Sinclair Technologies Inc.

Intermodulation Fundamentals

Dr. Raymond McArthur

1.0 Definition

Intermodulation (IM) or intermodulation distortion (IMD) is a frequency conversion process that occurs when two or more signals pass through a non-linear system or device(s)/component(s) within a system. The essential result of the process is that energy contained in the input signal of a non-linear system is transformed at its output into a set of frequency components at the original frequencies plus additional components at new frequencies that were not contained in the input. The IM phenomenon is often referred to as mixing.

For example, consider a signal composed of two fundamental tones f_1 and f_2 that could represent two transmitter signals co-located at a communications site. If this composite signal is passed through a non-linear device (of third-order), the most general form of the output signal will contain frequency components at dc, f_1, f_2, second-order products and harmonics as well as the third-order products at $2f_1 - f_2$, $2f_2 - f_1$. These last products are often troublesome because they fall closest to the original tones at f_1 and f_2. It is possible that the newly generated third-order products could fall close to or within the receive band of a communication system located at the same site, which could degrade the performance of the receiver.

As another example, consider the same two tones at f_1 and f_2 passing through a stronger non-linear device of fifth-order. The set of most potentially troublesome IM products that can be produced by a fifth-order non-linear system would fall at the frequencies

$2f_1 - f_2$, $2f_2 - f_1$, third-order products,
$3f_1 - 2f_2$, $3f_2 - 2f_1$, fifth-order products.

Note that the order of the non-linearity is determined by the sum of the coefficients. If the non-linearity were stronger still (such as a seventh-order), it would have an output containing the following most potential interfering carriers

$2f_1 - f_2$, $2f_2 - f_1$, third-order products,
$3f_1 - 2f_2$, $3f_2 - 2f_1$, fifth-order products,
$4f_1 - 3f_2$, $4f_2 - 3f_1$, seventh-order products.

With respect to the original tones at f_1 and f_2, the third-order components are closest, the fifth-order are the next closest and the seventh-order are furthest removed but still ‘close’ to f_1 and f_2. This pattern continues for devices of increasing non-linear severity.

When more than two tones of sufficient strength are present at a site, the generated IM products will consist of the set of tones occurring at all linear combinations of the original tones (up to the order of the non-linearity). Some of these IM tones will be potentially threatening to system performance, with the exact threat being dependent upon the particular frequencies and bandwidths of the receivers present at the site.

A more detailed mathematical analysis showing the spectrum widening effects of IM appears in Section 3.0 of this white paper.

2.0 Practical IM Discussion

In the context of transmit and receive components of wireless communication systems, there are basically two classes of IMD that are commonly discussed. The first is Active Intermodulation and the second class is Passive Intermodulation (PIM).

2.1 Active IM

Active Intermodulation is associated with active devices in a system. This can occur, for example, in active devices typically contained in the output stage of transmitters, any portion of an inline amplifier or the input stage of receivers.

In the case of a transmitter, active IM can occur if a signal from another transmitter enters the final stage of the transmitter under consideration. Since the finals are frequently biased as class-C, they are inherently non-linear and are prone to IMD if the interfering tone(s) are strong enough. A rule of thumb is to assume an IM conversion coefficient of about 10 dB for solid state transmitters.

If there is a choice at the initial stages of system design, it is wise to choose a set of transmit and receive frequencies that minimize the potential of generating IM products that fall in or close to the receive bands of interest. When choosing a frequency plan, a software tool in the form of an intermodulation analysis program should be run. After entering a set of potential transmit and receive frequencies, the code will determine the potential generated IM frequencies and given a representative receiver bandwidth can determine if there are any IM 'hits' on any of the chosen receive pass bands. It is important to consider the effects of IM products at least up to the seventh order. In some cases the frequency set can be juggled to minimize potential IM effects.

In the design of systems where a number of transmitters are co-located, isolation between transmitters can be achieved by placing each on its own dedicated antenna. Isolation is achieved by virtue of physical separation between antennas - separation in the vertical direction is much more effective than horizontal separation. Since the expense in terms of antennas, feed cables and especially tower space is at a premium, it is typically impractical to place each transmitter on its own antenna, so transmitter combining systems are employed to enable a number of transmitters to share a common antenna(s).

In transmitter combining systems where the transmitters are literally ‘wired together,’ it is imperative to provide sufficient isolation between the transmitter output stages using in-line devices such as isolators, cavity filters, attenuators, etc. A common practice (and requirement) at frequency congested sites is to have a minimum of a one or two-stage isolator and one or more cavity filters for each individual transmit line.

Isolators provide signal isolation in one direction and serve to block other energy from entering the transmitter through its output line. They also furnish the transmitter a constant VSWR even if the match conditions at the antenna change. Cavity filters provide isolation to energy at frequencies outside of its pass band - this inherently suppresses the second harmonic of the connected transmitter.

It is also possible for transmitter energy to be radiated from its associated antenna and be directly picked up somewhere in another transmitter(s). If the undesired signal is strong enough, this can again cause IM in the power amplifier and a portion of it will be radiated out of the antenna system of the offended transmitter(s). The IM energy can be picked up at the same site or radiated out to areas far removed from the original IM source(s).

In receivers, offending IM usually occurs when multiple signals of sufficient strength overload the first r.f. amplifier stage or the first mixer stage. Typical receiver designs employ narrow band pre-selectors to limit the band of signals entering the r.f. amplifier. The pre-selection (band pass function) provides immunity to signals out of the desired receive band. Examples of such are transmit carrier signals, out of receive band transmitter noise and generated IM signals removed from the receive band. If the offending IM or other interfering signal energy falls close enough to the desired receive frequency, it will pass through the pre-selector and enter the first stage of the r.f. amp.

To test if the receiver front end is the cause of IM, an attenuator can be placed in front of the receiver and the IM product energy observed. If the IM products drop by the value of the attenuator, the offending energy not caused by the receiver and other sources must be identified and dealt with. If the IM products drop by three or more times the value of the attenuator, then the receiver front end is generating IM and the system must be appropriately modified to drop the levels of the multiple tones entering the receiver.

Sinclair Systems Engineers are very experienced in dealing with problems of transmitter and receiver generated IM. The Sinclair Systems Engineering staff can run IM studies on a given frequency plan and use the results to appropriately design a system to minimize the potential of IM.

2.2 Passive IM

The other form of IM that has received much attention in recent years is that of Passive IM, also known as PIM. This is the production of new signal frequency components in a wide array of passive devices commonly found on and around radio sites.

In transmitter and transmit/receive combining systems, all components in the lines that contain multiple tones are potential sources of PIM. Assuming that the transmitters and receivers are not IM sources, all devices that carry the combined signals of different frequencies must be scrutinized. A partial list of common devices in combining systems that have the potential to produce PIM is:

- Cables
- Connectors
- Adapters
- Transmission line junctions such as tees and stars
- Filters
- Directional couplers
- Antennas
- Antenna feed lines
- Surge protectors.

These sources responsible for IM can often be deduced by systematically searching the combining system assemblies and components until the problem is isolated.

Other less obvious potential sources of PIM at a typical site can be:

- Any mounting brackets for antennas, feed lines, etc.
- Rusty or loose fasteners such as bolts, nuts and washers on the tower.
- Tower member joints.
- Tower guy wires.
- Nearby fences, gates and metal signs.
- Hazard/warning lights atop the tower (these can also produce active IM).

These IM sources are much harder to trace. After the combining system is thoroughly probed, sometimes IM sources can be found by ‘sniffing’ in various directions around the site using a highly directional antenna and a spectrum analyzer.

There are two general principles by which PIM is generated in passive devices - magnetic effects and joint/contact effects. These general mechanisms play a strong role in PIM generation *where significant multi-tone r.f. currents are present.*

Magnetic effects cause IM by virtue of hysteresis in magnetic or paramagnetic materials at locations carrying significant multi-tone, high frequency currents.

To combat magnetically generated PIM, avoid the use of ferrous materials in base materials and platings. (In some cases, magnetic material could be used for a base material *if* good plating is used and the plating is not damaged in any way during component or cable assembly.) Avoid the use of nickel as either a base material or outer/under platings.

Joint/contact effects result when metal oxide layers develop in between poor metal-metal contacts. These oxide junctions behave in a similar manner to a semi-conductor diode in terms of having a non-linear transfer characteristic. They are important in locations of significant currents.

To minimize PIM caused by poor contacts and joints, the total number of joints should be minimized. It is best for all metal joints to be welded, brazed or soldered. If these are used they must be implemented correctly using non-corrosive fluxes.

In cases where welding, brazing or soldering isn't possible, all contact points should be scrubbed clean and high contact pressure must be evenly maintained. Maximum contact area should be maintained between all mating surfaces that aren't welded/brazed/soldered.

All cables in a combining system that are required to pass multiple tones at high signal levels should meet the following conditions:

- Minimum use of braids for inner and outer conductor of transmission lines- solid conductors are best.
- If braids/strands are present, minimize cable flexing when in use.
- Soldered braids/strands are better than not soldered.
- Whenever possible avoid nickel anywhere in the cable material system. Sometimes nickel is 'hidden' as an underplating on braids and solid conductors.
- Minimum use of ferrous materials in outer/inner conductors – avoid if possible. If ferrous material is present it must be properly plated and the plating must not be compromised. It is best to avoid ferrous materials altogether.

Choice of connectors and adapters in a system is an important aspect of low-PIM construction. The total number of connectors and adapters should be minimized. Also important is that the total part count of connector sub-assemblies should be minimized.

Connectors will realize the best contact/lowest PIM generation when they are soldered in place. In cases when soldering isn't possible, scrubbed contacts, high mating force and maximum contact area should be maintained.

As in the case of cables and other components, no Nickel outer or under plating should be used. Silver and Gold are good plating materials - Gold also does not form an oxide layer. All material surfaces to be plated must be correctly prepared both mechanically and chemically. Plating must be implemented properly with minimum contamination. Avoid plating materials with excessive ‘brighteners’ – these can significantly decrease conductivity.

It is important to minimize the number of times that connectors are assembled and disassembled. Connectors that are designed to be accurately torqued are best for maintaining low PIM generation over long term. At critical locations in a system, there has been a strong trend in recent years to replace Type-N connectors with 7/16 DIN, as the design of the DIN type is inherently low PIM generating.

Joints between cables and connectors should be of low PIM character. When possible, minimize the use of dissimilar metals at contact points. Soldering is best - when one cannot solder, proper clamping is next best. Clamping, when implemented properly, is much better than crimping. Crimped joints can loosen over time, develop oxide layers and cause PIM.

The above outlined principles and techniques are used in the construction of all low PIM components in transmitter/receiver combining systems as well as in the antenna related portions of the systems. It is important to maintain low PIM generating conditions at all points where there are multiple carriers and where significant multi-carrier r.f. currents are present.

Sinclair has many years of experience in designing components and systems that minimize PIM. To test low PIM designs, Sinclair has a number of IM measuring systems in-house that are used to characterize PIM behavior.

3.0 Non-linear model and simple amplitude analysis

We begin the discussion with the choice of a simple non-linear device model and then proceed to illustrate the important effects of IM by carrying out a simple analysis of a two-tone signal passing through a non-linear system.

In common electrical engineering terminology, the signals may be any of the related quantities of voltage, current, fields, etc. For the purposes of our analysis, we will choose to use the voltage $V = V(t)$. To help shorten the expressions, we will suppress the parenthesis expressing the time dependence.

One of the simplest ways to understand the phenomenon is to consider two pure, single frequency sinusoidal tones and pass the sum of the two through a non-linear device. After carrying out the necessary manipulations, we will make observations of the output amplitudes in the frequency domain.

The first step is to define the type of non-linearity describing the system or device. There are an infinite number of non-linear transfer functions that could be considered but it is common to use one of the simplest models applicable to mild non-linear behavior. We will consider the following non-linear transfer function:

$$V_{out} = K_1 V_{in} + K_2 (V_{in})^2 + K_3 (V_{in})^3 + \ldots \quad . \qquad (1)$$

In this power series representation, the strength of the non-linearity depends upon the values of the complex coefficients $K_1, K_2, K_3, \ldots$. For mild non-linear behavior, we will assume that the series representation is convergent for values of voltages V_{in} under consideration.

To illustrate the frequency domain distortion effects, we will start by considering only the first few terms of the series. (This is a valid approximation if the series is convergent because the neglected terms will introduce only a small error compared to the exact representation of V_{out}.)

Neglecting the terms beyond the cubic, we get

$$V_{out} \sim K_1 V_{in} + K_2 (V_{in})^2 + K_3 (V_{in})^3 \ . \qquad (2)$$

Eq. (2) allows easy calculation of V_{out} to show the intermodulation effects in the frequency domain.

Consider now the pure two-tone input voltage

$$V_{in} = A \cos \omega_1 t + B \cos \omega_2 t, \qquad (3)$$

where $\omega_1 \neq \omega_2$ and it is understood that $\omega = 2\pi f$.

Inserting (3) into (2), expanding and collecting terms, we obtain the resulting equation (4)

$$\begin{array}{lll}
V_{out} \sim & \tfrac{1}{2} K_2 (A^2 + B^2) & \text{DC term} \\
& & \\
+ & (K_1 A + 3/2\, K_3 A B^2 + \tfrac{3}{4} K_3 A^3) \cos \omega_1 t & \text{Fundamental terms} \\
+ & (K_1 B + 3/2\, K_3 A^2 B + \tfrac{3}{4} K_3 B^3) \cos \omega_2 t & \\
& & \\
+ & (1/2\, K_2 A^2) \cos 2\omega_1 t & 2^{\text{nd}} \text{ harmonic terms} \\
+ & (1/2\, K_2 B^2) \cos 2\omega_2 t & \\
& & \\
+ & (K_2 AB) \cos (\omega_1 - \omega_2) t & 2^{\text{nd}}\text{-order products} \\
+ & (K_2 AB) \cos (\omega_1 + \omega_2) t & \\
& & \\
+ & (1/4\, K_3 A^3) \cos 3\omega_1 t & 3^{\text{rd}} \text{ harmonic terms} \\
+ & (1/4\, K_3 B^3) \cos 3\omega_2 t & \\
& & \\
+ & (3/4\, K_3 A^2 B) \cos (2\omega_1 + \omega_2) t & 3^{\text{rd}}\text{-order products} \\
+ & (3/4\, K_3 A B^2) \cos (2\omega_2 + \omega_1) t & \\
+ & (3/4\, K_3 A^2 B) \cos (2\omega_1 - \omega_2) t & \\
+ & (3/4\, K_3 A B^2) \cos (2\omega_2 - \omega_1) t \ . & (4)
\end{array}$$

In each case, the ‘order’ of the harmonic or product is determined by the sum of the non-negative coefficients (m + n) as they occur as arguments of each sinusoidal component in $\cos(\pm m\omega_1 \pm n\omega_2)t$.

From (4) we note a number of important facts. First, the amplitude of the two input sinusoids at the fundamental frequencies ω_1 and ω_2 are distorted by the *odd*-order coefficients K_1 and K_3. Each fundamental frequency component is affected differently by the corresponding initial component amplitudes A and B.

Next, we find that the output consists of a number of *additional signal frequency components that were not present in the initial two-tone input signal* – this is the well-known intermodulation problem. The new signal components appear in the form of DC, 2nd and 3rd harmonics as well as 2nd and 3rd-order ‘sum and difference’ products.

We see that the second harmonics and second-order products are affected by the K_2 power series coefficient only. As with the fundamental terms, each component $2\omega_1$ and $2\omega_2$ is affected differently depending upon how the amplitudes A and B enter into the coefficients. In other words, the harmonic at $2\omega_1$ is determined only by the amplitude A of the input tone at ω_1 and the harmonic at $2\omega_2$ is determined only by the amplitude B of the input tone at ω_2.

In the case of the third-order harmonics and third-order ‘sum and difference’ products, the same observations related to K_3 hold as with the second-order case with K_2. When viewed in the light of real communication systems, there is an additional consideration that is unique, however, to a subset of the third-order products in (4). If the two input tones comprising V_{in} are relatively close together (as would occur in the case of two transmitters operating in the transmit band of a cellular system), the products at the frequencies $(2\omega_1 - \omega_2)$ and $(2\omega_2 - \omega_1)$ will occur relatively close to the fundamentals ω_1 and ω_2. In real systems operating in the same band, all of the other extraneously generated signal components in (4) will either be out of the receiver pass bands or will be filtered out by various components of the system. If the signal components at $(2\omega_1 - \omega_2)$ and $(2\omega_2 - \omega_1)$ enter a nearby receiver, they can cause severe reception problems.

On some antenna sites where systems in different bands are co-located, it is possible for some of the higher-order harmonics to affect systems operating in a different band. Consider the case where a VHF system operating at 150 – 160 MHz is operating on the same site as a UHF system operating near 450 – 480 MHz. Depending upon the frequency plan, it is possible for third harmonics (and possibly the last set of third-order ‘sum and difference’ products) generated in the VHF system to enter the receive bands of the UHF system.

In the case where the amplitudes of the input tones are equal, A = B in (3) and therefore (4) shows that each subset of equal-order tones appearing in V_{out} will have equal amplitudes. For example, both coefficients of output tones at ω_1 and ω_2 are equal when

A = B. Considering the second-order signals, the same is true of the set of tones at $2\omega_1$ and $2\omega_2$. In the case of the generated third-order tones, the set of third harmonics have equal coefficients and the set of 'third-order sum and difference products' have identical amplitudes.

When A = B, it is common to compare the relative amplitudes of the fundamental output tones at ω_1 and ω_2, with the amplitudes of the closest third-order tones at $(2\omega_1 - \omega_2)$ and $(2\omega_2 - \omega_1)$. For a mild non-linearity, $| K_1| >> | K_3 |$ and the output amplitude at ω_1 and ω_2 will be nearly directly proportional to the input amplitude at ω_1 and ω_2. The third-order products at $(2\omega_1 - \omega_2)$ and $(2\omega_2 - \omega_1)$ have an amplitude that is proportional to the *cube* of the input amplitude at ω_1 and ω_2. On a log – log scale with input power on the abscissa and the output power on the ordinate, it is apparent that the third-order products have a slope three times greater than the slope of the fundamental. This means that on a decibel scale, the third-order products change three times faster than the fundamental tones at ω_1 and ω_2. In other words, a 6 dB change in input power will have a corresponding 6 dB change in output power at ω_1 and ω_2 but a 18 dB change in power at $(2\omega_1 - \omega_2)$ and $(2\omega_2 - \omega_1)$.

Although this is accepted as a common rule of thumb when measuring and comparing intermod. levels, it must be understood that there is the underlying assumption that $| K_1| >> | K_3 |$. This is held to be true for 1) a wide range of input power levels and 2) for each mechanism causing the non-linear transfer characteristic. This assumption is not necessarily true if one compares the different mechanisms of non-linearity due to magnetic hysteresis versus non-linearity caused by metal oxides occurring between a crimped connector in partial contact with the outer conductor of a cable. In systems which measure high levels of third and higher-order IM, the non-linearity is not necessarily amplitude and mechanism invariant. Although beyond the present discussion, some IM producing mechanisms have been found to be frequency dependent.

For the simple case of a weak cubic non-linear system or component, we see from (4) the general effect of taking the input spectrum energy at ω_1 and ω_2 and transforming the energy to a new spectrum that includes the original. Depending upon the frequency spacings between transmitters and receivers that are co-located or located near each other, some components of the new spectrum can cause interference in the receiving systems. Higher-order harmonics can cause problems on bands far removed from the original band that contains components that function as I.M. sources.

If we consider the case where the non-linearity can be expressed as

$$V_{out} \sim K_1 V_{in} + K_2 (V_{in})^2 + K_3 (V_{in})^3 + K_4 (V_{in})^4 + K_5 (V_{in})^5, \qquad (5)$$

the resulting output for the same two tone input (3) is a very large expression and will include

a) harmonics up to the fifth-order
b) 'sum and difference' products at the second through fifth-order.

The coefficients of the fundamental frequencies ω_1 and ω_2 present in the output V_{out} are

$$\begin{aligned} &(K_1A + 3/2\, K_3AB^2 + \tfrac{3}{4} K_3A^3 + 15/4\, K_5\, A^3B^2 + 15/8\, K_5\, AB^4 + 5/8\, K_5A^5) \cos \omega_1 t \\ + &(K_1B + 3/2\, K_3A^2B + \tfrac{3}{4} K_3B^3 + 15/4\, K_5\, A^2B^3 + 15/8\, K_5\, A^4B + 5/8\, K_5B^5) \cos \omega_2 t \,. \qquad (6) \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 1^{st}\text{-order} \end{aligned}$$

The IM products of primary importance are those that fall closest to the fundamental input frequencies. These occur at the odd orders only and are

$$\begin{array}{lll} & (3/4\, K_3A^2B + 15/8\, K_5A^2B^3 + 5/4\, K_5A^4B) \cos (2\omega_1 - \omega_2)t & 3^{rd}\text{-order} \\ + & (3/4\, K_3B^2A + 15/8\, K_5B^2A^3 + 5/4\, K_5B^4A) \cos (2\omega_2 - \omega_1)t & \\ & & \\ + & (5/8\, K_5\, A^3B^2) \cos (3\omega_1 - 2\omega_2)t & 5^{th}\text{-order} \\ + & (5/8\, K_5\, B^3A^2) \cos (3\omega_2 - 2\omega_1)t \qquad . & (7) \end{array}$$

If we write the relationship between the two input tones $ά_2$ and $ά_1$as

$$\omega_2 - \omega_1 = \Delta, \qquad (8)$$

then the generated additional frequencies closest to $ά_2$ and $ά_1$will occur at

$$\begin{aligned} &\omega_1 - \Delta, \\ &\omega_2 + \Delta, \qquad\qquad (9) \end{aligned}$$

for third-order non-linearity (2) above and

$$\begin{aligned} &\omega_1 - \Delta, \\ &\omega_2 + \Delta, \\ &\omega_1 - 2\Delta, \\ &\omega_2 + 2\Delta, \qquad\qquad (10) \end{aligned}$$

in the case of a fifth-order non-linearity in (5). This pattern continues for higher-order non-linearities.

In the fifth-order case, the 'rule of thumb' relating the relative change in product order w.r.t. input level change will hold if $|K_1| >> |K_3| >> |K_5|$ is true for a wide range of input amplitudes and IM generating mechanisms. From (6) and (7) we see that on a log-log scale, the fifth-order products will change at a rate five times faster than the fundamental for any given input change. For example, when the inequalities between K_1, K_3 and K_5 hold, a 6 dB change in input amplitude will correspond to a 18 dB change in third-order product levels and a 30 dB change in fifth-order product levels.

In the cases of 7th, 9th, 11th -order non-linearities and beyond, the general pattern for the important coefficients has been established in equations (4), (6) and (7). It is apparent that the odd-order coefficients in the model of the system/component non-linear behavior determine the magnitude of the IM generated additional frequencies that fall closest to the fundamental input tones.

The same frequency generating effects occur when more than two signals pass through a non-linear device or system. In these cases, each input tone will ‘beat’ with the other tones to produce all combinations of the ‘sum and difference frequencies’ associated with the degree of non-linearity under consideration. The analysis will exactly parallel that shown above and the associated expressions for the coefficients making up the output voltage will become extremely large. When viewed as expressed in (9) and (10), the various IM generated tones close to the input frequencies are apparent.

4.0 Conclusion

The subject of intermodulation is extremely important for modern day communication systems. An understanding of the fundamentals of this subject is important for equipment designers, system designers, network operators, or anyone else responsible for proper operation of a communication system. Sinclair openly welcomes further discussion of this subject with any interested parties.