac{2v_cv_m}{2!} + \frac{v_m^2}{2!} + \frac{v_c^3}{3!} - \frac{3v_c^2v_m}{3!} + \frac{3v_cv_m^2}{3!} - \frac{v_m{}^3}{3!} - \frac{v_c^3v_m}{4!} + \ldots \\ \hline = & 2v_m + 2v_cv_m + v_c^2v_m + \frac{v_m^3}{6} + \frac{v_c^3v_m}{24}\ldots \end{array}$$

The negative carrier part is illustrated in Figure 4. Here the polarity of v_c has changed and the direction of the currents are opposite to the positive case.

The analysis procedure itself is identical to the positive case. For the situation where diodes D_3 and D_4 are conducting

$$i_{d3} = I_S \exp\left(\frac{v_c}{2} + (-v_m) - \frac{-v_c}{2}\right) = I_S \exp\left(v_c - v_m\right)$$
$$i_{d4} = I_S \exp\left(\frac{v_c}{2} + v_m - \frac{-v_c}{2}\right) = I_S \exp\left(v_c + v_m\right)$$

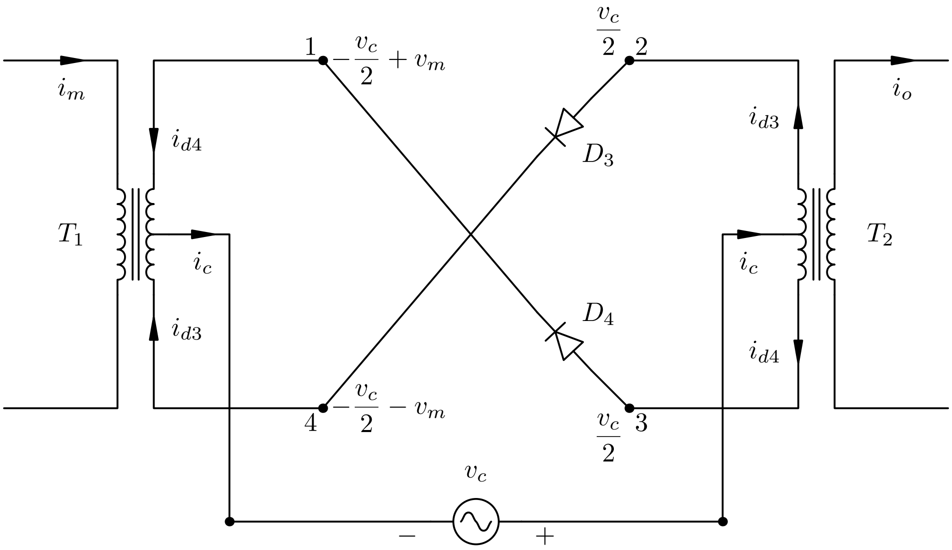

Figure 4: Case of negative carrier control voltage

The same situation with currents i_{d3} and i_{d4} in T_2 secondary will combine into

$$\begin{array}{rl} & 1+v_c-v_m+\frac{v_c^2}{2!}-\frac{2v_cv_m}{2!}+\frac{v_m^2}{2!}+\frac{v_c^3}{3!}-\frac{3v_c^2v_m}{3!}+\frac{3v_cv_m^2}{3!}-\frac{{v_m}^3}{3!}-\frac{v_c^3v_m}{4!}+\ldots \\ - & 1+v_c+v_m+\frac{v_c^2}{2!}+\frac{2v_cv_m}{2!}+\frac{v_m^2}{2!}+\frac{v_c^3}{3!}+\frac{3v_c^2v_m}{3!}+\frac{3v_cv_m^2}{3!}+\frac{{v_m}^3}{3!}+\frac{v_c^3v_m}{4!}+\ldots \\ \hline = & -2v_m-2v_cv_m-v_c^2v_m-\frac{v_m^3}{6}-\frac{v_c^3v_m}{24}\ldots \end{array}$$

The minus sign appears because of the direction of the currents i_{d3} and i_{d4} in the T_2 secondary.

To get a better feel of the remaining frequency components, the signal products must be written open when cosine (or sine) functions are inserted as v_c and v_m. Here we can use the trigonometric rule

$$A_c\cos\omega_c t A_m\cos\omega_m t=\frac{A_cA_m}{2}\left[\cos(\omega_c+\omega_m)t+\cos(\omega_c-\omega_m)t\right], \qquad (2)$$

where A_c and A_m depict the effective amplitudes of the carrier and modulated wave respectively. The main frequency components, based on the evaluation above, of the mixed signal are

$$2A_m\cos\omega_m t+A_cA_m\left[\cos(\omega_c+\omega_m)t+\cos(\omega_c-\omega_m)t\right]. \qquad (3)$$

Obviously later products will add up some excess amplitude to these components and bring in many more harmonic and mixed-mode components.

Since the non-cancelling components appear when v_m multiplies with a minus sign, it would be intuitive to assume that all the frequency components could be expressed as

$$\omega_m + [i\omega_c \pm j\omega_m], \tag{4}$$

where $i = 1, 2, 3, 4, ..., \infty$ and $j = 1, 3, 5, 7, ..., \infty$. That is, i goes through all positive integers and j goes through all *odd* positive integers for each i. Also the amplitudes of frequency components i and j will decrease as the indices increase. But I don't have the knowledge to prove this formula for certain.

And, in addition, because of the switching action which causes the phase changes, there surely will be additional (odd?) harmonics present in the complete frequency response. These harmonics due to the switching action were not considered during the previous analysis. An FFT analysis from the output signal would indicate how the switching affects the frequency content of the mixing product.

2 Diode ring mixer with output between centre-taps

In this configuration the carrier signal is applied to the transformer T_1 and the modulating signal to T_2. The output voltage is measured over the load resistor R_L. First we will analyse the situation when the carrier signal is positive, then the diodes D_1 and D_3 will conduct and D_2 and D_4 will be cut off. This is shown in Figure 5.

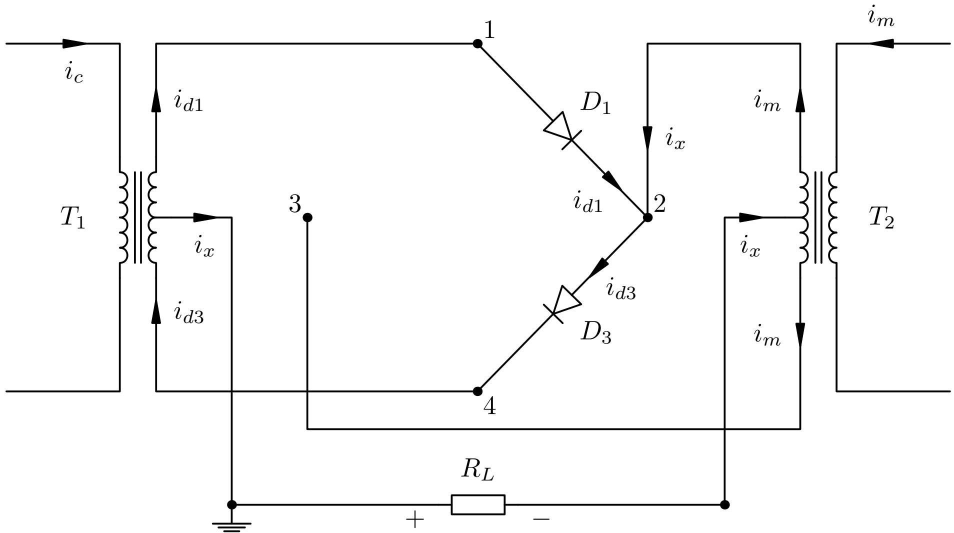

Figure 5: Positive side of the carrier signal

Considering that there is again two control voltages v_c and v_m for the diodes, the following current equations can be written:

$$i_{d1} = I_S \exp(v_c - v_m)$$
$$i_{d3} = I_S \exp(v_m - (-v_c)) = I_S \exp(v_c + v_m).$$

Because now there are distinct voltage/current nodes in the circuit, one can use the Kirchhoff rule, where the sum of currents entering and leaving a node will equal zero. Here we take the polarity assumption that entering current will be positive and leaving current negative. Using the notations in Figure 5, the current equation for node 2 is

$$i_{d1} - i_{d3} + i_x = 0. \tag{5}$$

The direction of the current i_x is still unknown, but after working out the calculation of equation (5), then the direction will be found. So, from equation (5),

$$i_x = I_S \exp(v_c + v_m) - I_S \exp(v_c - v_m). \tag{6}$$

This difference has been calculated already in the previous section, so we already know that the result is

$$i_x = 2v_m + 2v_c v_m + v_c^2 v_m + \frac{v_m^3}{6} + \frac{v_c^3 v_m}{24} + \ldots \tag{7}$$

The direction of i_x is therefore the same as drawn in Figure 5. This current flows through the load resistor and creates the output voltage accordingly.

For the negative case of v_c, diodes D_2 and D_4 will conduct, and v_m will come out the other side of the centre-tap having opposite polarity $-v_m$ and this will change the polarity of i_x so that

$$i_x = -2v_m - 2v_c v_m - v_c^2 v_m - \frac{v_m^3}{6} - \frac{v_c^3 v_m}{24} \ldots \tag{8}$$

This will switch the polarity of the output voltage over the load resistor. As a conclusion, the output signal is essentially the same using this input-output configuration of the diode ring modulator.

3 Some experimental results from calculations

The following two plots show some examples on using some random input waveforms to get some clean-looking output waveform. The equation used for drawing the waves is not the approximate series expansion, but a numerical evaluation using the exact exponential form as $e^{\cos(x)} - e^{\cos(y)}$, which will automatically contain all the possible harmonics.

Figure 6 shows a typical example, where the modulator is used 'backwards' for creating basic amplitude modulation. The modulating frequency is 10 times the carrier frequency. Notice that every time when the carries crosses 0, the output waveform experiences a 180 degree phase shift.

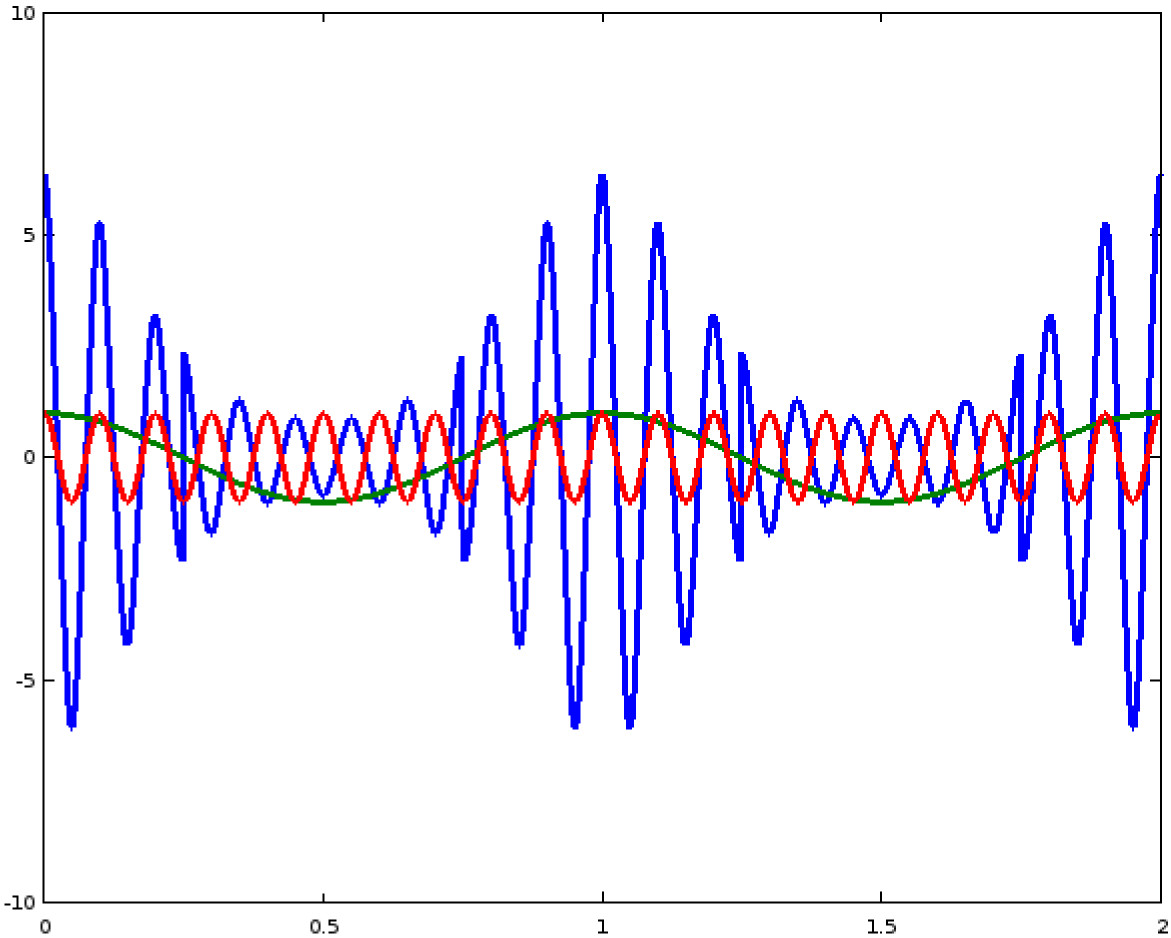

Figure 6: $\omega_m = 10\omega_c$, carrier on green, modulating on red and output on blue

Figure 7 shows an example, where the carrier is of double the frequency of the modulating signal. Again note the 180 degree phase shift at the output signal when the carrier crosses 0.

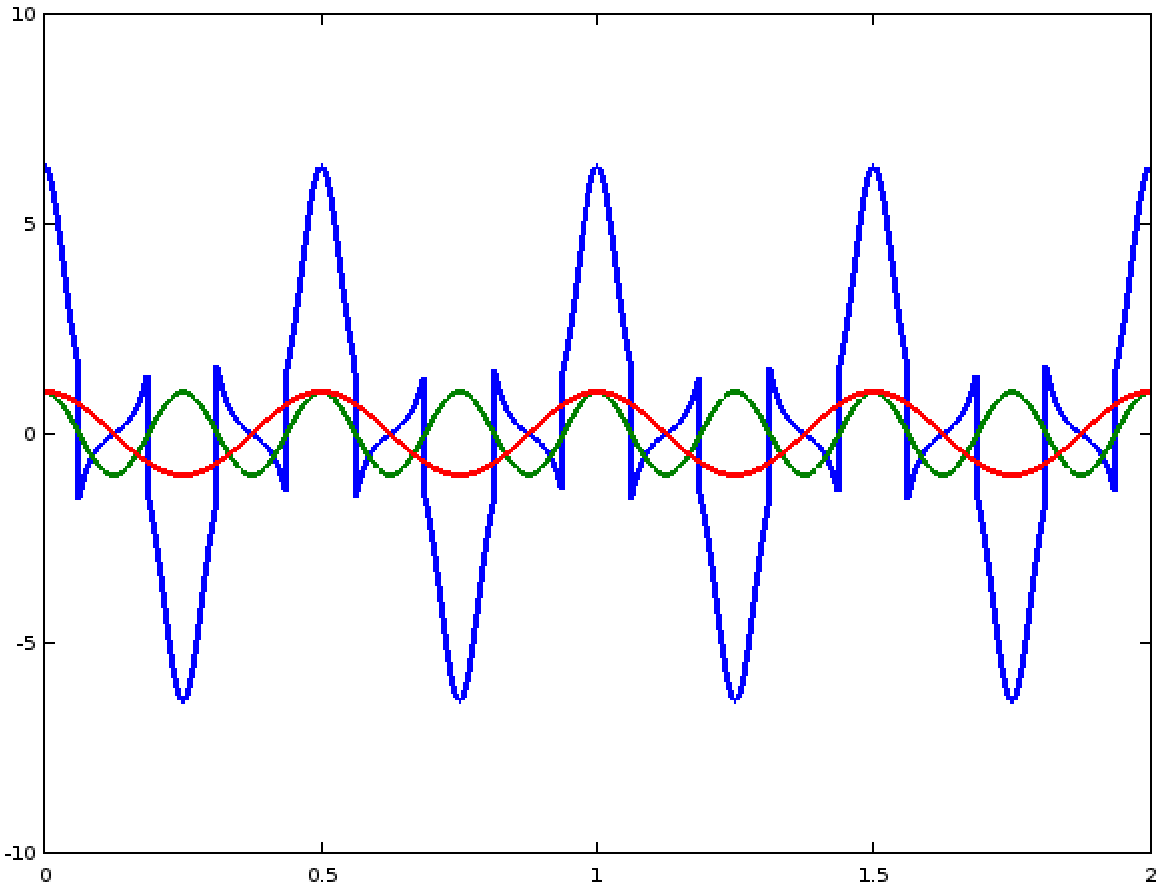

Figure 7: $\omega_c = 2\omega_m$, carrier on green, modulating on red and output on blue