REPORT ITU-R SM.2021*

PRODUCTION AND MITIGATION OF INTERMODULATION PRODUCTS IN THE TRANSMITTER

(2000)

TABLE OF CONTENTS

_____

* Radiocommunication Study Group 1 made editorial amendments to this Report the year 2017 in accordance with Resolution ITU-R 1.

1   Introduction

There are various types of intermodulation that can be found. In radio systems, these are manifested in a number of ways and defined as the following five types in Rec. ITU-R SM.1446:

Type 1:   Single channel intermodulation: where the wanted signal is distorted by virtue of non-linearities in the transmitter.

Type 2:   Multichannel intermodulation: where the wanted signals of multi channels are distorted by virtue of non-linearities in the same transmitter.

Type 3:   Inter transmitter intermodulation: where one or more transmitters on a site intermodulate, either within the transmitters themselves or within a non-linear component on site to produce intermodulation products.

Type 4:   Intermodulation due to active antennas: the multicarrier operating mode of an active antenna, along with the non-linearity of amplifiers, originates spurious emissions under the form of intermodulation signals.

Type 5:   Intermodulation due to passive circuits: where transmitters share the same radiating element and intermodulation occurs due to non-linearities of passive circuits.

The generation and mitigation of these intermodulation products are described in the following sections in more detail. Some examples of intermodulation products generated at radio sites are given. Measurement techniques are referred in Rec. ITU-R SM.1446. A comprehensive list of useful literature is attached at the end of the Report including references for the measurement of intermodulation Types 1 to 3 [ETSI, 1997; Shahid *et al.*, 1996; Bhargava *et al.*, 1981; ITU-R Handbook on satellite communications – fixed-satellite service (Appendix 2-1, § 5); Heathman, 1989; Bond et Meyer, 1970; Shimbo, 1971; Saleh, 1982; Wassermann *et al.*, 1983; Tondryk, 1991; Kaeadar, 1986; IESS, 1996; ETSI, 1995].

Instead of intermodulation products the expression intermodulation noise is also used in order to reflect digital modulation formats.

2   Generation of intermodulation

Intermodulation has classically been a major determinant of transmitter performance for amplitude modulated services, such as single sideband (SSB) or independent sideband (ISB). Theoretically, it does not apply to any constant envelope transmission, although in practice, practical implementation limitations lead to some of such modulation techniques not providing absolutely constant envelope modulation, and thus requiring linear amplification if spectral regrowth is to be avoided.

2.1   Intermodulation products due to discrete frequencies

The following approach [Chadwick, 1986] is classical and a complete analysis for input signal which can be represented by discrete frequencies like all analogue signals in the time domain. It may be also helpful for the basic understanding of the generation of intermodulation products.

An amplifier can be characterized by a Taylor series of the generalized transfer function [Chadwick, 1986]

$$i_0 + k_1 e_{IN} + k_2 e_{IN}^2 + k_3 e_{IN}^3 + k_4 e_{IN}^4 + k_5 e_{IN}^5 + ...$$

where i_0 is the quiescent output current, k_1, k_2, etc. are coefficients and e_{IN} represents the input signal. When two sinusoidal frequencies $\omega_1 = 2\pi f_1$ and $\omega_2 = 2\pi f_2$ of the amplitude a_1 and a_2 are applied to the input of the amplifier, the input signal is:

$$e_{IN} = a_1 \cos \omega_1 t + a_2 \cos \omega_2 t$$

and the output i_{OUT} may be shown to be the sum of the DC components:

$$i_{OUT} = i_0 + \frac{k_2}{2} \left(a_1^2 + a_2^2\right) + \frac{k_4}{8} \left(3a_1^4 + 12a_1^2 + 3a_2^4\right)$$

fundamental components,:

$$+\left(k_1\,a_1\;+\;\frac{3}{4}\,k_3\,a_1^3\;+\;\frac{3}{2}\,k_3\,a_1\,a_2^2\;+\;\frac{5}{8}\,k_5\,a_1^5\;+\;\frac{15}{4}\,k_5\,a_1^3\,a_2^2\;+\;\frac{15}{8}\,k_5\,a_1\,a_2^4\right)\cos\omega_1\,t$$

$$+\left(k_1\,a_2\;+\;\frac{3}{4}\,k_3\,a_2^3\;+\;\frac{3}{2}\,k_3\,a_1^2\,a_2\;+\;\frac{5}{8}\,k_5\,a_2^5\;+\;\frac{15}{4}\,k_5\,a_1^2\,a_2^3\;+\;\frac{15}{8}\,k_5\,a_1^4\,a_2\right)\cos\omega_2\,t$$

2nd order components:

$$+\left(\frac{1}{2}\,k_2\,a_1^2\;+\;\frac{1}{2}\,k_3\,a_1^4\;+\;\frac{3}{2}\,k_4\,a_1^2\,a_2^2\right)\cos 2\omega_1\,t$$

$$+\left(\frac{1}{2}\,k_2\,a_2^2\;+\;\frac{1}{2}\,k_3\,a_2^4\;+\;\frac{3}{2}\,k_4\,a_1^2\,a_2^2\right)\cos 2\omega_2\,t$$

$$+\left(k_2\,a_1\,a_2\;+\;\frac{3}{2}\,k_4\,a_1^3\,a_2\;+\;\frac{3}{2}\,k_4\,a_1\,a_2^3\right)\cos\left(\omega_1\pm\omega_2\right)t$$

3rd order components:

$$+\left(\frac{1}{4}\,k_3\,a_1^3\;+\;\frac{5}{16}\,k_5\,a_1^5\;+\;\frac{5}{4}\,k_5\,a_1^3\,a_2^2\right)\cos 3\omega_1\,t$$

$$+\left(\frac{1}{4}\,k_3\,a_2^3\;+\;\frac{5}{16}\,k_5\,a_2^5\;+\;\frac{5}{4}\,k_5\,a_1^2\,a_2^3\right)\cos 3\omega_2\,t$$

$$+\left(\frac{3}{4}\,k_3\,a_1^2\,a_2\;+\;\frac{5}{4}\,k_5\,a_1^4\,a_2\;+\;\frac{15}{8}\,k_5\,a_1^2\,a_2^3\right)\cos\left(\omega_1\pm 2\omega_2\right)t$$

$$+\left(\frac{3}{4}\,k_3\,a_1\,a_2^2\;+\;\frac{5}{4}\,k_5\,a_1\,a_2^4\;+\;\frac{15}{8}\,k_5\,a_1^3\,a_2^2\right)\cos\left(\omega_2\pm 2\omega_1\right)t$$

4th order components:

$$+\;\frac{1}{8}\,k_4\,a_1^4\,\cos 4\omega_1\,t\;+\;\frac{1}{8}\,k_4\,a_2^4\,\cos 4\omega_2\,t$$

$$+\;\frac{1}{2}\,k_4\,a_1^3\,a_2\,\cos\left(3\omega_1\pm\omega_2\right)t\;+\;\frac{3}{4}\,k_4\,a_1^2\,a_2^2\,\cos\left(2\omega_1\pm 2\omega_2\right)t\;+\;\frac{1}{2}\,k_4\,a_1\,a_2^3\,\cos\left(\omega_1\pm 3\omega_2\right)t$$

and 5th order components:

$$+\;\frac{1}{16}\,k_5\,a_1^5\,\cos 5\omega_1\,t\;+\;\frac{1}{16}\,k_5\,a_2^5\,\cos 5\omega_2\,t$$

$$+\;\frac{5}{16}\,k_5\,a_1^4\,a_2\,\cos\left(4\omega_1\pm\omega_2\right)t\;+\;\frac{5}{8}\,k_5\,a_1^3\,a_2^2\,\cos\left(3\omega_1\pm 2\omega_2\right)t$$

$$+\;\frac{5}{8}\,k_5\,a_1^2\,a_2^3\,\cos\left(2\omega_1\pm 3\omega_2\right)t\;+\;\frac{5}{16}\,k_5\,a_1\,a_2^4\,\cos\left(\omega_1\pm 4\omega_2\right)t$$

The series may be expanded further for terms in $k_6 e_{IN}^6$ etc. if desired. The relationships between the different products are shown in Fig. 1. It can be seen from this Figure and the equations that all the even order terms produce outputs at harmonics of the input signal and that the sum and difference products are well removed in frequency far from the input signal. The odd order products, however, produce signals near the input frequencies $f_1 \pm 2f_2$ et $f_2 \pm 2f_1$. Therefore, the odd order intermodulation products cannot be removed by filtering, only by improvement in linearity.

FIGURE 1

Unscaled intermodulation (IM) products (bold lines)
related to the fundamentals (bold dashed lines)
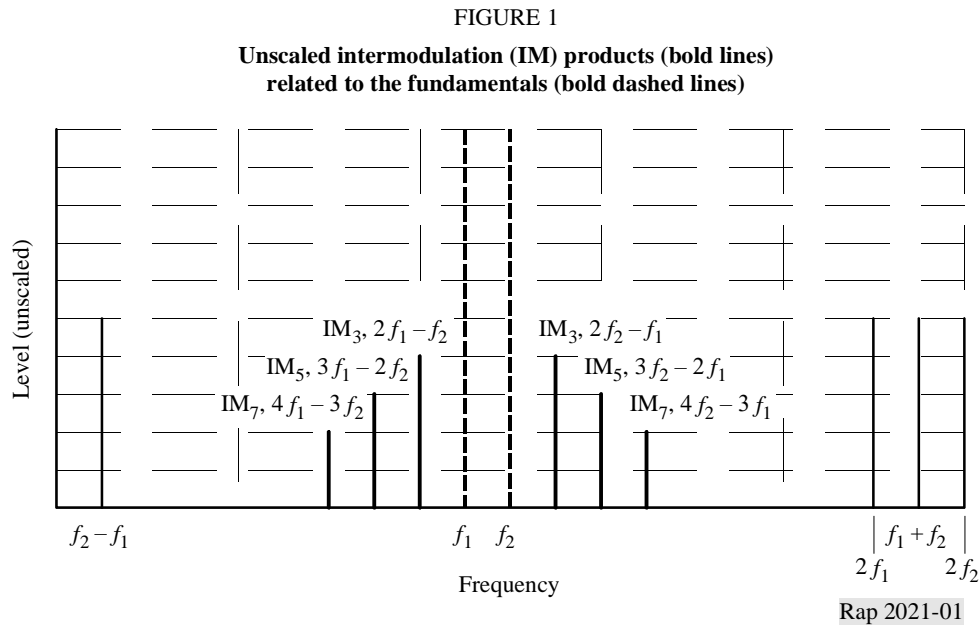
Assuming class A operation, $a_1 = a_2$ and k_4, k_5 are very small. The 3rd order intermodulation product IM_3 becomes proportional to a^3. That means that the cube of the input amplitude and the graph of the intermodulation products will have a slope of 3 in logarithmic scale while the wanted signal will have the slope of 1 (see Fig. 2). Second order products IM_2 can be similarly calculated, and the graph for these has a slope of two. The points where these graphs cross are called 3rd order intercept point IP_3 and 2nd order intercept point IP_2, respectively. IP_3 is the point where the intermodulation product is equal to the fundamental signal. This is a purely theoretical consideration, but gives a very convenient method of comparing devices. For example, a device with intermodulation products of –40 dBm at 0 dBm input power is to be compared with one having intermodulation products of –70 dBm for –10 dBm input. By reference to the intercept point, it can be seen that the two devices are equal.

As the level of the input signal increases, a point is eventually reached at which the output cannot increase, dB for dB, with the input. This is gain compression, and is important in defining the dynamic range of the device. For example, assuming an amplifier with 20 dBm intercept point and at 0 dBm input to obtain a 40 dB intermodulation ratio, but because the devices' input/output characteristics are not linear at this input level, the expected intermodulation ratio is not obtained. If, however, the compression point is a few dB higher, then the intermodulation ratio of 40 dB could be obtained. In the case of class AB operation different characteristics may occur as plotted in Fig. 2, especially at lower input signals.

FIGURE 2

**Examples of 2nd and 3rd order IM products,
gain compression in class A operation**
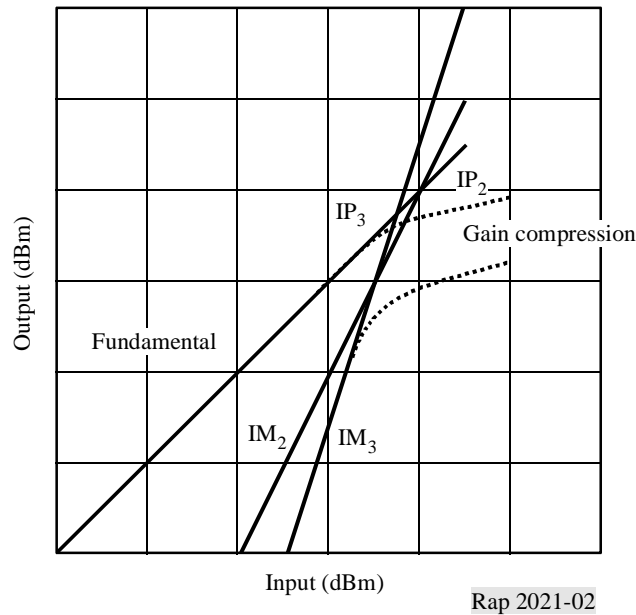
2.2 Intermodulation noise due to continuous frequency spectrum

The classical description of intermodulation at analogue radio systems deals with a two-frequency input model to a memoryless non-linear device. This non-linear characteristic can be described by a function $f(x)$, which yields the input-output relation of the element device. The function, f, is usually expanded in a Taylor-series and thus produces the harmonics and as well the linear combinations of the input frequencies. This classical model is well suited to analogue modulation schemes with dedicated frequency lines at the carrier frequencies. The system performance of analogue systems is usually measured in terms of signal-to-noise (S/N) ratio, and the distorting intermodulation signal can adequately be described by a reduction of S/N.

With digital modulation methods the situation is changed completely. Most digital modulation schemes have a continuous signal spectrum without preferred lines at the carrier frequencies. The system degradation due to intermodulation is measured in terms of bit error ratio (BER) and depends on a variety of system parameters, e.g. the special modulation scheme which is employed.

For estimation of the system performance in terms of BER a rigorous analysis of non-linear systems is required. There are two classical methods for the analysis and synthesis of non-linear systems: the first one carries out the expansion of the signal in a Volterra series [Schetzen, 1980]. The second due to Wiener uses special base functionals for the expansion.

Both methods lead to a description of the non-linear system by higher order transfer functions having n input variables depending on the order of the non-linearity. A more detailed description and two examples are given in Annex 1.

The block diagram of example 1 is shown in Fig. 3. The two data signals $x_1(t)$ and $x_2(t)$ are linearly filtered by the devices with the impulse responses $h_a(t)$ and $h_b(t)$ in adjacent frequency bands. The composite summed signal y is hereafter distorted by an imperfect square-law device which might model a transmit-amplifier. The input-output relation of the non-linear device is given by:

$$z(t) = y(t) + a y^2(t)$$

The input signals $x_1(t)$ and $x_2(t)$ are originated from a single signal $x(t)$, because of the spectral separation by the filters $h_a(t)$ and $h_b(t)$.

FIGURE 3

Signals $y_1(t)$ and $y_2(t)$ in adjacent channels subjected to an imperfect square-law device
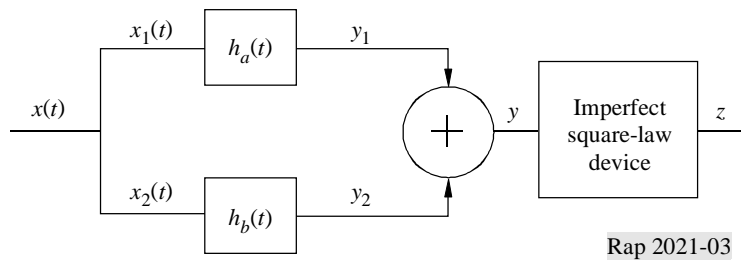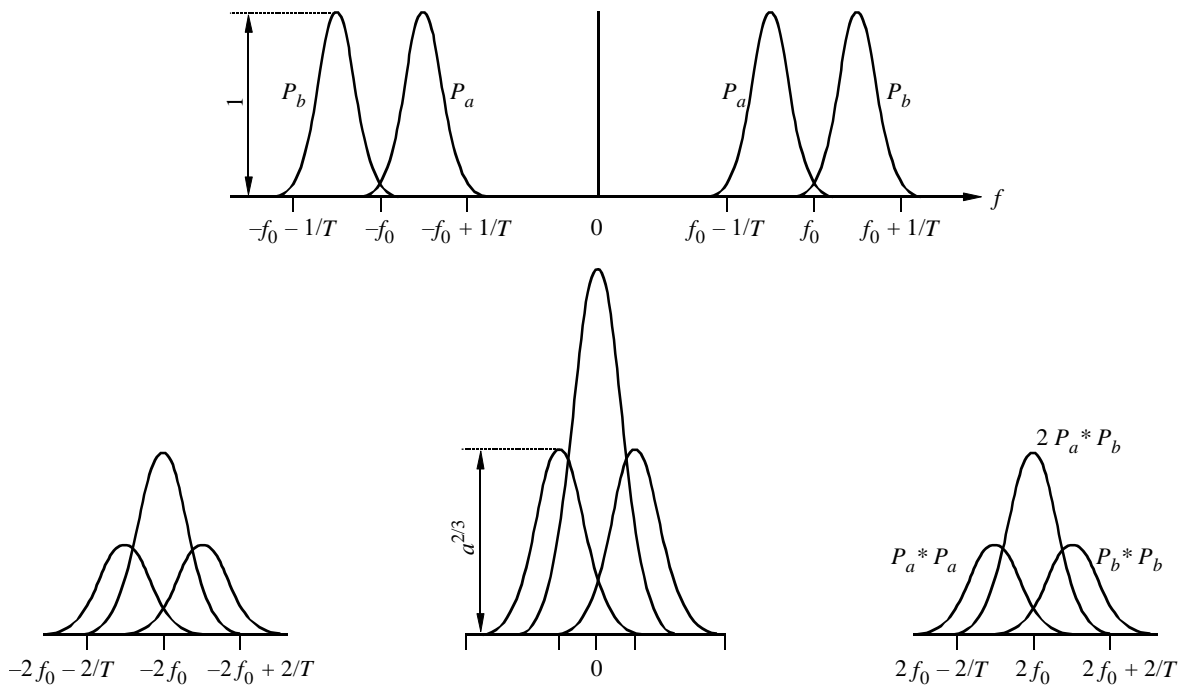

Rap 2021-03

The output signal $z(t)$ including the intermodulation noise is plotted in Fig. 4. For RF-modulated signals the intermodulation distortion in the proper frequency band is caused by non-linearities of third order. For this reason the imperfect square-law device in Fig. 3 is now replaced by an imperfect cubic device with the input-output relation:
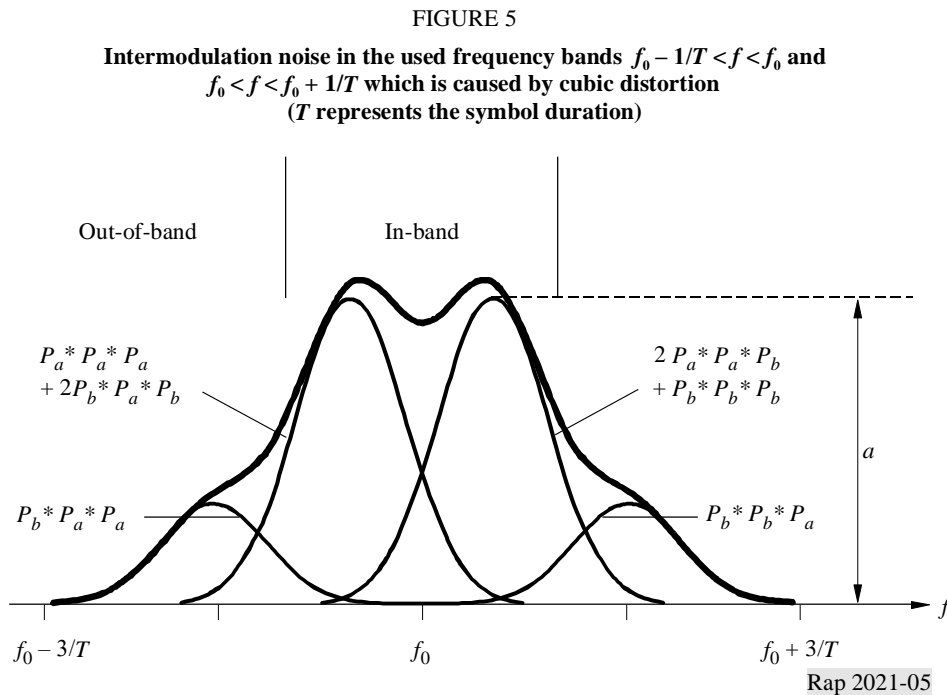
$$z(t) = y(t) + ay^3(t)$$

FIGURE 4

Upper part: power spectra of the signals $y_1(P_a)$ and $y_2(P_b)$ according to Fig. 3, lower part: spectral contributions of the intermodulating second-order term

Rap 2021-04

there are several contributions of the intermodulation noise falling into the used channels near f_0. The different parts $P_a * P_a * P_a \dots P_b * P_b * P_b$ are plotted in Fig. 5. The thick line shows the sum of the distortions.

FIGURE 5

Intermodulation noise in the used frequency bands $f_0 - 1/T < f < f_0$ and
$f_0 < f < f_0 + 1/T$ which is caused by cubic distortion
(T represents the symbol duration)

Rap 2021-05

3 Mitigation techniques

A number of techniques have been developed to reduce intermodulation in transmitter power amplifiers, and some of these are briefly described. However, they are not considered as an exhaustive list.

In certain standards [ETSI, 1994 and 1998], distinction is often made between active intermodulation due to non-linear components in the transmitters themselves and passive intermodulation. Passive intermodulation is caused, e.g., by metallic contacts in masts. Antenna hardware can be a source of problems when high field strengths occur on the site and cause them to radiate intermodulation products which disturb equipment at the same site, or at a neighbouring site. No standards exist that are able to specify reasonable limits for such effects. The intermodulation factors used to make these calculations are affected by many parameters and they also depend on electrical resonances in the components which make up the mast and antenna arrays. These products can be radiated from the site.

The overall loss, A_{CI}, between a transmitter providing the unwanted emissions giving rise to the intermodulation product is given by the sum of:

$$A_{CI} = A_C + A_I$$

where A_C is the coupling loss defined as the ratio of the power emitted from one transmitter to the power level of that emission at the output of another transmitter which may produce the unwanted intermodulation product. The intermodulation conversion loss, A_I, is the ratio of power levels of the interfering signal from an external source and the intermodulation product, both measured at the output of the transmitter (ITU-R Report M.739 – Interference due to intermodulation products in the land mobile service between 25 and 1 000 MHz).

Using this definition, mitigation of intermodulation products means increasing the overall loss A_{CI}. It is obvious that a reduction of the non-linearity, particularly of the odd-numbered orders, will improve the overall performance and increase the value of intermodulation conversion loss. Such techniques considered in § 3.1 can be used to reduce the intermodulation products of the Types 1 and 2 and can be implemented in the transmitters themselves. Possibilities for increasing the coupling loss, e.g. by increasing the spatial separation, are described in § 3.2. These mitigation measures which are known as radio site shielding engineering are applicable for intermodulation Type 3. In contrast to § 3.1 these measures are not used in the transmitters themselves.

Other mitigation measures are briefly quoted in § 3.3. Some examples of intermodulation Type 3 at the transmitter site are depicted in § 3.4 for illustration.
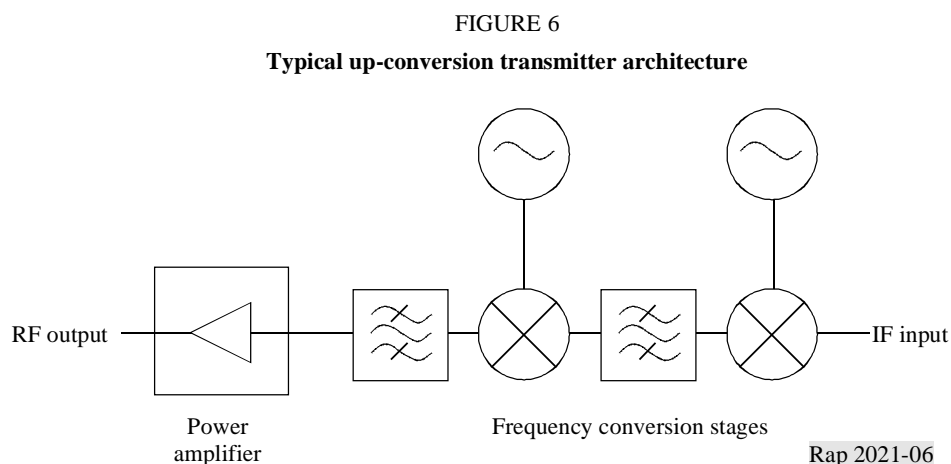
3.1 Suppression at transmitters

These intermodulation products are part of the unwanted emissions as defined in No. 1.145 of the Radio Regulations. In the following a typical transmitter architecture is introduced before discussing mitigation techniques.

3.1.1 Transmitter architecture

The RF architecture of radio transmitters often takes the form shown in the simplified block diagram of Fig. 6. The modulated input signal is generated at an IF, then frequency translated in one or more mixing and filtering stages to the final frequency.

A common problem with this arrangement is that each mixing process will produce many spurious products, as well as the main sum and difference frequency components. These arise through mixing of the local oscillator (LO) harmonics with harmonics of the IF input, often referred to as $m \times n$ products. Although the LO harmonics are unavoidable due to the modulating action of the mixer LO port, the IF harmonics can be reduced by ensuring that the IF port is operated well below compression. However, in practice, a compromise must be reached between linearity and S/N, so the spurii can never be completely eliminated. Spurious products which fall at offsets far removed from the wanted frequency can be suppressed through filtering, but those close to the carrier will not be attenuated.

FIGURE 6

Typical up-conversion transmitter architecture

RF output — Power amplifier — Frequency conversion stages — IF input

Rap 2021-06

One way of mitigating this problem is to generate the wanted signal directly at the final frequency using a vector modulator, as shown in Fig. 7. In this case, in-phase and quadrature (I and Q) baseband signals are used to directly modulate a carrier at the output frequency. Although spectral spreading of the signal into the adjacent channels can still occur, the harmonic mixing effect is eliminated, since there is only a single carrier component applied to the mixers.

FIGURE 7

Vector modulator transmitter architecture

Rap 2021-07
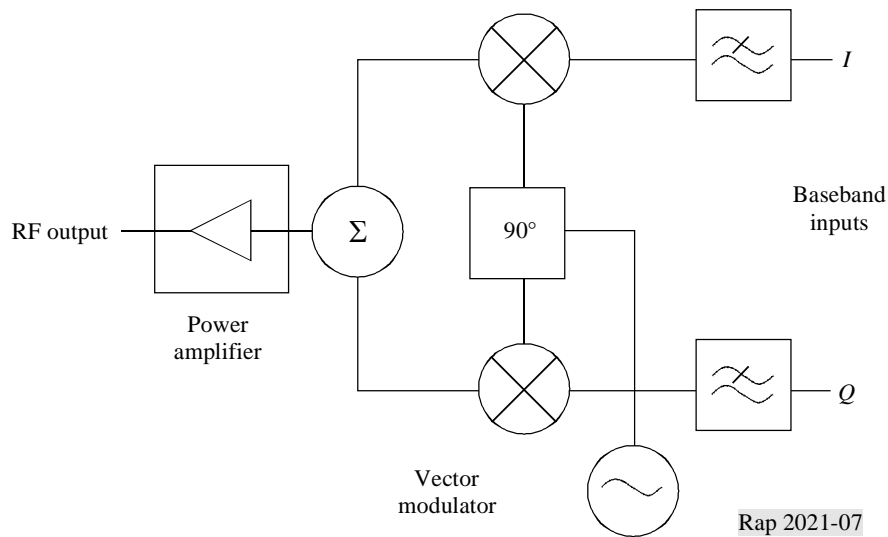
A drawback with this arrangement is that there will be a finite carrier leakage to the output, typically suppressed by about 30 dB relative to the wanted signal. Usually this is of no consequence, but in cases where better carrier suppression is required, it is necessary to adjust the d.c. bias on the I and Q inputs to null the carrier.

While the arrangement illustrated in Fig. 7 utilizes two bi-phase AM modulators, it is equally possible to use four uniphase modulators, and four orthogonal channels.

A more complex, but more flexible, approach is to use a single channel incorporating a digitally controlled attenuator, and a digitally controlled phase-shifter. These two components are driven by the baseband input by way of a look-up table, allowing the direct generation of virtually any (digital) modulation scheme. It might be noted that the carrier-frequency amplitude/phase shifter is the same component required for use in active antenna beam-forming arrays.

3.1.2 Filtering

Filtering (generally bandpass filtering) of the transmitter output can be used in conjunction with the other techniques discussed in this Report to reduce the residual spurious output levels. The choice of the type of filter to be used is, as usual, a compromise between a number of interacting, usually conflicting, requirements such as out-of-band rejection, bandpass attenuation, time domain response, size, weight, cost, etc.

Filter designs are usually based on the classical analytically derived categories such as Butterworth, Chebyshev, etc. Some of these categories are optimized for one of its characteristics at the expense of others, and some provide compromizes between characteristics as in Table 1:

TABLE 1

Category	Optimized parameter	Sacrificed parameter
Butterworth	Bandpass amplitude flatness	Out-of-band rejection
Chebyshev	Out-of-band rejection	Bandpass amplitude flatness and attenuation
Bessel	Bandpass delay flatness	Out-of-band rejection
Elliptic (Cauer)	Close-in out-of-band rejection (theoretically infinite at spot frequencies)	Out-of-band rejection away from spot frequencies

Other categories provide compromises between characteristics. For example, the so-called linear phase filter can be designed to provide a bandpass flatness approaching that of the Bessel filter, but with improved out-of-band rejection. Similarly, transitional filters have a near linear phase shift and smooth amplitude roll-off in the bandpass, with improved out-of-band rejection compared to a Bessel filter (but still significantly less than a Chebyshev filter). As well as the characteristics described above, another factor which defines the performance of any filter is its order of complexity, which is related to the number of poles and/or zeros in its transfer function. In general, increasing the order of complexity improves the performance of the optimized characteristic at the expense of degrading the performance of the sacrificed characteristics.

Figure 8 shows examples of the out-of-band rejection (which is the main performance parameter of interest in the context of this study) for Butterworth, Chebyshev and Elliptic filters of order of complexity $n = 3$. Note that the low-pass response is shown; in a practical design the band-pass response would be derived from this by suitable scaling of the frequency axis. The Figure therefore illustrates the relative performance of these filter types.

FIGURE 8

Comparison of Butterworth, Chebyshev and Elliptic filters, $n = 3$
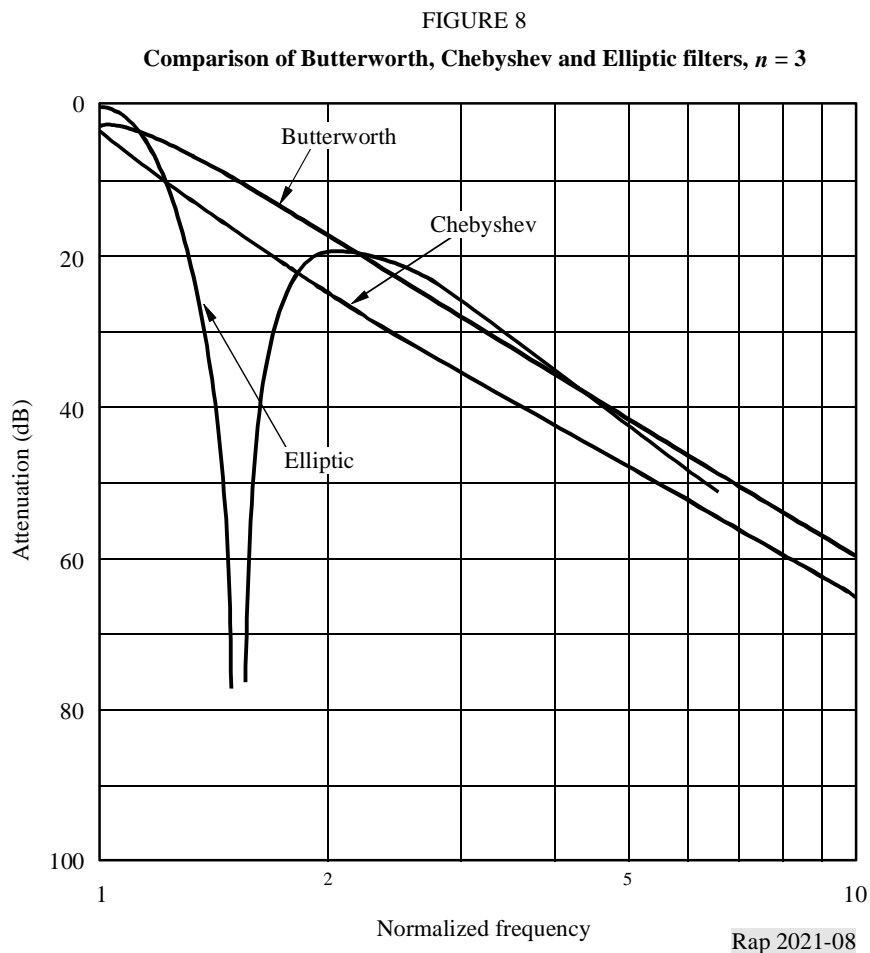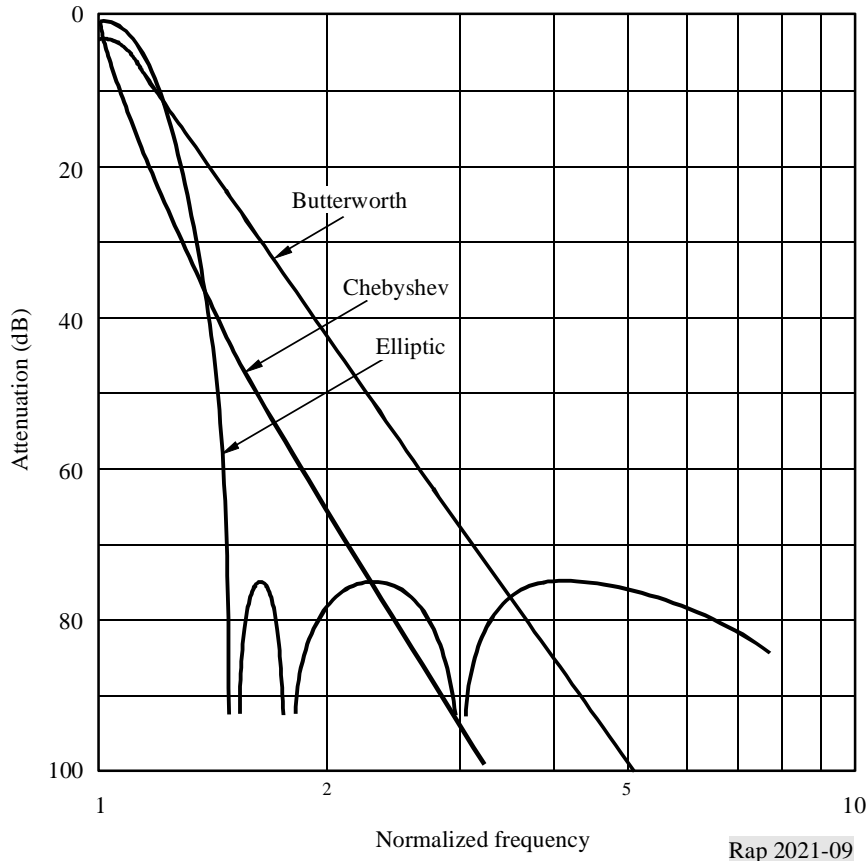

Rap 2021-08

Figure 9 shows examples of the out-of-band rejection for similar filters of order of complexity $n = 7$. The improved performance of these filters compared with those in Fig. 8 can only be obtained at the expense of increased implementation complexity and, in practice, increased insertion loss in the wanted frequency band.

FIGURE 9

Comparison of Butterworth, Chebyshev and Elliptic filters, *n* = 7

Rap 2021-09

Transmitter output filtering nearly always requires the use of resonant elements such as tuned circuits or transmission lines to form filter structures. Although surface acoustic wave (SAW) filters have been produced for operation at up to 2 GHz, these have relatively low power handling. The insertion loss of SAW filters also tends to be quite high, up to 6 dB for SAW resonator filters, and up to 30 dB for transversal (delay line) filters.

At frequencies up to a few hundred MHz, LC (inductor capacitor) filters are usually used to achieve bandwidths of 10% or more. Narrower bandwidths are possible, but the unloaded Q, tolerances and temperature stability of the components generally preclude significant further reduction.

At higher frequencies, up to a few GHz, the commonest filter technologies are printed microstrip and silver plated ceramic. Microstrip filters are generally limited to bandwidths no less than a few per cent, due to tolerances of the dielectric constant, substrate thickness and etching variability. The unloaded Q of microstrip resonators (typically < 200) also limits the minimum practical bandwidth due to insertion loss considerations.

The use of silver plated ceramic technology can achieve better performance owing to the higher unloaded Q and excellent stability of the materials used. The digital cellular and cordless telephone industry in particular, has prompted the development of very high dielectric constant, low loss ceramics for use in miniature coupled resonator filters. For example, a typical 2-pole 1.9 GHz filter can achieve an insertion loss of 0.8 dB with a bandwidth of 1%.

At frequencies of several GHz and above, the resonant elements tend to be cavities or transmission lines with an air dielectric. A common configuration is the interdigital filter, where several resonant fingers are positioned within a single cavity to give the desired coupling, and hence overall filter response. Performance is comparable with that of silver plated ceramic filters, with bandwidths available as low as 0.2%.

3.1.3   Linearization

Linearization of a transmitter system may be accomplished by a number of methods:

–   *Feedforward linearization:* This technique compares the amplified signal with an appropriately delayed version of the input signal and derives a difference signal, representing the amplifier distortions. This difference signal is in turn amplified, and subtracted from the final HPA output. The main drawback of the method is the requirement for a 2nd amplifier – the technique can, however, deliver an increase in output power of some 3 dB when used with a TWT.

–   *Feedback linearization:* In audio amplifiers, linearization may readily be achieved by the use of feedback, but this is less straightforward at high RF frequencies due to limitations in the available open-loop amplifier gain. It is possible, however, to feedback a demodulated form of the output, to generate adaptive pre-distortion in the modulator. It is clearly not possible to apply such an approach in a bent-pipe transponder, however, where the modulator and HPA are rather widely separated.

–   *Predistortion:* Rather than using a method that responds to the actual instantaneous characteristics of the HPA, it is common to pre-distort the input signal to the amplifier, based on *a priori* knowledge of the transfer function. Such pre-distortion may be implemented at RF, IF or at baseband. Baseband linearizers, often based on the use of look-up tables held in firmware memory are becoming more common with the ready availability of VLSI techniques, and can offer a compact solution. Until recently, however, it has been easier to generate the appropriate pre-distortion function with RF or IF circuitry.

RF amplifier linearization techniques can be broadly divided into two main categories:

–   Open-loop techniques, which have the advantage of being unconditionally stable, but have the drawback of being unable to compensate for changes in the amplifier characteristics.

–   Closed-loop techniques, which are inherently self-adapting to changes in the amplifier, but can suffer from stability problems.

3.1.3.1   Predistortion

This involves placing a compensating non-linearity into the signal path, ahead of the amplifier to be linearized, as shown in Fig. 10. The signal is thus predistorted before being applied to the amplifier. If the predistorter has a non-linearity which is the exact inverse of the amplifier non-linearity, then the distortion introduced by the amplifier will exactly cancel the predistortion, leaving a distortionless output.

FIGURE 10

Predistortion concept

Predistorter   Power amplifier

Rap 2021-10
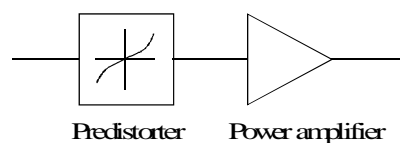
In its simplest analogue implementation, a practical predistorter can be a network of resistors and non-linear elements such as diodes or transistors. Several examples of this technique have appeared in the literature, e.g. [Gray, 1980], where the reduction in 3rd order intermodulation distortion that has been reported is typically in the range 7 to 15 dB. The poor performance is due to the fact that the amplifier characteristics are not constant, but vary with time, frequency, power level, supply voltage and environmental conditions.

Better results have been reported in [Aghvami and Robertson, 1993], where a pair of FET amplifiers is used as the predistorter, as shown in Fig. 11. In this arrangement, the input signal is unequally split between the two amplifiers, such that one of them is driven into compression. The compressed output is then scaled and subtracted from the linear output to produce the inverse of the compression characteristic, as required. Reduction in intermodulation distortion of around 20 dB has been measured using this technique, but only when the main amplifier is operated with at least 1 dB of back-off.

FIGURE 11

Soft-limiter predistortion

Rap 2021-11
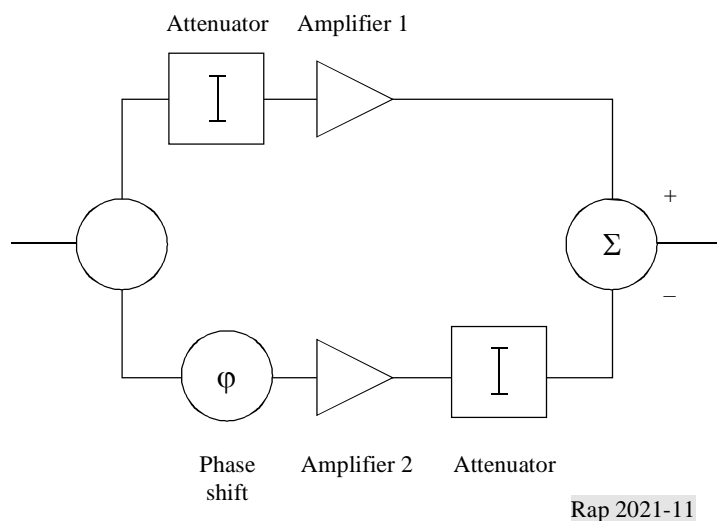
Although adaptive predistortion schemes have been reported, where the non-linearity is implemented in digital signal processing (DSP), they tend to be very computationally or memory intensive, and power hungry.

3.1.3.2   Feedforward

This technique is widely used in cellular base stations where a number of transmitter outputs are amplified in common providing intermodulation cancellations greater than 30 dB in 20 MHz bandwidth.

3.1.3.2.1   Theory

Feedforward [Black, 1928] is a distortion cancellation technique for power amplifiers. Figure 12 shows its basic configuration.

The error signal generated in the power amplifier is obtained by summing the loosely coupled signal and a delayed inverted input signal, so that the input signal component is cancelled. This circuit is called the signal cancelling loop.

The error signal is amplified by an auxiliary amplifier, and is then subtracted from the delayed output signal of the power amplifier, so that the distortion at the output is cancelled. This circuit is called the error cancelling loop.

FIGURE 12

Feedforward

Power amplifier

Input

Delay

Output

+ Σ

−

Signal
cancelling
loop

Error
cancelling
loop

+

Delay   − Σ

Auxiliary amplifier

Rap 2021-12
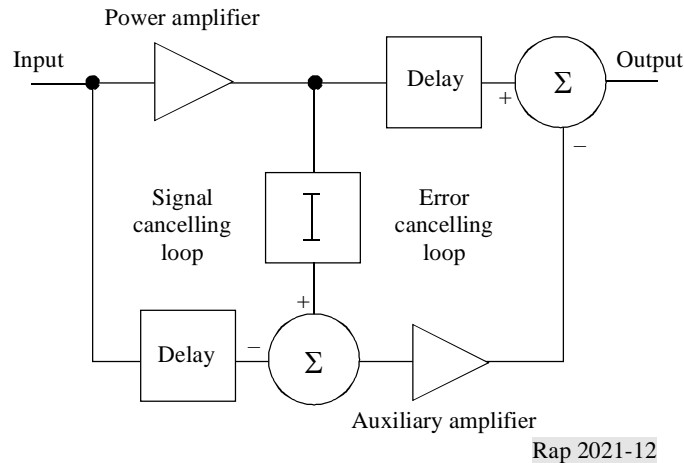
3.1.3.2.2 Accuracy necessary for each cancellation loop and expected cancellation

Signal cancelling loop

It is necessary to attenuate the input signal component lower than the error signal at the input of the auxiliary amplifier, so that the residual main signal does not cause overloading of the auxiliary amplifier, or does not cancel the main signal itself at the equipment output. For example, when the intermodulation ratio of the power amplifier is about −30 dB, the loop accuracy of the order of 0.2 dB in the amplitude and 1.5° in the phase are required in order to satisfy the above conditions.

Error cancelling loop

For example, to obtain 35 dB cancellation requires a loop accuracy of the order of 0.1 dB in the amplitude and 1° in the phase are required.In order to secure the above accuracy, each cancelling loop is usually automatically controlled against the variation of the gain, the phase of the amplifier, etc. Therefore, deep cancellation can be achieved over a limited frequency band. However, due to non-flatness of the frequency response of the circuits for the amplitude or the phase in the frequency band concerned, these cancellations are restricted in general according to the relative bandwidth or other factors.

3.1.3.2.3 Auxiliary amplifier

Output power level of the auxiliary amplifier is less than 10% of the equipment output power in general. For example, when the intermodulation ratio of the power amplifier is −30 dB and insertion loss of the auxiliary amplifier output is 16 dB, it is −30 + 16 = −14 dB (4%) of the equipment output power.

It is necessary to have a higher intermodulation ratio for the auxiliary amplifier at this power level than the inverse of its expected distortion cancellation (dB). For example, to achieve 35 dB cancellation, an intermodulation ratio of about −40 dB is required.

3.1.3.2.4 Summary

More than 30 dB of distortion cancellation is possible within a certain bandwidth. Some design trade-offs include:

– increase of the overall power consumption of the order of 30~40% is unavoidable, due to the loss by the delay circuit and by others in the error cancellation loop, and due to power consumption of the auxiliary amplifier;

– circuit configuration is rather complicated.

The above features mean that the feedforward technique is suitable for power amplifiers which need high cancellation of the order of 20~40 dB, accepting some increase of the overall power consumption. (It will not be suitable to apply it to high power amplifiers like in the HF transmitters whose power consumption is an important factor and required cancellation of the distortion is not so high as shown in [Bennet and Clements, 1974].)

3.1.3.3 Feedback

Negative feedback [Black, 1937] is the most well known linearization technique and is widely used in low frequency amplifiers, where stability of the feedback loop is easy to maintain. With multi-stage RF amplifiers however, it is usually only possible to apply a few dB of overall feedback before stability problems become intractable [Mitchell, 1979]. This is mainly due to the fact that, whereas at low frequency it can be ensured that the open-loop amplifier has a dominant pole in its frequency response (guaranteeing stability), this is not feasible with RF amplifiers because their individual stages generally have similar bandwidths.

Of course, local feedback applied to a single RF stage is often used, but since the distortion reduction is equal to the gain reduction, the improvement obtained is necessarily small because there is rarely a large excess of open loop gain available.

3.1.3.4 Modulation feedback

At a given centre frequency, a signal may be completely defined by its amplitude and phase modulation. Modulation feedback exploits this fact by applying negative feedback to the modulation of the signal, rather than to the signal itself. Since the modulation can be represented by baseband signals, we can successfully apply very large amounts of feedback to these signals without the stability problems that beset direct RF feedback.

Early applications of modulation feedback used amplitude (or envelope) feedback only, applied to valve amplifiers [Arthanayake and Wood, 1971], where amplitude distortion is the dominant form of non-linearity. With solid-state amplifiers however, phase distortion is highly significant and must be corrected in addition to the amplitude errors. The first successful practical implementation of simultaneous amplitude and phase feedback was demonstrated by Petrovic and Gosling [1979], and is known as the Polar Loop technique.

3.1.3.5 The Polar Loop technique

The Polar loop technique is based around the principle of envelope elimination and restoration (EER), first proposed in 1952 by Kahn [1952], but modified to allow feedback to be applied. A block diagram of the Polar Loop technique is shown in Fig. 13.

FIGURE 13

The Polar Loop technique
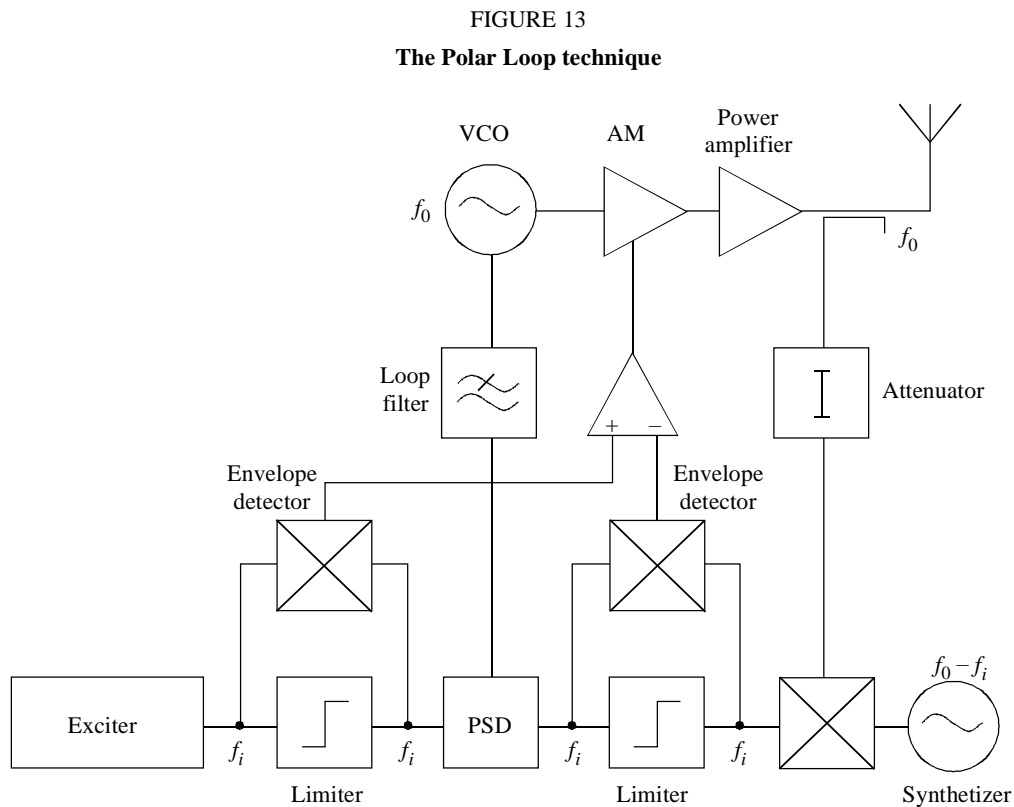
VCO: voltage controlled oscillator
PSD: phase sensitive detector

Rap 2021-13

The RF stages of the system are particularly simple. They consist of a VCO running at the output frequency, which generates the phase component of the output signal, an amplitude modulated stage which generates the amplitude component, and the main power amplifier.

The input signal to the Polar Loop is first generated at intermediate frequency, f_i, and at low power level (shown as the exciter block in the diagram). It is then resolved into polar coordinate form by envelope detection to produce the amplitude component, and hard-limiting to give the phase component. The envelope detection is conveniently achieved by multiplying the input signal by the limiter output in a double balanced mixer (a process equivalent to full-wave rectification). A sample of the final RF output is translated (usually down) to the same frequency as the input signal, and is similarly resolved into its polar coordinates. The two envelope signals are then compared in a high gain differential amplifier which in turn controls the AM, forming an envelope feedback system. The two phase modulated signals are phase compared in a PSD, and the amplified error signal controls the VCO forming a phase locked loop (PLL). The overall effect is that two orthogonal feedback loops are formed, which by suitable choice of loop gain and bandwidth, attempt to make the amplitude and phase of the output signal closely approach that of the IF input.

The two main limiting factors in the performance of a Polar Loop system are:

–   The balance between the two polar resolving circuits (limiters + mixers).

–   The relative bandwidths of the feedback loops and the amplitude and phase spectra (which determines the amount of negative feedback available).

In practical Polar Loop transmitters designed for narrow-band (5 kHz) applications, it has been found that the balance of the resolving circuits is the main problem, and this sets a minimum value to the residual 3rd order intermodulation distortion of around –60 dBc [Smith, 1986]. For wider bandwidth signals, it is the finite amount of feedback which is the main restriction. This is particularly true for signals where the envelope can fall to zero, as the zero-crossing often results in a sharp discontinuity in both the envelope and phase waveforms, and consequently produces envelope and phase spectra which are considerably wider than the composite signal bandwidth.

An alternative approach to modulation feedback, which overcomes both of the above problems, is known as the Cartesian Loop technique invented by Smith and Petrovic [1982] and this is covered in the next section.

### 3.1.3.6   The Cartesian Loop technique

The Cartesian Loop technique makes use of the fact that a modulated RF signal can be represented in complex (*I* and *Q*) baseband form as well as by amplitude and phase functions.

If negative feedback is applied to *I* and *Q* rather than *A* and *f*, this leads to the configuration shown in Fig. 14.

The principal of operation is as follows:

Complex baseband signals, I_{mod} and Q_{mod}, are used to modulate in-phase and quadrature local oscillator signals in double balanced mixers, and the combined output forms the input to the driver and power amplifier. A sample of the PA output is fed to a 2nd pair of mixers configured as demodulators which use the same local oscillators. The RF output is thus coherently demodulated back down to *I* and *Q* baseband. These signals, I_{fb} and Q_{fb}, are then feedbacked and compared with the input signals, I_{in} and Q_{in}, in high gain differential amplifiers, the outputs of which form the inputs to the modulators, I_{mod} and Q_{mod}. Just as in the Polar Loop, two orthogonal feedback loops are thus formed which attempt to make the *I* and *Q* demodulated outputs closely approach the *I* and *Q* inputs. Note that because of the coherent nature of the feedback, the technique is identically equivalent to RF feedback, but because dominant loop poles are introduced by the differential amplifiers, a good phase margin of stability may be easily maintained, even when very large amounts of feedback are applied.

FIGURE 14

The Cartesian Loop technique
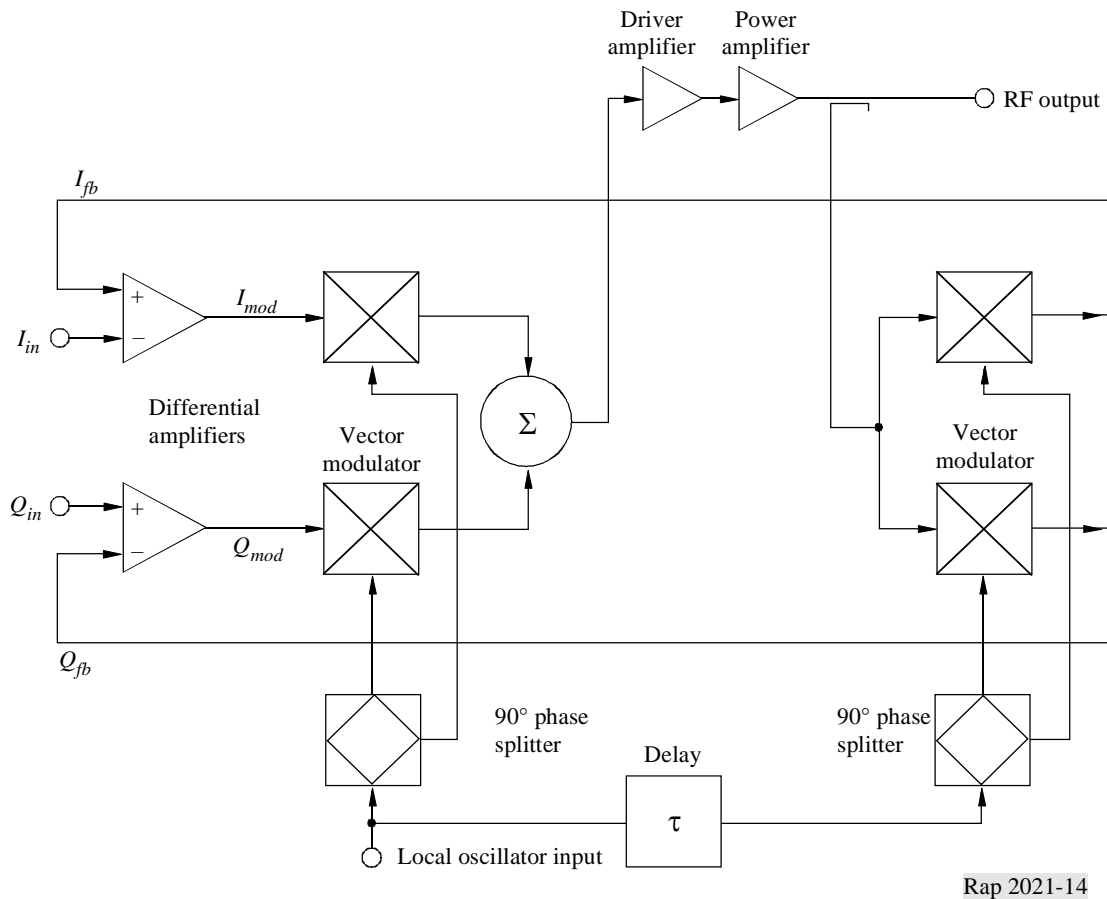
Rap 2021-14

The delay element shown in the diagram is to ensure that the RF output and the demodulating carriers are at the correct relative phase. Perfect alignment is not necessary owing to the compensating action of the loops.

The effectiveness of the Cartesian Loop depends on two factors:

– The ratio of the feedback loop bandwidths to the *I* and *Q* input bandwidths (determines the amount of feedback).

– Linearity of the demodulators (since the *I* and *Q* demodulated outputs must be a linear representation of the RF output).

Note that unlike the Polar Loop, the RF output bandwidth is simply twice the *I* and *Q* bandwidth. We do not have the problem of generating wideband *A* and *f* signals.

Practical Cartesian Loop transmitters have been constructed which operate at up to 900 MHz [Cole, 1989] for relatively narrow band signals (< 5 kHz bandwidth) and these have achieved excellent results. On a two-tone test, 3rd order intermodulation products are typically reduced by 40 dB, compared to the same transmitter with the power amplifier run open loop.

3.2 Site-shielding for inter transmitter intermodulation

The reduction of the intermodulation products generated in the output stages of different transmitters in a communal site is best achieved by minimising the coupling rather than attempting to reduce the level at the mixing junction. The latter can only be achieved by making the output stage more linear and this reduces efficiency.

Antennas are an important factor for radio site engineering to reduce intermodulation Type 3. Antennas are specified by the gain, voltage standing wave ratio (VSWR), radiation pattern, balance ratio, tolerable input power bandwidth and intermodulation performance. The following values [Petrovic and Gosling, 1979] for intermodulation performances are desirable:

– for single frequency transmit and receiving applications: −100 dBc

– for multiple frequency transmission: −130 dBc

– for multiple frequency transmission and reception: −143 dBc

Similar considerations hold for other non-linear components like circulator, combiner, isolators, etc., installed between output stage of the amplifier and the transmitting antenna. Examples for mobile radio are given in [RA, 1987].

3.2.1 Antenna spacing

Antenna spacing is a very appropriate measure to increase the coupling attenuation. Assuming far field condition, $d \gg \lambda$, the isolation, L, is given by the free space equation:

$$L = -27.5 + 20 \log \frac{d}{\text{m}} + 20 \log \frac{f}{\text{MHz}} - G_1\left(\varphi_1, \vartheta_1\right) - G_2\left(\varphi_2, \vartheta_2\right) \qquad \text{dB}$$

$$L = L_0 - G_1\left(\varphi_1, \vartheta_1\right) - G_2\left(\varphi_2, \vartheta_2\right) \qquad \text{dB}$$

where d is the spatial separation of the antenna (m), f is the frequency (MHz) and G_1, G_2 are the antenna gains referred to isotropic antennas and depending on incident angles. For isotropic and very small antennas, the isolation can be approximated by the basic free space loss: $L \approx L_0$. Regarding real antenna patterns in elevation and azimuth the actual isolation will be larger.

Examples for VHF/UHF broadcasting transmitters are given in Fig. 15 [Pye Telecom, 1978]. The graphs based on vertical antenna polarization indicate which attenuations are possible in the VHF/UHF bands for horizontal and vertical separation. It is shown that maximum attenuation is always easier to obtain when antennas are separated vertically.

FIGURE 15

**Isolation versus separation of vertically polarized antennas**
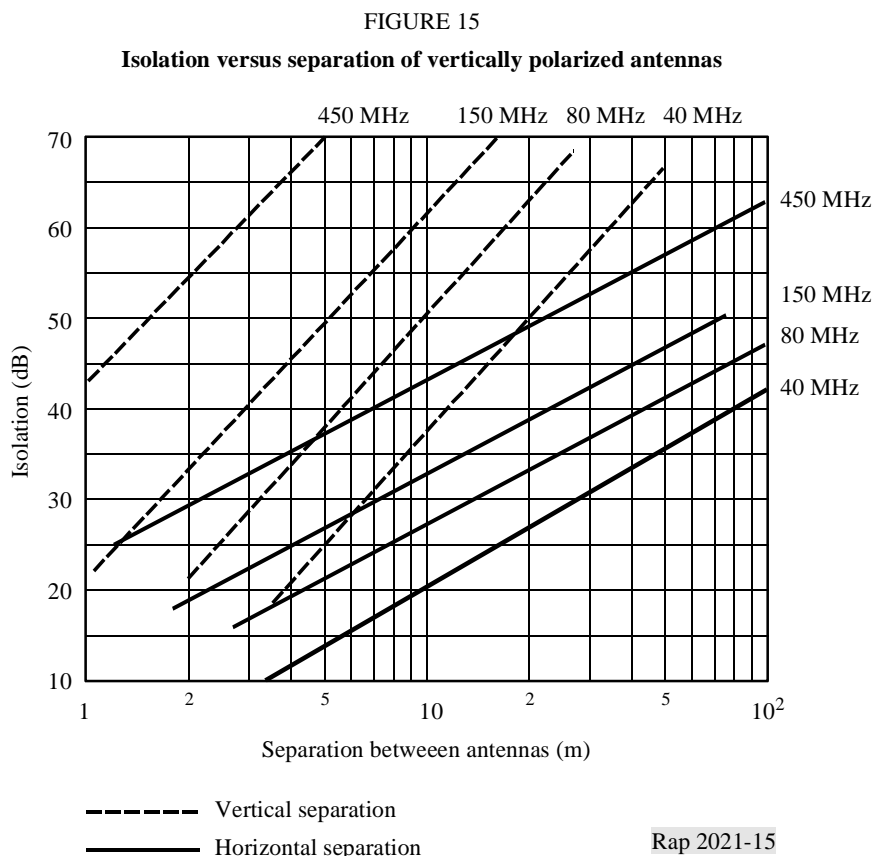
It is to consider that for lower frequencies or huge and closely located antenna arrays, the far field condition is not valid which yields to smaller values of isolation. If obstacles like buildings, the mast or even the antenna hardware itself are located in the main lobes of the antennas, reflections are likely to occur and the isolation is also reduced.

3.2.2 Antenna pattern

The isolation is also influenced by the antenna patterns deployed at the radio site, e.g., using notches in the pattern can increase significantly the isolation.

For illustration, in Fig. 16 the isolation gained by an dipole and 8 element Yagi antenna are plotted versus the normalized wavelength λ [RA, 1987]. The diagrams indicate that a minimum isolation is achieved for parallel spacing of the antennas and the isolation of an omnidirectional dipole is about 10 dB worse than for the 8 element Yagi antenna having a directional pattern. About 8 to 10 dB larger isolation is obtained by collinear installation. The difference between dipole and Yagi is about 2 dB, but the large value is gained by the dipole due to larger attenuation at 90° elevation.

FIGURE 16

Typical isolations between antennas
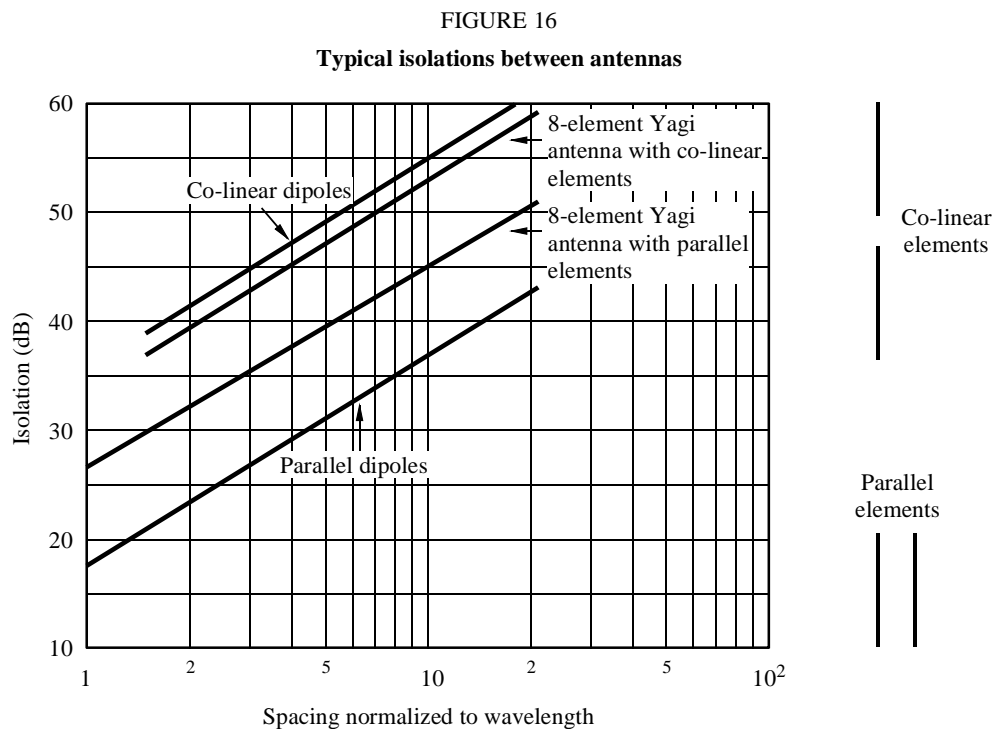
Rap 2021-16

3.3 Other mitigation measures

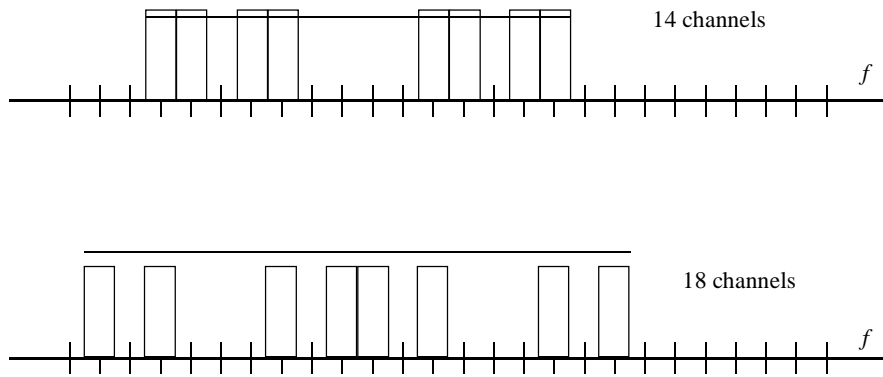### 3.3.1 Reduction of intermodulation products in receivers

Filtering should be used at the front-end of receivers to reject adjacent-band energy. Similar techniques as described for transmitters in § 3.1 may be applied.

If the input level of the wanted signal is high enough, an additional, variable RF attenuator should be inserted between the antenna feeder and receiver input. It permits the reduction of the incoming interference signal below a sufficient level.

3.3.2 Frequency arrangements

The frequencies, i.e. radio channels, to be used should be arranged so that no receiver on the interfering intermodulation frequencies is required to operate in an area where the wanted signals may produce an intermodulation product of sufficient level to disturb the service. In systems operating a number of frequency channels most cases of harmful intermodulation within the system can be alleviated by the proper choice of channel sets. Two illustrative examples for the possible frequency arrangements are depicted in Fig. 17 for low power devices [Wood, 1998].

FIGURE 17

Possible channel allocations to minimize 3rd order intermodulation products
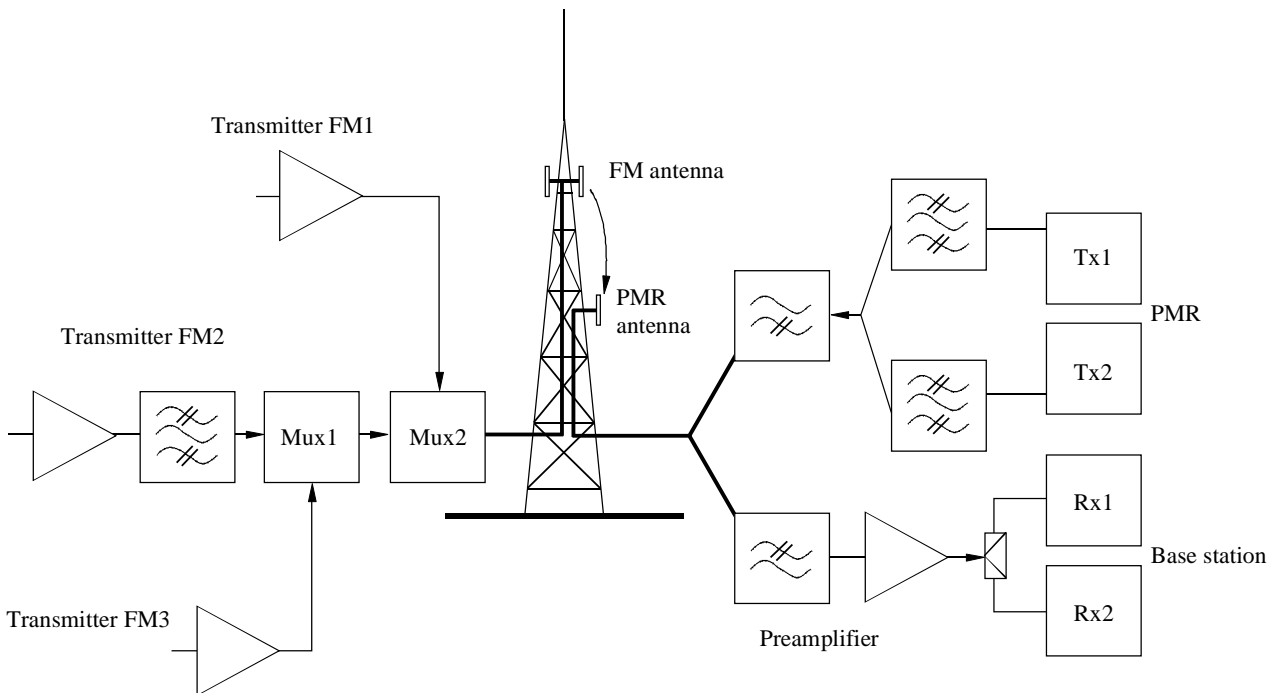

Rap 2021-17

3.4 Examples of intermodulation products generated on a radio site with FM and public mobile radio (PMR)

Figure 18 represents a simplified radio site where three FM transmitters and one PMR base station are installed.

FIGURE 18

Generalized block diagram of FM and PMR transmitters at one radio site

Rap 2021-18

The three FM transmitters are combined via directional multiplexers (Mux) made of 3 dB couplers and filters (see Fig. 19):

FIGURE 19

Block diagram of a directional multiplexer

Rap 2021-19
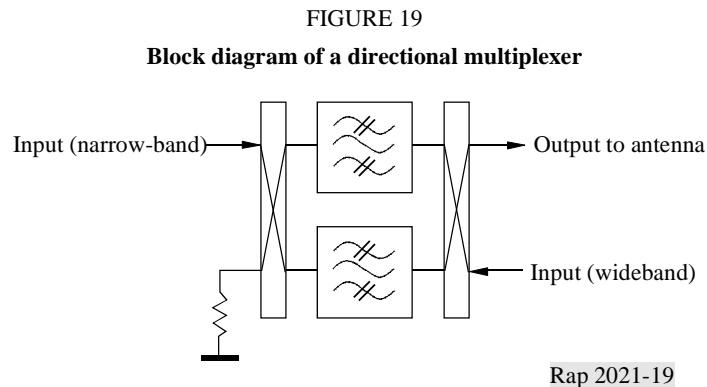
The PMR base station is composed of two duplex couples of channels, Rx1-Tx1 and Rx2-Tx2. Downlink channels, Tx1 and Tx2, are combined via a classical start point combiner and uplink channels, Rx1 and Rx2, are combined via a passive splitter. Resulting downlink and uplink multiplexes are then combined via high pass/low pass start point combiner, considering here that the downlink frequencies are (see above). The following frequencies will be used in examples:

FM: FM1: $f_1 = 88$ MHz,   FM2: $f_2 = 92.7$ MHz and   FM3: $f_3 = 95.7$ MHz

PMR: Rx1: 83.3 MHz,   Tx1: $f_4 = 86.3$ MHz

           Rx2: 83.4 MHz,   Tx 2: $f_5 = 86.4$ MHz

3.4.1 Intermodulation between FM transmitters

The deep selectivity brought by the directional multiplexers used for the FM transmitters will lead to considerable reduction of the following products which will nevertheless appear on the site and will be radiated (intermodulation source in bold) see Figs. 20 to 22.
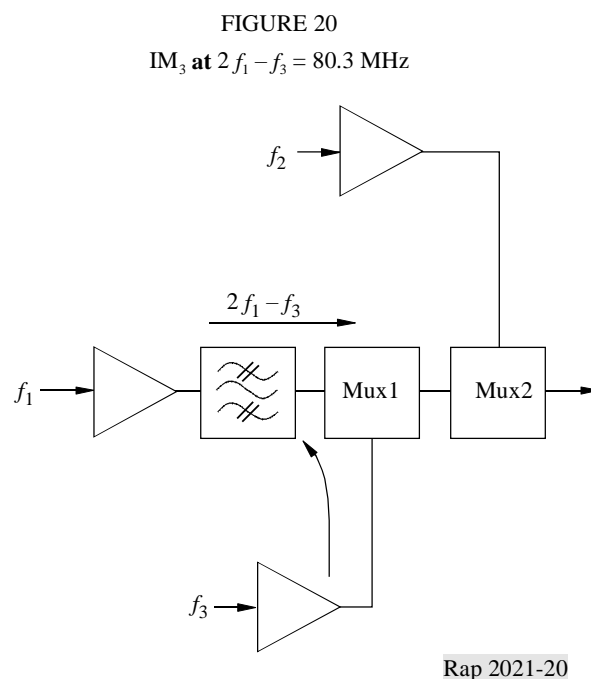
FIGURE 20

IM$_3$ **at** $2 f_1 - f_3 = 80.3$ MHz

Rap 2021-20

FIGURE 21

IM_3 **at** $2 f_1 - f_2 = 83.3$ MHz **(interfering Rx1)**
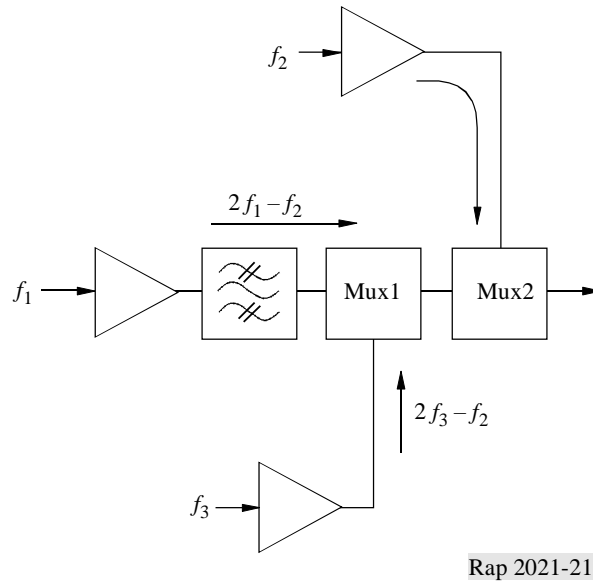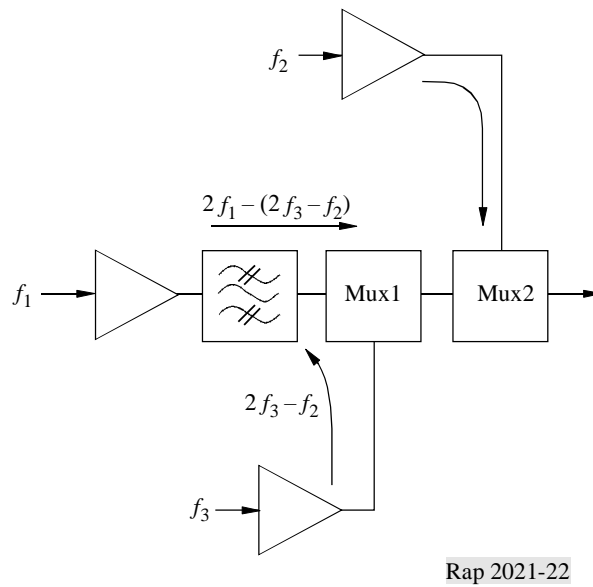**and at** $2 f_3 - f_2 = 98.7$ MHz

Rap 2021-21

FIGURE 22

IM_5 **at** $2 f_1 - 2 f_3 + f_2 = 77.3$ MHz

Rap 2021-22

This latter case is particularly interesting as it shows that 5th order products can result from two steps of 3rd order intermodulation.

3.4.2 Intermodulation between PMR base station transmitters

The transmitters of the PMR base station also generate intermodulation products that will be radiated from the site (see Fig. 23). Those are limited by the spurious emissions limit of the PMR base (–36 dBm) that is considered as a single equipment.

FIGURE 23

IM_3 **at** $2 f_5 - f_4 = 86.5$ MHz

Rap 2021-23
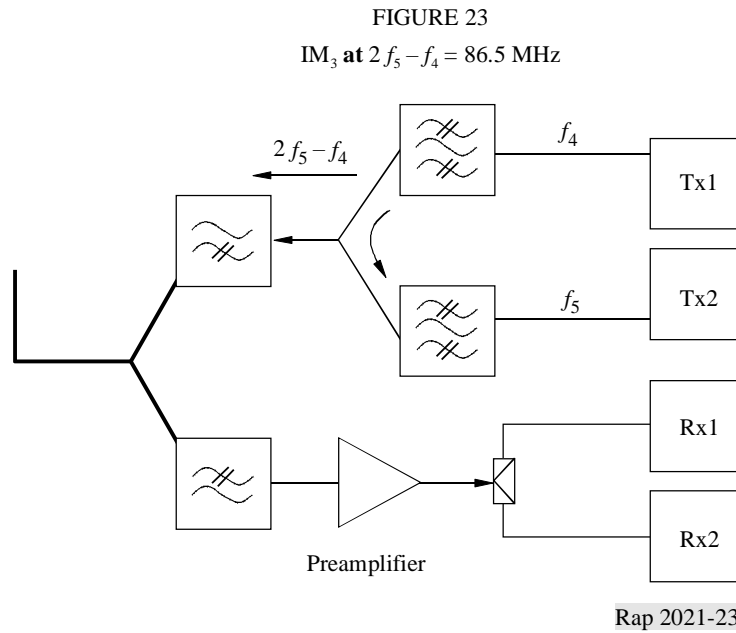
3.4.3 Intermodulation at the input of the PMR base station

The RF preamplifier of the PMR base station is exposed to both FM and PMR transmitted signals. Hence, FM/FM and FM/PMR intermodulation products will be generated (see Figs. 24 and 25).

FIGURE 24

IM_3 **(FM/FM) at** $2 f_1 - f_2 = 83.3$ MHz **(interfering Rx1)**
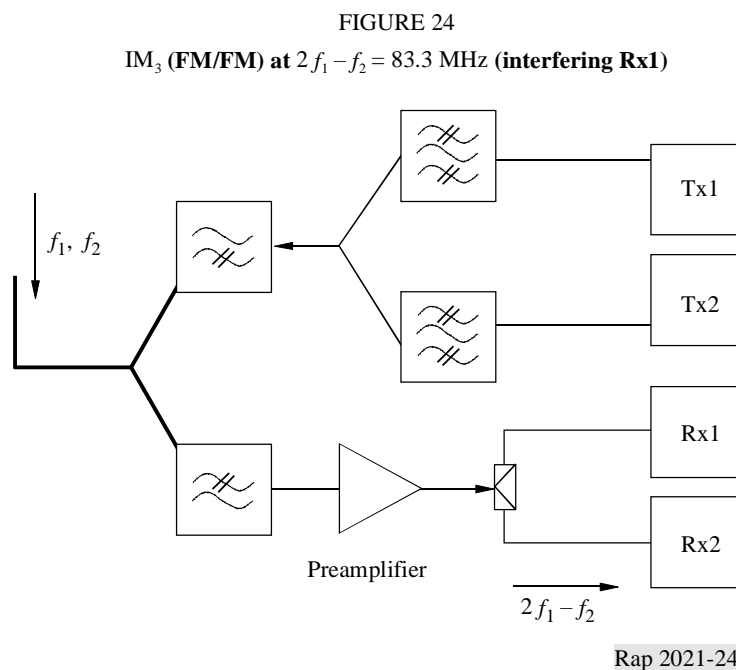
Rap 2021-24

FIGURE 25

IM$_3$ **(FM/PMR) at** $f_4 + f_2 - f_3 = 83.3$ MHz **and** $f_5 + f_2 - f_3 = 83.4$ MHz
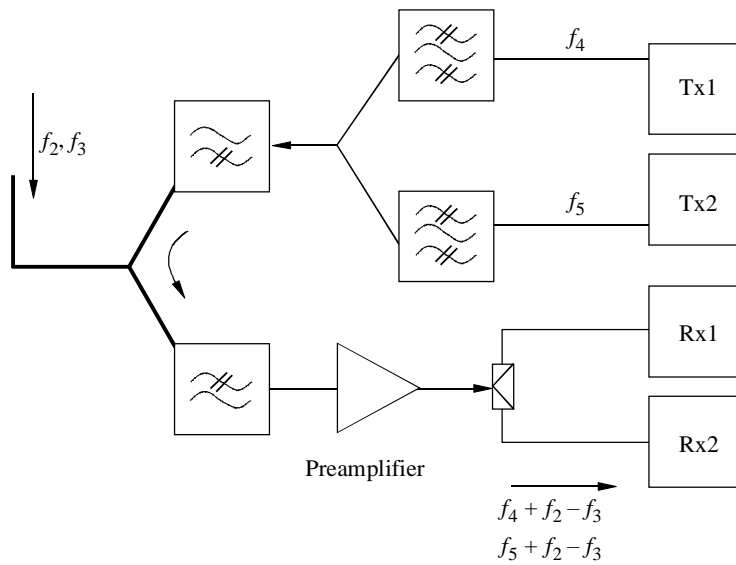
Rap 2021-25

The above example of perturbation is met on radio sites where the frequency separation between 2 FM transmitted is equal to the duplex frequency separation of a PMR base station.

3.4.4 Intermodulation between FM and PMR transmitters

This mechanism is less critical than the previous one but can lead to similar products (see Fig. 26).

FIGURE 26

IM$_3$ **(FM/PMR) at** $f_4 + f_2 - f_3 = 83.3$ MHz **and** $f_5 + f_2 - f_3 = 83.4$ MHz
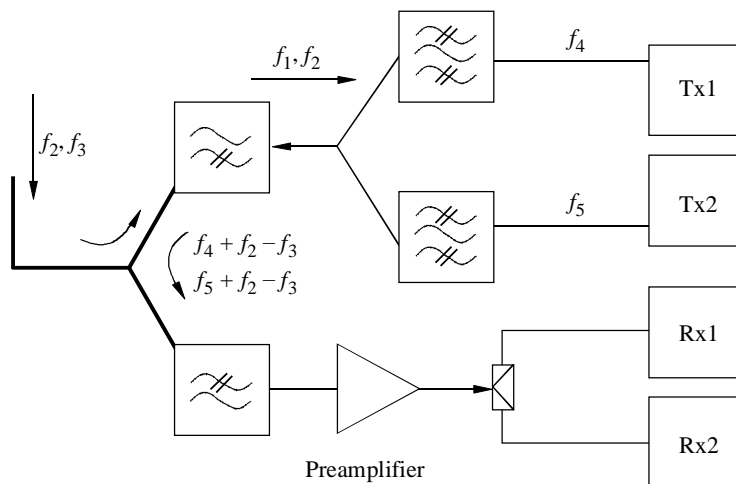
Rap 2021-26

REFERENCES

AGHVAMI, A. H. and ROBERTSON, I. D. [April, 1993] Power limitation and high-power amplifier non-linearities in on-board satellite communications systems. *Electron. and Comm. Engin. J.*

ARTHANAYAKE, T. and WOOD, H. B. [8 April, 1971] Linear amplification using envelope feedback. *Elec. Lett*.

BENNET, T. J. and CLEMENTS, R. F. [May, 1974] Feedforward – An alternative approach to amplifier linearisation. *Radio and Electron. Engin.*

BHARGAVA, V. K. *et al.* [1981] *Digital Communications by Satellite*, John Wiley and Sons.

BLACK, H. S. [October, 1928] Translating system. U.S. Patent No. 1686792.

BLACK, H. S. [December, 1937] Wave translating system. U.S. Patent No. 2102671.

BOND F. E. and MEYER, H. F. [April, 1970] Intermodulation effects in limiter amplifier repeaters. *IEEE Trans. Comm.*, Vol. COM-18, p. 127-135.

CHADWICK, P. [1986] *Wideband Amplifier Applications Book*, Edition 2, Plessey Semiconductor.

COLE, R. A. [December, 1989] Linearisation of a power amplifier using Cartesian Loop feedback. Report No. 72/89/R/451/C. *Roke Manor Res*.

ETSI [August, 1994] Standard ETR 132. Radio broadcasting systems; Code of practice for site engineering VHF FM sound broadcasting transmitters. European Telecommunications Standards Institute, Sophia Antipolis, F-06291, Valbonne Cedex, France.

ETSI [January, 1995] European Standard ETS 300 384. Radio broadcasting systems; Very high frequency (VHF), frequency modulated, sound broadcasting transmitters. European Telecommunications Standards Institute, Sophia Antipolis, F-06291, Valbonne Cedex, France.

ETSI [March, 1997] European Standard ETS 300 113. Radio equipment and systems (RES); Land mobile service; Technical characteristics and test conditions for radio equipment intended for the transmission of data (and speech) and having an antenna connector. European Telecommunications Standards Institute, Sophia Antipolis, F-06291, Valbonne Cedex, France.

ETSI [June, 1998] Standard ETR 053 Ed 3 – Radio site engineering for equipment and systems in the mobile service. European Telecommunications Standards Institute, Sophia Antipolis, F-06291, Valbonne Cedex, France.

GRAY, L. F. [1980] Application of broadband linearisers to satellite transponders. IEEE Conf. Proc. ICC'80.

HEATHMAN, A. C. [1989] Methods for intermodulation prediction in communication systems. Ph. D. Thesis, University of Bradford, United Kingdom.

IESS [November, 1996] IESS-401 (Rev. 4). Performance requirements for intermodulation products transmitted from INTELSAT earth stations. Intelsat Earth Station Standard (IESS).

KAEADAR, K. [December, 1986] Gaussian white-noise generation for digital signal synthesis. *IEEE Trans. Inst. and Meas.*, Vol. IM-35, **4**.

KAHN, L. R. [July, 1952] SSB transmission by envelope elimination and restoration. *Proc. IRE*.

MITCHELL, A. F. [November, 1979] A 135 MHz feedback amplifier. IEEE Colloq. Broadband High Frequency Amplifiers.

PETROVIC, V. and GOSLING, W. [10 May, 1979] Polar loop transmitter. *Elec. Lett*.

Pye Telecom [November, 1978] Intermodulation in VHF and UHF radio systems – locating and minimizing the effects. Engineering Notes, Pub. Ref. No. TSP480/1, United Kingdom.

RA [April, 1987] Code of practice for radio site engineering. MPT 1331. Radiocommunications Agency (RA) United Kingdom.

SALEH, A. M. [mai 1982] Intermodulation analysis of FDMA satellite systems employing compensated and uncompensated TWT's'. *IEEE Trans. Comm.*, Vol. COM-30, **5**.

SCHETZEN, M. [1980] *The Volterra and Wiener Theories of Non-linear Systems*. Wiley & Sons.

SHAHID, M., SHEPHERD, S. J., LIN, B., KHAIRRUDDIN, I., and BARTON, S. K. [December, 1996] Study of methods of measuring multi-carrier intermodulation performance Report No. 581, Purchase Order No. 142379 d'ESA, with University of Bradford, United Kingdom.

SHIMBO, O. [February, 1971] Effects of intermodulation, AM-PM conversion, and additive noise in multicarrier TWT systems. *Proc. IEEE*, Vol. 59, p. 230-238.

SMITH, C. N. [1986] Application of the polar loop technique to UHF SSB transmitters. Ph.D. Thesis, University of Bath.

SMITH, C. N. and PETROVIC, V. [1982] Cartesian loop transmitter. Internal Research Report, University of Bath, School of Electrical and Electronic Engineering.

TONDRYK, W. [1991] Intermodulation testing of the INMARSAT payload – Response to system PDR. Marconi Space Systems Ltd.

WASSERMANN, M. *et al*. [1983] Study and breadboarding of an L-band high power linearized TWT amplifier. Final Report, ESTEC contract No. 5459/83/NL/GM.

WOOD, A. [Octobre 1998] Radio interference: Sources and solutions. *LPRA NEWS*, p. 21.

BIBLIOGRAPHY

PAVLIOUK, A. [1977] Unification of measurement procedures for out-of-band emission spectra and peak envelope power of single-sideband radio transmitter measurements. Proc. of the NIIR, **4** (in Russian).

ANNEX 1

Mathematical description of the generation of intermodulation noise in the transmitter

For estimation of the system performance in terms of BER a rigorous analysis of non-linear systems is required. There are two classical methods for the analysis and synthesis of non-linear systems: the first one carries out the expansion of the signal in a Volterra series [Schetzen, 1980]. The second due to Wiener uses special base functionals for the expansion. These are the Wiener **G** -functionals which are orthogonal if white Gaussian noise excites the system. It is the special autocorrelation property of the white Gaussian noise which makes it so attractive for the analysis of non-linear systems. The filtered version of AWGN, the Brownian movement or the Wiener process, has special features of its autocorrelation which are governed by the rules for mean values of the products of jointly normal random variables.

1   Low-order operators

The non-linear system output signal $y(t)$ can be expressed by a Volterra series:

$$y(t) \;=\; \boldsymbol{H}_0 \;+\; \boldsymbol{H}_1 \;+\; \boldsymbol{H}_2 \;+\; .....$$

where $\boldsymbol{H}_i$ is the abbreviated notation of the Volterra operator operating on the input $x(t)$ of the system. The first three operators are given in the following. The convolution integrals are integrated from $-\infty$, to $+\infty$.

$$\boldsymbol{H}_0\big[x(t)\big] \;=\; h_0$$

$$\boldsymbol{H}_1\big[x(t)\big] \;=\; \int h_1(\tau)\, x(t-\tau)\, \mathrm{d}\tau$$

$$\boldsymbol{H}_2\big[x(t)\big] \;=\; \iint h_2(\tau_1,\tau_2)\, x(t-\tau_1)\, x(t-\tau_2)\, \mathrm{d}\tau_1\, \mathrm{d}\tau_2$$

.....

The kernels of the integral operator can be measured by a variation of the excitation time of input pulses, e.g. for the second order kernel $h_2(\tau_1, \tau_2)$: $x(t) = \delta(t - \tau_1)\, \delta(t - \tau_2)$. A better method is the measurement of the kernel by the cross-correlation of exciting white Gaussian noise $n(t)$ as input signal with the system output $y_i(t)$. These equations hold, if:

$$\Phi_{nn}(\tau) = A\delta(\tau)$$

is the autocorrelation function of the input signal $x(t) = n(t)$ (white Gaussian noise) where A is the noise power spectral density. The first three kernels are given then by:

$$h_0 \;=\; \overline{y_0(t)}$$

$$h_1(\sigma) \;=\; \frac{1}{A}\, \overline{y_1(t)\, n(t-\sigma)}$$

$$h_2(\sigma_1,\sigma_2) \;=\; \frac{1}{2A^2}\, \overline{y_2(t)\, n(t-\sigma_1)\, n(t-\sigma_2)}$$

The overline denotes the expected value, or temporal mean value for ergodic systems. The method can be expanded to higher order systems by using higher order Volterra operators $\boldsymbol{H}_n$. However, the Volterra operators of different order are not orthogonal and, therefore, some difficulties arise at the expansion of an unknown system in a Volterra series.

These difficulties are circumvented by the Wiener **G** -functionals, which are orthogonal to all Volterra operators with lower order, if white Gaussian noise excites the system.

The first three Wiener **G** -functionals are:

$$G_0\big[x(t)\big] = k_0$$

$$G_1\big[x(t)\big] = \int k_1(\tau_1)\, x(t - \tau_1)\, d\tau_1$$

$$G_2\big[x(t)\big] = \iint k_2(\tau_1, \tau_2)\, x(t - \tau_1)\, x(t - \tau_2)\, d\tau_1\, d\tau_2 \; - \; A\int k_2(\tau_1, \tau_1)\, d\tau_1$$

$$G_3\big[x(t)\big] = \iiint k_3(\tau_1, \tau_2, \tau_3)\, x(t - \tau_1)\, x(t - \tau_2)\, x(t - \tau_3)\, d\tau_1\, d\tau_2\, d\tau_3 - 3A\iint k_3(\tau_1, \tau_2, \tau_2)\, x(t - \tau_1)\, d\tau_1\, d\tau_2$$

For these functionals hold:

$$\overline{\boldsymbol{H}_m\big[n(t)\big]\ \boldsymbol{G}_n\big[n(t)\big]} = 0 \quad \text{for} \quad m < n$$

if the input signal $n(t)$ is white Gaussian noise.

2   Example 1: Imperfect square-law device

In Fig. 3 the two data signals $x_1(t)$ and $x_2(t)$ are linearly filtered by the devices with the impulse responses $h_a(t)$ and $h_b(t)$ in adjacent frequency bands. The composite summed signal y is hereafter distorted by an imperfect square-law device which might model a transmit-amplifier. The input-output relation of the non-linear device is given by:

$$z(t) = y(t) + ay^2(t)$$

The input signals $x_1(t)$ and $x_2(t)$ are originated from a single signal $x(t)$, because of the spectral separation by the filters $h_a(t)$ and $h_b(t)$.

The output signal $z(t)$ is therefore determined by:

$$z(t) = \int \big[h_a(\tau) + h_b(\tau)\big] x(t - \tau)\, d\tau + a\left\{\int \big[h_a(\tau) + h_b(\tau)\big] x(t - \tau)\, d\tau\right\}^2$$

The first and second order Volterra-operators $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$ for this example are accordingly determined by the kernels:

$$h_1(\tau) = h_a(\tau) + h_b(\tau)$$

and

$$h_2(\tau_1, \tau_2) = h_a(\tau_1)\big[h_a(\tau_2) + h_b(\tau_2)\big] + h_b(\tau_1)\big[h_a(\tau_2) + h_b(\tau_2)\big]$$

This kernel $h_2(\tau_1, \tau_2)$ is symmetric, so that:

$$h_2(\tau_1, \tau_2) = h_2(\tau_2, \tau_1)$$

The second order kernel transform $H_2(\omega_1, \omega_2)$ is obtained by the two-dimensional Fourier-transform with respect to τ_1 and τ_2, and can be obtained as:

$$H_2(\omega_1, \omega_2) = \big\{H_a(\omega_1)\big[H_a(\omega_2) + H_b(\omega_2)\big] + H_b(\omega_1)\big[H_a(\omega_2) + H_b(\omega_2)\big]\big\}$$

by elementary manipulations. $H_a(\omega)$ and $H_b(\omega)$ are the Fourier-transforms of $h_a(t)$ and $h_b(t)$. With the transform $X(\omega)$ of the input signal $x(t)$, an artificial two dimensional transform $Z_2(\omega_1, \omega_2)$ is obtained:

$$Z_{(2)}(\omega_1, \omega_2) = H_2(\omega_1, \omega_2) \, X(\omega_1) \, X(\omega_2)$$

with the two-dimensional inverse $Z_2(t_1, t_2)$. The output signal $z(t)$ is:

$$z(t) = z_{(2)}(t, t)$$

The transform $Z(\omega)$ of $z(t)$ can be obtained by convolution:

$$Z(\omega) = \frac{1}{2\pi} \int Z_{(2)}(\omega_1, \omega - \omega_1) \, d\omega_1$$

where the integration is carried out from $-\infty$ to $+\infty$.

The output $z(t)$ can be as well represented by use of the Wiener **G**-functionals:

$$z(t) = \mathbf{G}_0 + \mathbf{G}_1 + \mathbf{G}_2 + \dots$$

where $\mathbf{G}_i$ is the simplified notation of $\mathbf{G}_i[x(t)]$. The first two operators are:

$$\mathbf{G}_0[x(t)] = -A \int [h_a(\tau) + h_b(\tau)]^2 d\tau = \text{const}$$

$$\mathbf{G}_1[x(t)] = \int [h_a(\tau) + h_b(\tau)] \, x(t - \tau) \, d\tau$$

The operator $\mathbf{G}_i$ equals $\mathbf{H}_i$ in this example. For $x(t)$ equal white Gaussian noise $x(t) = n(t)$:

$$\overline{\mathbf{G}_1[n(t)] \; h_0} \qquad \text{holds for all } h_0$$

Especially:

$$\overline{\mathbf{G}_1 \mathbf{G}_0} = 0$$

$$\mathbf{G}_2[x(t)] = \int [h_a(\tau_1) \, h_a(\tau_2) + h_a(\tau_1) \, h_b(\tau_2) + h_b(\tau_1) \, h_a(\tau_2) + h_b(\tau_1) \, h_b(\tau_2)]$$

$$x(t - \tau_1) \, x(t - \tau_2) \, d\tau_1 \, d\tau_2 - A \int [h_a(\tau_1) + h_b(\tau_1)]^2 \, d\tau_1$$

The consequence is:

$$\overline{\mathbf{G}_2 h_0} = h_0 \int [h_a(\tau_1) \, h_a(\tau_2) + h_a(\tau_1) \, h_b(\tau_2) + h_b(\tau_1) \, h_a(\tau_2) + h_b(\tau_1) \, h_b(\tau_2)]$$

$$\overline{n(t - \tau_1) \, n(t - \tau_2)} \, d\tau_1 \, d\tau_2 - h_0 A \int [h_a(\tau_1) + h_b(\tau_1)]^2 \, d\tau_1$$

and

$$\overline{\mathbf{G}_2 h_0} = 0 \quad \text{because of} \quad \overline{n(t - \tau_1) \, n(t - \tau_2)} = A\delta(\tau_1 - \tau_2)$$

and similarly:

$$\overline{\mathbf{G}_2 \mathbf{H}_1} = 0 \quad \text{for all } \mathbf{H}_1$$

This equation involves the mean of the product of three zero mean jointly Gaussian random variables, which is zero.

The Wiener kernels can be determined by exciting the system with white Gaussian noise and taking the average of some products of the system output and the exciting noise process $n(t)$:

$$k_0 = \overline{z(t)}$$

$$k_1(\tau) = \frac{1}{A} \overline{z(t)\ n(t-\tau)}$$

and

$$k_2(\tau_1,\tau_2) = \frac{1}{2A^2} \overline{z(t)\ n(t-\tau_1)\ n(t-\tau_2)}$$

Figure 4 shows in the upper part the power spectra of the signals y_1 and y_2 according to Fig. 3, where the filters h_a and h_b form adjacent frequency bands for transmitting in two channels. We assume here a Gaussian shape of the spectra P_a and P_b. The lower part of Fig. 4 illustrates the different spectral contributions which are produced by a square-law device and consist of $P_a * P_a$, $P_a * P_b$ and $P_b * P_b$. No part of the distorted spectrum falls into the original frequency bands $f_0 - 1/T \le f \le f_0$ and $f_0 \le f \le f_0 + 1/T$.

3   Example 2: A third order system

For RF-modulated signals the intermodulation distortion in the proper frequency band is caused by non-linearities of third order. For this reason the imperfect square-law device in Fig. 3 is now replaced by an imperfect cubic device with the input-output relation:

$$z(t) = y(t) + ay^3(t)$$

If only the intermodulation term which distorts the signal in its own frequency band is considered, the kernel transform of the third-order Volterra operator $Z_{(3)}(\omega_1, \omega_2, \omega_3)$ becomes then:

$$Z_{(3)}(\omega_1,\omega_2,\omega_3) = a\prod_{i=1}^{3} \left[H_a(\omega_i) + H_b(\omega_i) \right] X(\omega_i)$$

The intermodulation part in the spectrum of $z(t)$ is now given by:

$$Z(\omega) = \frac{1}{(2\pi)^2} \iint Z_{(3)}(\omega - \mu_1,\mu_1 - \mu_2,\mu_2)\ d\mu_1\ d\mu_2$$

For a cubic device replacing the squarer, however, there are several contributions of the intermodulation noise falling into the used channels near f_0. The different parts $P_a * P_a * P_a \ldots P_b * P_b * P_b$ are plotted in Fig. 5. The thick line shows the sum of the distortions.

REFERENCES

SCHETZEN, M. [1980] *The Volterra and Wiener Theories of Non-linear Systems*. Wiley & Sons.

_____