Analog Linear Modulation and Demodulation

Goal

The goal of this experiment is to study and analyze the analog linear modulation and demodulation techniques in communication systems.

Theory

Analog linear modulation

In electronics and telecommunications, modulation is the process of varying one or more properties of a periodic waveform, called the carrier signal, with a modulating signal that typically contains information to be transmitted.

In telecommunications, modulation is the process of conveying a message signal, for example a digital bit stream or an analog audio signal, inside another signal that can be physically transmitted. Modulation of a sine waveform transforms a baseband message signal into a passband signal.

The aim of analog modulation is to transfer an analog baseband (or lowpass) signal, for example an audio signal or TV signal, over an analog bandpass channel at a different frequency, for example over a limited radio frequency band or a cable TV network channel.

In linear modulation, the amplitude of the transmitted signal varies linearly with the modulating digital signal, $m(t)$. In this experiment, the AM, DSBSC and SSB modulation techniques are considered.

Since in all of the mentioned modulation techniques the envelope of the signal is transmitted, first we should elaborate the definition of the envelope.

Envelope definition

When we talk of the envelopes of signals we are concerned with the appearance of signals in the time domain. Textbooks are full of drawings of modulated signals, and you already have an idea of what the term “envelope” means. It will now be given a more formal definition.

Qualitatively, the envelope of a signal $y(t)$ is that boundary within which the signal is contained, when viewed in the time domain. It is an imaginary line.

This boundary has an upper and lower part. You will see these are mirror images of each other. In practice, when speaking of the envelope, it is customary to consider only one of them as “the envelope” (typically the upper boundary).

Although the envelope is imaginary in the sense described above, it is possible to generate, from $y(t)$, a signal $e(t)$, having the same shape as this imaginary line. The circuit which does this is commonly called an envelope detector.

1. AM

a) AM modulation

In the early days of wireless, communication was carried out by telegraphy, the radiated signal being an interrupted radio wave. Later, the amplitude of this wave was varied in sympathy with (modulated by) a speech message (rather than on/off by a telegraph key), and the message was recovered from the envelope of the received signal. The radio wave was called a 'carrier', since it was seen to carry the speech information with it. The process and the signal were called ***amplitude modulation***, or 'AM' for short.

In the context of radio *communication*s, near the end of the 20th century, few modulated signals contain a significant component at 'carrier' frequency. However, despite the fact that a carrier is not radiated, the need for such a signal at the transmitter (where the modulated signal is generated), and also at the receiver, remains fundamental to the modulation and demodulation process respectively. The use of the term 'carrier' to describe this signal has continued to the present day.

As distinct from radio communications, present day radio *broadcasting* transmissions do have a carrier. By transmitting this carrier the design of the demodulator, at the receiver, is greatly simplified, and this allows significant cost savings.

The most common method of AM generation uses a 'class C modulated amplifier'; such an amplifier is not available in the BASIC TIMS set of modules. It is well documented in text books. This is a 'high level' method of generation, in that the AM signal is generated at a power level ready for radiation. It is still in use in broadcasting stations around the world, ranging in powers from a few tens of watts to many megawatts.

Unfortunately, text books which describe the operation of the class C modulated amplifier tend to associate properties of this particular method of generation with those of AM, and AM generators, in general. This gives rise to many misconceptions. The worst of these is the belief that it is impossible to generate an AM signal with a depth of modulation exceeding 100% without giving rise to serious RF distortion.

You will see in this experiment that there is no problem in generating an AM signal with a depth of modulation exceeding 100%, and without any RF distortion whatsoever. But we are getting ahead of ourselves, as we have not yet even defined what AM is!

The amplitude modulated signal is defined as:

$$AM = E[1 + m * cos(\mu t)]\cos(\omega t) \tag{1}$$

$$= A[1 + m * \cos(\mu t)] * Bcos(\omega t) \tag{2}$$

$$= [low\ frequency\ term\ a(t)] * [high\ frequency\ term\ c(t)] \tag{3}$$

Here:

'**E**' is the AM signal amplitude from eqn. (1). For modelling convenience eqn. (1) has been written into two parts in eqn. (2), where $(A * B) = E$.

'**m**' is a constant, which, as you will soon see, defines the 'depth of modulation'. Typically m < 1. Depth of modulation, expressed as a percentage, is 100.m. There is no inherent restriction upon the size of 'm' in eqn. (1). This point will be discussed later.

'**μ**' and '**ω**' are angular frequencies in rad/s, where $\mu/2\pi$ is a low, or message frequency, say in the range 300 Hz to 3000 Hz; and $\omega/2\pi$ is a radio, or relatively high, 'carrier' frequency. In TIMS the carrier frequency is generally 100 kHz.

Notice that the term $a(t)$ in eqn. (3) contains both a DC component and an AC component. As will be seen, it is the DC component which gives rise to the term at ω - the 'carrier' - in the AM signal. The AC term '$m *$

$cos(\mu t)$' is generally thought of as the message, and is sometimes written as $m(t)$. But strictly speaking, to be compatible with other mathematical derivations, the whole of the low frequency term $a(t)$ should be considered the message. Thus:

$$a(t) = DC + m(t) \tag{4}$$

Figure 1 below illustrates what the oscilloscope will show if displaying the AM signal. A block diagram representation of eqn. (2) is shown in Figure 6 below. In the experiment you will model eqn. (2) by the arrangement of Figure 2. The depth of modulation will be set to exactly 100% (m = 1). You will gain an appreciation of the meaning of 'depth of modulation', and you will learn how to set other values of 'm', including cases where m > 1.

The signals in eqn. (2) are expressed as voltages in the time domain. You will model them in two parts, as written in eqn. (3).

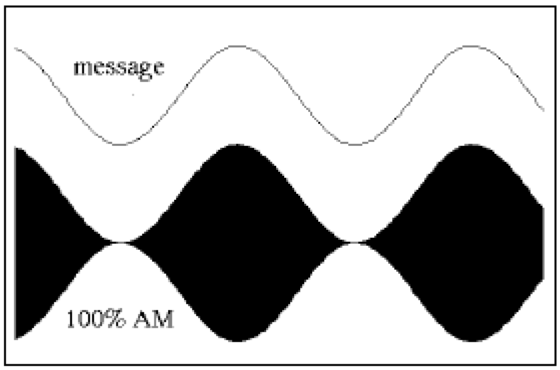

Figure 1- AM, with m = 1, as seen on the oscilloscope

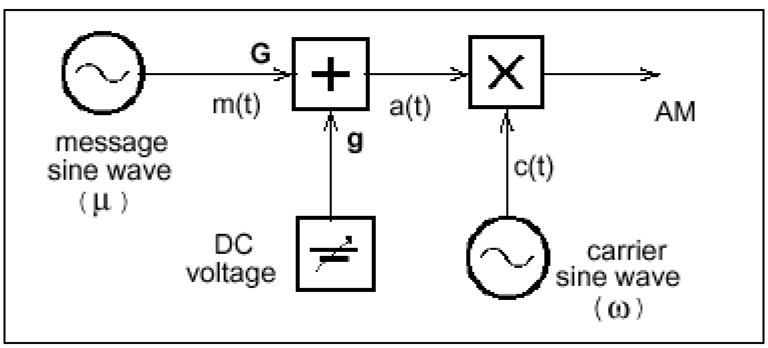

Figure 2: Generation of eqn. (2)

Depth of modulation

100% amplitude modulation is defined as the condition when $m = 1$. Just what this means will soon become apparent. It requires that the amplitude of the DC ($= A$) part of $a(t)$ is equal to the amplitude of the AC part $(= A * m)$. This means that their ratio is unity at the *output* of the ADDER, which forces 'm' to a magnitude of exactly unity.

By aiming for a ratio of unity it is thus not necessary to know the absolute magnitude of A at all.

The magnitude of 'm' can be measured directly from the AM display itself. Thus,

$$m = \frac{(P - Q)}{(P + Q)} \tag{5}$$

where P and Q are as defined in Figure 3.

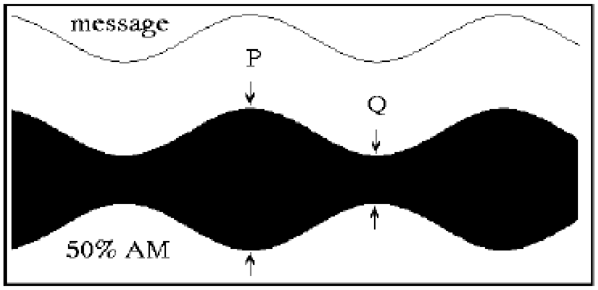

Figure 3 : Oscilloscope display for the case m = 0.5

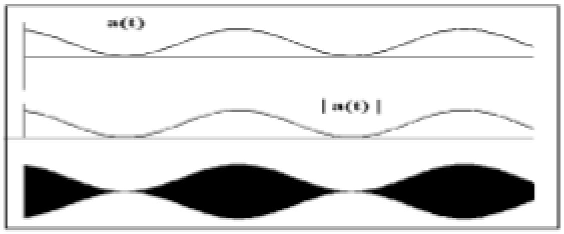

Figure 4 : AM with m = 1

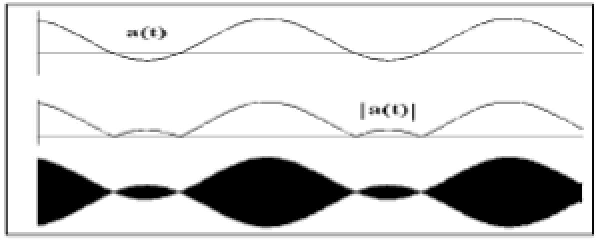

Figure 5 : AM with m = 1.5

b) AM demodulation

Envelope detector

Although the envelope is imaginary in the sense described above, it is possible to generate, from y(t), a signal e(t), having the same shape as this imaginary line. The circuit which does this is commonly called an envelope detector. A better word for envelope detector would be envelope generator, since that is what these circuits do.

In this experiment, you will model circuits which will generate these envelope signals.

Diode detector

The ubiquitous diode detector is the prime example of an envelope generator. It is well documented in most textbooks on analog modulation. It is synonymous with the term 'envelope demodulator' in this context. But remember: the diode detector is an approximation to the ideal. We will first examine the ideal circuit.

Ideal envelope detector

The ideal envelope detector is a circuit which takes the absolute value of its input, and then passes the result through a low pass filter. The output from this low pass filter is the required envelope signal. See Figure 6.

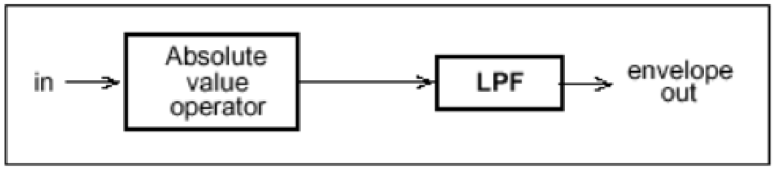

Figure 6 : Ideal envelope recovery arrangement

The absolute value operation, being non-linear, must generate some new frequency components. Among them are those of the wanted envelope. Presumably, since the arrangement actually works, the unwanted components lie above those wanted components of the envelope.

It is the purpose of the low pass filter to separate the wanted from the unwanted components generated by the absolute value operation.

The analysis of the ideal envelope recovery circuit, for the case of a general input signal, is not a trivial mathematical exercise, the operation being non-linear. So it is not easy to define beforehand where the unwanted components lie.

Ideal rectifier

A circuit which takes an absolute value is a full wave rectifier. Note carefully that the operation of rectification is non-linear. The so-called ideal rectifier is a precision realization of a rectifier, using an operational amplifier and a diode in a negative feedback arrangement. It is described in text books dealing with the applications of operational amplifiers to analog circuits. An extension of the principle produces an ideal full wave rectifier.

You will find a half wave rectifier is generally adequate for use in an envelope recovery circuit.

Envelope bandwidth

You know what a low pass filter is, but what should be its cut-off frequency in this application? The answer: ‘the cut-off frequency of the low pass filter should be high enough to pass all the wanted frequencies in the envelope, but no more’. So you need to know the envelope bandwidth.

In a particular case, you can determine the expression for the envelope from the definition given before, and the bandwidth by Fourier series analysis. Alternatively, you can estimate the bandwidth, by inspecting its shape on an oscilloscope, and then applying rules of thumb which give quick approximations.

An envelope will always include a constant, or DC, term.

This is inevitable from the definition of an envelope - which includes the operation of taking the absolute value. It is inevitable also in the output of a practical circuit, by the very nature of rectification.

The presence of this DC term is often forgotten. For the case of an AM signal, modulated with music, the DC term is of little interest to the listener. But it is a direct measure of the strength of the carrier term, and so is used as an automatic gain control signal in receivers.

It is important to note that it is possible for the bandwidth of the envelope to be much wider than that of the signal of which it is the envelope. In fact, except for the special case of the envelope modulated signal, this is generally so. An obvious example is that of the DSBSC signal derived from a single tone message.

2. DSBSC

a) DSBSC modulation

Consider two sinusoids, or cosinusoids, $\cos(\mu t)$ and $\cos(\omega t)$. A double sideband suppressed carrier signal, or DSBSC, is defined as their product, namely

$$DSBSC = E\cos(\mu t)\cos(\omega t) \tag{6}$$

Generally, and in the context of this experiment, it is understood that

$$\omega \gg \mu \tag{7}$$

Eqn. (6) can be expanded to give:

$$E\cos(\omega t)\cos(\mu t) = \frac{E}{2}\cos(\omega - \mu)\,t + \frac{E}{2}\cos(\omega + \mu)t \tag{8}$$

Eqn. (8) shows that the product is represented by two new signals, one on the sum frequency $(\omega + \mu)$, and one on the difference frequency $(\omega - \mu)$. See Figure 7.

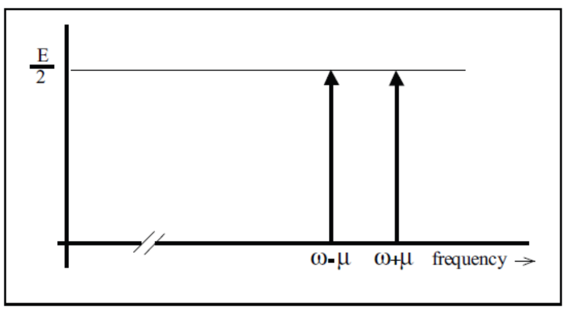

Figure 7: Spectral components

Remembering the inequality of eqn. (7). the two new components are located close to the frequency ω rad/s, one just below, and the other just above it. These are referred to as the *lower* and *upper* sidebands respectively. These two components were derived from a 'carrier' term on ω rad/s, and a message on μ rad/s. Because there is no term at carrier frequency in the product signal it is described as a double sideband *suppressed* carrier (DSBSC) signal.

The term 'carrier' comes from the context of 'double sideband amplitude modulation' (commonly abbreviated to just AM). AM is introduced later (although, historically, AM preceded DSBSC).

The time domain appearance of a DSBSC, eqn. (6), in a text book is generally as shown in Figure 8.

Notice the waveform of the DSBSC in Figure 8, especially near the times when the message amplitude is zero. The fine detail differs from period to period of the message. This is because the ratio of the two frequencies μ and ω has been made non-integral.

Although the message and the carrier are periodic waveforms (sinusoids), the DSBSC itself need not necessarily be periodic.

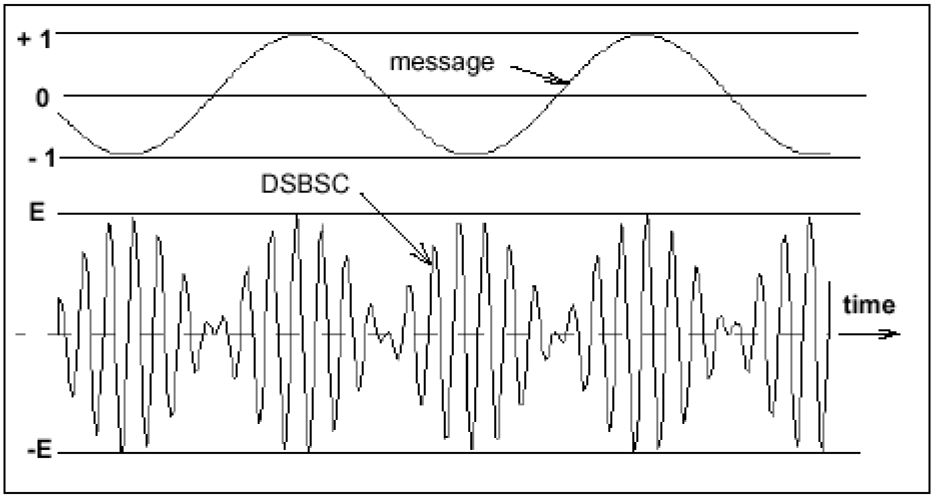

Figure 8: eqn. (6) – a DSBSC – seen in the time domain

A block diagram, showing how eqn. (6) could be modeled with hardware, is shown in Figure 9 below.

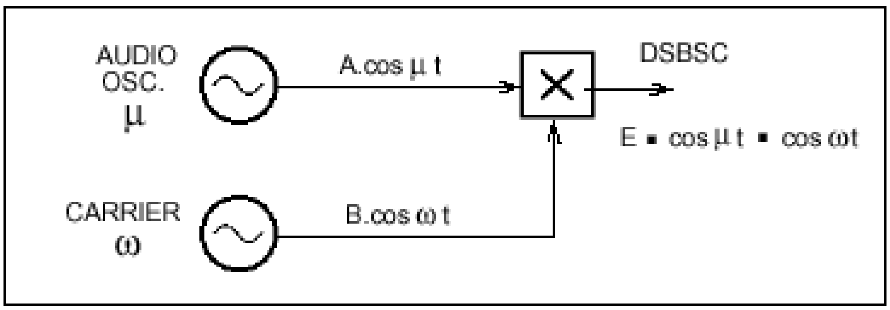

Figure 9: Block **Error! Reference source not found.**diagram to generate eqn. (6) with hardware.

Multi-tone message

The DSBSC has been defined in eqn. (1), with the message identified as the low frequency term. Thus:

$$message = \cos(\mu t) \tag{9}$$

For the case of a multi-tone message, $m(t)$, where

$$m(t) = \sum_i a_i \cos(\mu_i t) \tag{10}$$

then the corresponding DSBSC signal consists of a band of frequencies below ω, and a band of frequencies above ω. Each of these bands is of width equal to the bandwidth of $m(t)$. The individual spectral components in these sidebands are often called side frequencies.

If the frequency of each term in the expansion is expressed in terms of its difference from ω, and the terms are grouped in pairs of sum and difference frequencies, then there will be ‘n’ terms of the form of the right hand side of eqn. (8).

Note it is assumed here that there is no DC term in $m(t)$. The presence of a DC term in $m(t)$ will result in a term at ω in the DSB signal; that is, a term at 'carrier' frequency. It will no longer be a double sideband suppressed carrier signal. A special case of a DSB with a significant term at carrier frequency is an amplitude modulated signal, which will be examined later.

A more general definition still, of a DSBSC, would be:

$$DSBSC = E * m(t)\cos(\omega t) \tag{11}$$

where $m(t)$ is any (low frequency) message. By convention $m(t)$ is generally understood to have a peak amplitude of unity (and typically no DC component).

b) DSBSC demodulation

Productive demodulation

All of the modulated signals you have seen so far may be defined as narrow band. They carry message information. Since they have the capability of being based on a radio frequency carrier (suppressed or otherwise) they are suitable for radiation to a remote location. Upon receipt, the object is to recover -demodulate - the message from which they were derived.

In the discussion to follow the explanations will be based on narrow band signals. But the findings are in no way restricted to narrow band signals; they happen to be more convenient for purposes of illustration.

Frequency translation

When a narrow band signal $y(t)$ is multiplied with a sine wave, two new signals are created – on the 'sum and difference' frequencies.

Figure 10 illustrates the action for a signal $y(t)$, based on a carrier f_c, and a sinusoidal oscillator on frequency f_0.

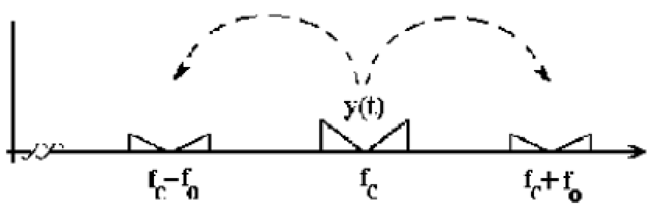

Figure 10: 'sum and difference frequencies'

Each of the components of $y(t)$ was moved up an amount fo in frequency, and down by the same amount, and appear at the output of the multiplier.

Remember, neither $y(t)$, nor the oscillator signal, appears at the multiplier output. This is not necessarily the case with a 'modulator'.

A filter can be used to select the new components at either the sum frequency (BPF preferred, or an HPF) or difference frequency (LPF preferred, or a BPF).

The combination of MULTIPLIER, OSCILLATOR, and FILTER is called a frequency translator.

Synchronous demodulator

When the frequency translation is down to baseband the frequency translator becomes a demodulator.

Synchronous demodulator $f_0 = f_c$:

For successful demodulation of DSBSC and AM the synchronous demodulator requires a 'local carrier' of exactly the same frequency as the carrier from which the modulated signal was derived, and of fixed relative phase, which can then be adjusted, as required, by the phase changer shown.

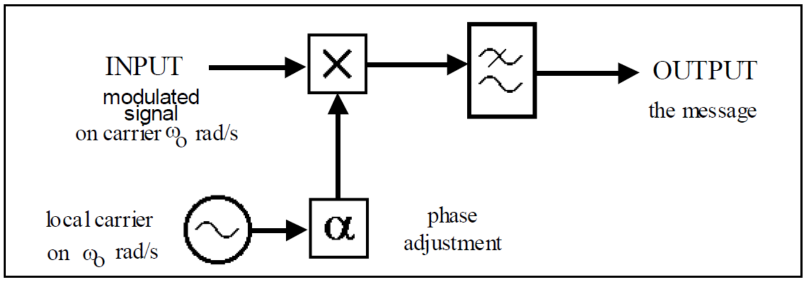

Figure 11: Synchronous demodulator

3. SSB

a) SSB modulation

There are three well known methods of SSB generation using analog techniques, namely the filter method, the phasing method, and Weaver's method. This experiment will study the phasing method.

Filter method

We have already modeled a DSBSC signal. An SSB signal may be derived from this by the use of a suitable bandpass filter – commonly called, in this application, an SSB sideband filter. This, the filter method, is probably the most common method of SSB generation. Mass production has given rise to low cost, yet high performance, filters. But these filters are generally only available at 'standard' frequencies (for example 455 kHz, 10.7 MHz) and SSB generation by the filter method at other frequencies can be expensive. For this reason, TIMS no longer has a 100 kHz SSB filter module, although a decade ago these were in mass production and relatively inexpensive.

Phasing method

The phasing method of SSB generation, which is the subject of this experiment, does not require an expensive filter, but instead an accurate phasing network, or quadrature phase splitter (QPS). It is capable of acceptable performance in many applications. The QPS operates at baseband, no matter what the carrier frequency (either intermediate or final), in contrast to the filter of the filter method.

SSB signal

Recall that, for a single tone message $\cos(\mu t)$, a DSBSC signal is defined by:

$$DSBSC = A * \cos(\mu t)\cos(\omega t) \quad (12)$$

$$= \frac{A}{2}\cos(\omega - \mu)\,t + \frac{A}{2}\cos(\omega + \mu)\,t \quad (13)$$

(14)

$$= lower\ sideband + upper\ sideband$$

When, say, the lower sideband (LSB) is removed, by whatever method, then the upper sideband (USB) remains.

$$USB = \frac{A}{2}\cos(\omega + \mu)\, t \tag{15}$$

This is a single frequency component at frequency $(\omega + \mu)/(2\pi)$Hz. It is a (co)sine wave. Viewed on an oscilloscope, with the time base set to a few periods of ω, it looks like any other sine wave. What is its envelope?

The USB signal of eqn. (15) can be written in the following form:

$$USB = a(t)\cos[(\omega + \mu)t + p(t)] \tag{16}$$

The envelope has been defined as:

$$envelope = |a(t)| = A/2 \tag{17}$$

Thus, the envelope is a constant (ie., a straight line) and the oscilloscope, correctly set up, will show a rectangular band of color across the screen.

This result may seem at first confusing. One tends to ask: 'where is the message information'?

Answer: the message amplitude information is contained in the amplitude of the SSB, and the message frequency information is contained in the frequency offset, from w, of the SSB.

An SSB derived from a single tone message is a very simple example. When the message contains more components the SSB envelope is no longer a straight line. Here is an important finding!

An ideal SSB generator, with a single tone message, should have a straight line for an envelope.

Any deviation from this suggests extra components in the SSB itself. If there is only one extra component, say some 'leaking' carrier, or an unwanted sideband not completely suppressed, then the amplitude and frequency of the envelope will identify the amplitude and frequency of the unwanted component.

A most important characteristic of any SSB generator is the amount of out-of-band energy it produces, relative to the wanted output. In most cases this is determined by the degree to which the unwanted sideband is suppressed. A ratio of wanted-to-unwanted output power of 40 dB was once considered acceptable commercial performance; but current practice is likely to call for a suppression of 60 dB or more, which is not a trivial result to achieve.

Phasing generator

The phasing method of SSB generation is based on the addition of two DSBSC signals, so phased that their upper sidebands (say) are identical in phase and amplitude, whilst their lower sidebands are of similar amplitude but opposite phase.

The two out-of-phase sidebands will cancel if added; alternatively, the in-phase sidebands will cancel if subtracted. The principle of the SSB phasing generator is illustrated in Figure 12.

Notice that there are two 90° phase changers. One operates at carrier frequency, the other at message frequencies.

The carrier phase changer operates at a single, fixed frequency, ω rad/s.

The message is shown as a single tone at frequency μ rad/s. But this can lie anywhere within the frequency range of speech, which covers several octaves. A network providing a constant 90° phase shift over this frequency range is very difficult to design. This would be a wideband phase shifter, or Hilbert transformer.

In practice a wideband phase splitter is used. This is shown in the arrangement of Figure 13.

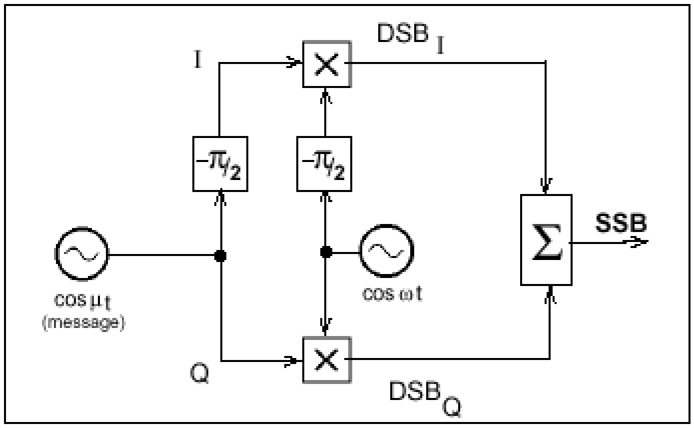

Figure 12: Principle of the SSB Phasing Generator

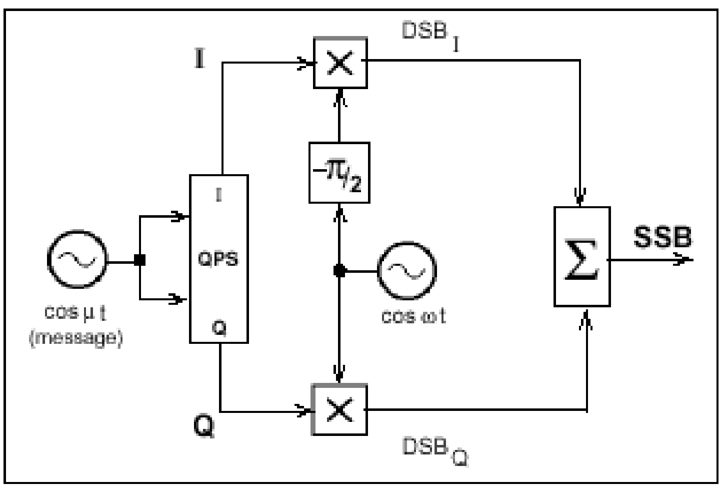

Figure 13: Practical realization of the SSB Phasing Generator

The wideband phase splitter consists of two complementary networks -say I (in-phase) and Q (quadrature). When each network is fed from the same input signal the phase difference between the two outputs is maintained at 90°.

Note that the phase difference between the common input and either of the outputs is not specified; it is not independent of frequency. Study Figure 8 and Figure 9 to ensure that you appreciate the difference.

At the single frequency μ rad/s the arrangements of Figure 12 and Figure 13 will generate two DSBSC. These are of such relative phases as to achieve the cancellation of one sideband, and the reinforcement of the other, at the summing output.

You should be able to confirm this. You could use graphical methods (phasors) or trigonometrical analysis.

The QPS may be realized as either an active or passive circuit, and depends for its performance on the accuracy of the components used. Over a wide band of audio frequencies, and for a common input, it maintains a phase difference between the two outputs of 90 degrees, with a small frequency-dependent error (typically equiripple).

Performance Criteria

As stated earlier, one of the most important measures of performance of an SSB generator is its ability to eliminate (suppress) the unwanted sideband. To measure the ratio of wanted-to-unwanted sideband suppression directly requires a SPECTRUM ANALYSER. In commercial practice these instruments are very expensive, and their purchase cannot always be justified merely to measure an SSB generator performance.

As always, there are indirect methods of measurement. One such method depends upon a measurement of the SSB envelope, as already hinted.

Suppose that the output of an SSB generator, when the message is a single tone of frequency μ rad/s, consists only of the wanted sideband W and a small amount of the unwanted sideband U.

It may be shown that, for U << W, the envelope is nearly sinusoidal and of a frequency equal to the frequency difference of the two components. Thus, the envelope frequency is 2μ rad/s.

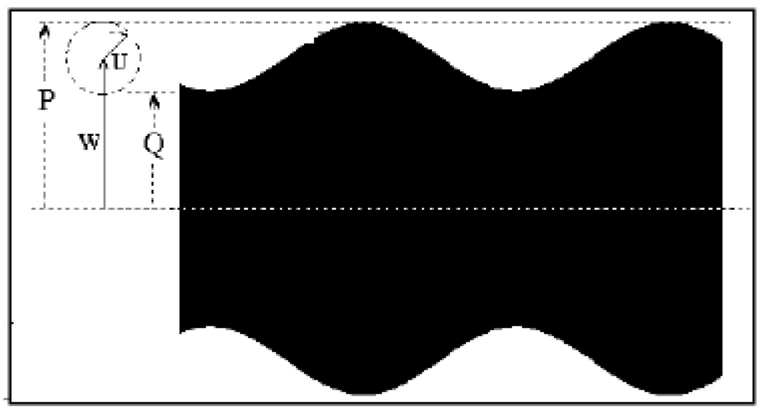

Figure 14: Measuring sideband suppression via the envelope

It is a simple matter to measure the peak-to-peak and the trough-to-trough amplitudes, giving twice P, and twice Q, respectively. Then:

$$P = W + U \tag{18}$$

$$Q = W - U \tag{19}$$

as seen from the phasor diagram. This leads directly to

$$sideband\ suppression = 20\log_{10}\left[\frac{P+Q}{P-Q}\right]\ (dB) \tag{20}$$

U is in fact the sum of several small components then an estimate of the wanted to unwanted power ratio can still be made. Note that it would be greater (better) than for the case where U is a single component.

A third possibility, the most likely in a good design, is that the envelope becomes quite complex, with little or no stationary component at either μ or μ/2; in this case the unwanted component(s) are most likely system noise.

Make a rough estimate of the envelope magnitude, complex in shape though it may well be, and from this can be estimated the wanted to unwanted suppression ratio. This should turn out to be better than 26 dB in TIMS, in which case the system is working within specification. The TIMS QPS module does not use precision components, nor is it aligned during manufacture. It gives only moderate sideband suppression, but it is ideal for demonstration purposes.

Within the 'working frequency range' of the QPS the phase error from 90 between the two outputs will vary with frequency (theoretically in an equi-ripple manner).

SSB demodulation

The demodulation types of SSB are the same as DSBSC demodulation.