modulation systems

The recent energy crises have underlined the need to forge ahead with the development of new communication systems, not only to alleviate the ever-increasing waveband congestion, but also to make more economical use of transmitter power.

It is hoped that this article will give an insight into the various modulation systems now in use and will also explain the design philosophy of a new transceiver developed by Elektor.

Communication systems

The purpose of a communication system is to convey information from one location to another (distant) location. The block diagram of a communication system is given in figure 1. It comprises three parts:

- an encoder to convert the information into a form suitable for transmission via the medium.
- the medium.
- a decoder to convert the information back into its original form.

One of the oldest communication systems utilises the human voice. Information from the brain is encoded into mechanical vibrations by the vocal system, transmitted via the air and reconverted by the aural system of the listener into information in the brain.

This system, although still widely used, has its drawbacks. Notably that the range is limited by the power of the lungs. The system is also subject to interference from nagging wives, mothers-in-law etc. and prone to breakdowns due to laryngitis and other complaints.

As another example of a communication system consider the postal system. Information from the brain is encoded in the form of writing, transmitted via the postal system (the medium), and decoded by the optical system of the recipient.

These two examples both require direct human intervention in the transmission and reception of the information, but this is not always necessary. Two computers, connected by a data link could carry on a meaningful dialogue without human interference or an unmanned meteorological station might transmit data to a remote terminal. Communication systems may therefore be divided into at least two categories:

- systems in which information perceptible to human sense organs is transmitted and in which the ultimate receiver is a human sense organ which decodes the information in conjunction with the brain. Radio broadcasting and television fall into this category.
- systems in which human senses play no part in the decoding process.

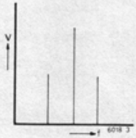
The difference between the two stems from the fact that human senses can operate very selectively so that the desired information can be extracted in the presence of large amounts of unwanted information (noise & c.). This faculty may be further improved by training, so that a radio operator can frequently distinguish signals that would be unintelligible to the layman.

It is thus possible to subdivide the first category into two sub-groups:

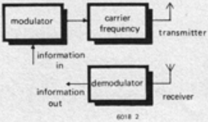
- systems in which the impairment of the transmitted information must be as small as possible, for instance high-fidelity f.m. broadcasting and television.

1

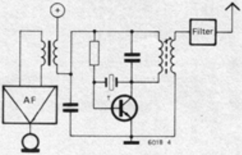
3



2



4



- systems whose functional capability is little impaired by distortion of the information content or even by omission of a large part of it.

Systems such as the telephone, which in general convey only speech, fall into this category. Speech is still intelligible even after removal of a large amount of the information by restricting the bandwidth or by other means.

The concept of Modulation

When electromagnetic radiation serves as a medium for the transmission (i.e. as a carrier) it is necessary to impress the information onto the carrier by changing one or more of its parameters (i.e. to modulate it). The decoder at the receiving and reconverts these changes into information.

If the discussion is confined to analogue modulation there are two important types:

- amplitude modulation (AM)
- frequency modulation (FM)

In amplitude modulation the variable parameter representing the information is the amplitude of the carrier, whereas for frequency modulation the variable is the frequency of the carrier.

The more important forms of amplitude modulation are as follows:

- double sideband with carrier (DSB)
- double sideband suppressed carrier (DSSC)
- single sideband suppressed carrier (SSB)
- carrier position modulation (CPM)

DSB

In a DSB system a low-frequency signal is used to change the amplitude of a radio-frequency carrier. In the absence of modulation the carrier continues to be radiated at a certain level. At maximum modulation the minimum value that the carrier amplitude can assume is nil, which occurs on the 'troughs' of the modulating waveform. It therefore follows that if the modulation is linear the maximum amplitude is twice the amplitude of the unmodulated carrier on the peaks of the modulating waveform. A DSB signal is shown in photo 1. The mathematical expression for this form of modulation with a sinusoidal modulating signal is as follows:

$$v=[1+m\cdot\cos(\omega_{AF}t)]\cdot v_0\cos(\omega_{RF}t) \qquad (1)$$

where m is the modulation index
$v_0\cos(\omega_{RF}t)$ is the carrier, of which v_0 is the peak unmodulated value
$\cos(\omega_{AF}t)$ is the modulating signal.

The modulation index can have values between zero (no modulation) and unity (maximum modulation). The depth of modulation is frequently expressed as a percentage in which case 100% corresponds to a modulation index of 1. In commercial broadcast transmitters the depth of modulation is around 30%, which occurs when the AF signal reaches its maximum value, that is $\cos(\omega_{AF}t)=1$. The mean value must therefore necessarily be lower.

Multiplying out equation (1) gives:

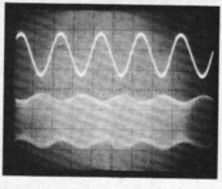
$$v=v_0\cdot\cos(\omega_{RF}t)+\frac{mvo}{2}\cdot[\cos(\omega_{RF}+\omega_{AF})t+\cos(\omega_{RF}-\omega_{AF})t] \qquad (2)$$

From this equation it can be seen that the low-frequency information appears in two sidebands, placed symmetrically above and below the carrier frequency. Figure 3 shows the frequency spectrum of a DSB signal. Of course with a complex modulating waveform the sidebands are not single frequencies but a spectrum of frequencies occupying a bandwidth equal to ± the highest modulating frequency on each side of the carrier. It can be seen from the equation that even with a modulation index of 1, half the energy radiated is at the carrier frequency and contains no information. In fact commercial broadcast transmitters operate with a mean modulation depth considerably less than 100%. It is therefore apparent that transmitter power amounting to many gigawatts is being radiated uselessly into space by transmitters around the world. Apart from the waste of energy other undesirable phenomena occur, such as cross-modulation in the ionosphere (the Luxembourg effect). Furthermore, the system is inefficient in its use of bandwidth since it uses two sidebands each containing the complete LF information. One of these is clearly redundant.

It seems legitimate to ask why, in the face of all these objections, DSB is the most common system in use at the present time. There are two reasons. Firstly it has the stamp of antiquity. DSB is the oldest modulation system in use and consequently much capital is invested in transmission and receiving equipment. Secondly, it is the simplest system to implement, whereas more economical systems (in terms of power and bandwidth) are considerably less economical in terms of equipment cost, though viewed in the long term not unduly so. The circuit of a simple modulator is shown in figure 4. The supply potential of the transistor oscillator, and hence its output, is varied by the output of the modulation amplifier.

Demodulation, or detection, of a DSB signal is simply accomplished by means of a diode and a low-pass RC filter. The diode rectifies the modulated waveform so that only the negative half-cycles appear at its cathode. This output contains one-half of the original envelope, that is the original modulating signal. The carrier is simply removed by the low-pass filter and only the original modulation appears at the output superimposed on a d.c. potential corresponding to the amplitude of the unmodulated carrier. To increase the useful radiated power of DSB transmissions dynamic range compression is often employed. This means that the range between the loudest and softest sounds of the modulating signal is reduced, or to put it another way, pianissimo is boosted and fortissimo reduced. This means that the variation in the modulation depth is reduced.

A simple compressor is shown in figure 6. The signal at the collector of the transistor is rectified by D_1/D_2 and the potential on the capacitor at point A is applied to the base of the transistor to provide bias. If the signal through the transistor is increased the potential at point A decreases, reducing the base bias of the transistor so that the working point is shifted to a point where the gain is less. Choice of suitable time constants in the



5

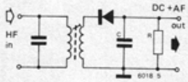


Figure 1. Block diagram of a communication system comprising an encoder, transmission medium and encoder.

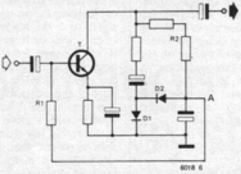
Figure 2. When electromagnetic radiation serves as the transmission medium encoding takes place in the transmitter and decoding in the receiver.

Figure 3. The frequency spectrum of a DSB signal with single frequency sinusoidal modulation.

Figure 4. DSB signals may be generated very simply as this diagram shows.

Figure 5. Simple DSB demodulator (envelope detector) using a diode as the non-linear element.

6



7

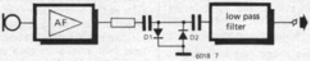


Figure 6. A dynamic compressor is simply an amplifier whose gain decreases as the input signal amplitude increases.

Figure 7. Block diagram of an AF clipper using a pair of diodes.

Figure 8. A symmetrical balanced modulator for the production of DSSC signals using an IC. The components in the IC are shown in the shaded portion of the diagram.

Figure 9. A simple product detector which may be used where good rejection of the input signals is not essential, or where this is done elsewhere in the circuit.

Figure 10. A 'universal' demodulator which will demodulate all existing narrow-band modulated signals both AM and FM.

Figure 11. Generation of an SSB signal by the filter method.

rectifier circuitry means that the distortion unavoidable with compressors is reduced to a minimum.

Where a communication system is to be used for speech only, intelligibility is more important than fidelity and large amounts of distortion may be tolerated. Modulation depth may then be controlled in a much more effective way by 'clipping'. In figure 7 the maximum output of the AF amplifier is limited to the forward voltage of the diodes, so any peaks in excess of this are clipped. After clipping the signal contains a lot of harmonics, so a low-pass filter is included to limit the bandwidth and thus make more economical use of waveband space.

Double Sideband Modulation with Suppressed Carrier (DSSC)

As the name suggests, with this type of modulation only the information-carrying sidebands are radiated. The waveform produced with a sinusoidal modulating signal using this type of modulation is shown in photograph 2. Suppressing the carrier obviously saves a great deal of transmitter power, but it is apparent from the photograph that the envelope of the resulting waveform is not the original modulating signal, which makes detection more difficult. Generation of DSSC signals is accomplished fairly easily by a number of circuit arrangements, the most effective being the symmetrical balanced modulator, of which a full range is available in IC form.

The circuit of a typical balanced modulator using such an IC is given in figure 8. T_1/T_2 and T_3/T_4 form two differential pairs. The carrier is fed in through the input transformer, but in the absence of a modulating input the outputs of the two differential pairs cancel and no carrier appears at the output. T_5 and T_6 form a differential pair into which the modulating signal is fed. This causes the pairs T_1/T_2 and T_3/T_4 to deviate from the balanced condition and a signal appears at the output which is proportional to the product of the LF and RF signals.

i.e. $$V_{out} = V_1 \cdot V_2$$

Demodulation of DSSC signals is accomplished by means of a product detector whose output is the product of the two input voltages.

So that for

$$V_1 = \cos(\omega_{AF}t) \cdot \cos(\omega_{RF}t)$$

(the DSSC signal)

and

$$V_2 = \cos(\omega_{RF}t+\varphi)$$

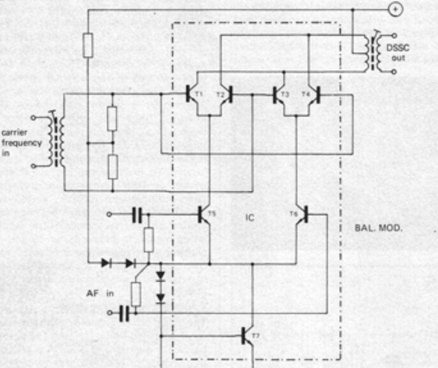
(the regenerated carrier)

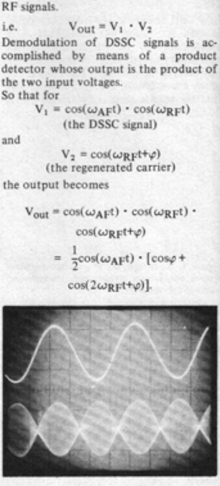
the output becomes

$$V_{out} = \cos(\omega_{AF}t) \cdot \cos(\omega_{RF}t) \cdot \cos(\omega_{RF}t+\varphi)$$

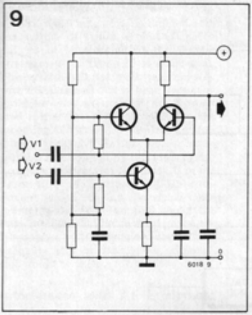
$$= \frac{1}{2}\cos(\omega_{AF}t) \cdot [\cos\varphi + \cos(2\omega_{RF}t+\varphi)].$$

8

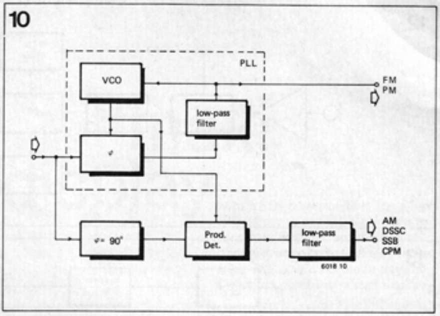




9



10



A simple low-pass filter will then remove the high-frequency component leaving the LF signal output:

$$V = \frac{1}{2}\cos(\omega_{AF}t) \cdot \cos\varphi \qquad (3)$$

The restored carrier required for demodulation can be derived from the sidebands but the practical difficulties are considerable. For this reason a small fraction of the original carrier is radiated as a pilot frequency to facilitate regeneration of the carrier at the receiving end. In the receiver this so-called residual or vestigial carrier has its level raised to the value required for demodulation. A phase-locked loop system is used because of the stringent phase criteria which have to be met. For instance, looking at equation (3) it is apparent that if the restored carrier is shifted in phase by 90° from the original carrier the LF output will be zero. Although it is easy to achieve the correct phase with several systems, a PLL system is one of the few which will maintain a phase relationship with time and temperature changes.

For the product detector the type of IC used for a balanced modulator may also be used. As the high-frequency component is quite easy to suppress at the output less complex circuits are generally used for demodulation. The more simple circuits are, however, prone to fading and RF interference.

In the Elektor laboratories a number of experiments were carried out to compare the performance of various product detector circuits and it was found that IC's with circuitry similar to figure 9 gave the best results.

Figure 10 is the block diagram of a demodulator for DSSC signals. When a signal is tuned in the phase-locked loop regenerates the vestigial carrier. Since most practical PLL's operate with a 90° phase shift the signal is shifted by 90° before being fed to the product detector. The output of the product detector is fed to a low-pass filter which removes the high-frequency components. This system will also demodulate normal DSB signals and although it may seem a little over-engineered compared with a diode detector it does offer significant advantages, particularly in the presence of interference. Furthermore, it will become apparent that this system will demodulate SSB and CPM signals plus frequency and phase modulation. In fact it is a universal demodulator for all practical forms of analogue modulation.

Single-sideband Modulation with Suppressed Carrier (SSB)

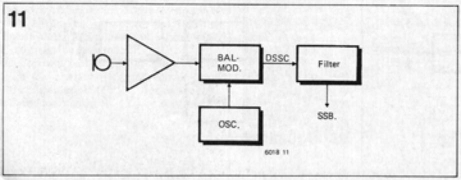
In DSSC modulation the carrier (which contains no information) is suppressed to save transmitter power, but this makes no economies in the bandwidth required to transmit the information as compared to DSB. The transmitted signal still has two sidebands above and below the carrier frequency and since each sideband contains all the LF information one of them may be discarded with a consequent halving of the required bandwidth. This is what happens with SSB, hence the name. For any given bandwidth twice as many SSB transmissions may be accommodated as compared with double-sideband transmissions.

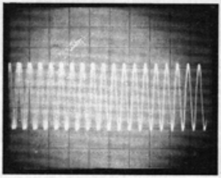
There are various ways of generating an SSB signal. The simplest way is to start with a DSSC signal and to suppress one of the sidebands by filtering so that only one sideband appears at the filter output. This method offers the choice of radiating either the upper sideband (USB), or the lower sideband (DSB) depending on the choice of filter parameters.

The filter method of generating an SSB signal is shown in figure 11. Since the two sidebands are separated only by twice the lowest frequency of the modulating signal the filter must have a sharp cutoff if adequate rejection of the unwanted sideband is to be achieved. Since filter slopes are quoted in terms of dB/octave (an octave above or below a frequency is twice and half that frequency respectively) it follows that the lower the carrier frequency the further apart are the sidebands in terms of octaves, and the easier it is to filter out the unwanted sideband. For this reason the signal is often modulated onto a carrier frequency much less than the transmitter frequency and after filtering out the unwanted sideband the frequency is raised by frequency conversion to the frequency to be transmitted. Carrier suppression is also easier at low frequencies. Currently available ceramic filters can give up to 50dB rejection of the unwanted sideband.

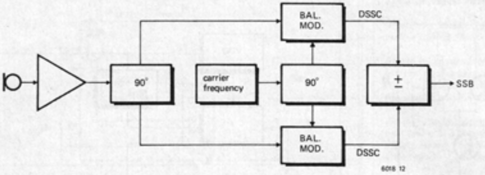
A second method of generating SSB

11

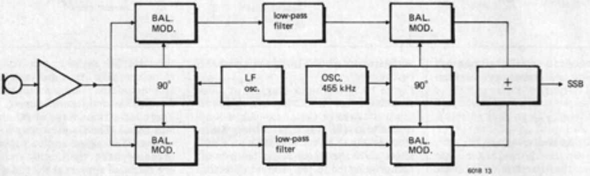




12



13



signals is shown in figure 12, but is less common. In this arrangement the LF signal is split into two components with equal amplitude but with a 90° phase shift with respect to one another and the carrier is dealt with in a similar fashion. The LF and RF signals are then fed to two balanced modulators. The two DSSC signals so produced are displaced in phase so that if a sideband of one signal is in phase with the corresponding sideband of the other signal then the other two sidebands will be 180° out of phase. Adding the two DSSC signals will therefore cancel one sideband, and subtracting them will cancel the other, so the desired sideband may easily be selected. Since accurate wide-band phase-shifters are often difficult to realise in practice a third method of producing an SSB signal is shown in figure 13. This is a two-stage modulation procedure. The LF signal is modulated onto two sub-carriers displaced in phase by 90°. The upper sidebands are rejected by the filters and the two signals are then processed as were the LF signals in figure 12.

Figure 12. The phase method of producing an SSB signal avoids the use of steep-slope filters.

Figure 13. A 'double modulation' system for the generation of SSB signals. Modulation of the AF signals onto a sub-carrier avoids the use of wide-band phase shifters.

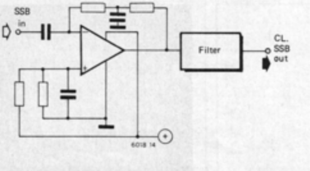
Figure 14. An RF limiter 'speech-processor' which increases the average radiated power.

Figure 15. Demodulation of an SSB signal with a beat-frequency oscillator (BFO) and product detector.

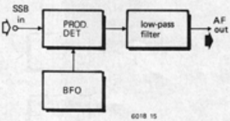
Speech Processing

The envelope of an SSB signal bears no resemblance to the original modulating waveform and attempts to raise the average level of the transmitted signal by low-frequency processing such as compressors or speech clippers are doomed to failure. The most effective way of raising the average level of an SSB transmission is by limiting of the RF signal itself. This causes harmonics, widening the frequency spectrum so that the limiter must be followed by a filter to remove them. If the SSB signal is produced by the filter method then a similar filter may be used for the removal of the harmonics after clipping. Figure 14 shows the block diagram of a typical RF clipper. These devices are available commercially in various forms under the name 'Speech Processor'.

14



15



Demodulation

An SSB signal may be demodulated with the system of figure 10, but if speech only is to be transmitted the simpler arrangement of figure 15 may be used. This system will not work satisfactorily with DSSC signals due to beat frequency problems. For example, if the regenerated carrier is not synchronised to the incoming carrier but is displaced in frequency by say 100 Hz then the sidebands will also be displaced by this amount but in opposite directions. This results in the production of a strong 200 Hz component which renders the signal quite unintelligible. With an SSB signal the only result would be to displace all the frequencies by 100 Hz and speech would probably still be intelligible, although this would be useless for music.

Interference with AF Equipment

All AM transmissions possess one common characteristic, that there is a correlation between the amplitude of the radiated signal and the amplitude of the modulating signal. Feeding such signals through an envelope detector will therefore result in an AF output, though with DSSC and SSB this will be unintelligible. In principle any non-linear element will function as an envelope detector provided the amplitude of the RF signal is sufficient. For this reason interference with domestic electronic equipment such as television and Hi-fi equipment can be a problem. Any of the semiconductor junctions in such equipment (and even dry joints, dirty plugs and the like!) could demodulate an unwanted RF signal although the most frequent cause of trouble is in the input stages of Hi-fi amplifiers. Radio amateurs are often unjustly blamed for such interference, but their equipment generally complies with the regulations and the fault is in the design of the equipment in which the interference is occurring.

Constant Amplitude Systems

Transmission systems in which the amplitude of the carrier remains constant rarely give rise to interference in domestic equipment. One possible exception is where an audio amplifier is blocked completely by a strong RF signal, but this occurrence is rare. Such systems are not necessarily more wasteful of transmitter power, since when used for speech voice-operated switches may be used so that the transmitter operates only when an LF signal is present. Another advantage of constant-amplitude systems is that automatic gain control (AGC) is much easier to include and indeed may sometimes be omitted altogether.

The second part of this article will deal with the characteristics of carrier position modulation (CPM), frequency and phase modulation. (To be continued)

M