Amplifiers and Transient Intermodulation Distortion

These days, total harmonic distortion (THD) in audio amplifiers is a parameter of minor importance — rendered so by design techniques developed over many years. What occupies the thoughts of audiophiles these days is Transient Intermodulation Distorition (TID or TIM) — a dynamic performance characteristic. Correspondent Wally Parsons discusses traditional design techniques and the modern approach to reducing TID.

WE KNEW ABOUT Transient Intermodulation Distortion in the '50.s — but we didn't call it that. Fortunately our technology was in some respects too advanced, and in others too primitive for the phenomenon to occur very often!

Valves in use then could handle high voltage outputs — so there was no excuse for an amplifier whose early stages overloaded. At the same time, the use of output transformers, good as many of them were, rendered absurd the idea of 40-60 dB feedback. A great deal of attention went into designing high performance into *each* amplifier stage, and into developing stabilizing feedback circuits — techniques still used in transistor amplifiers today.

Distortion

The transistor eliminated some problems (mostly those caused by the output transformers) but brought others of its own.

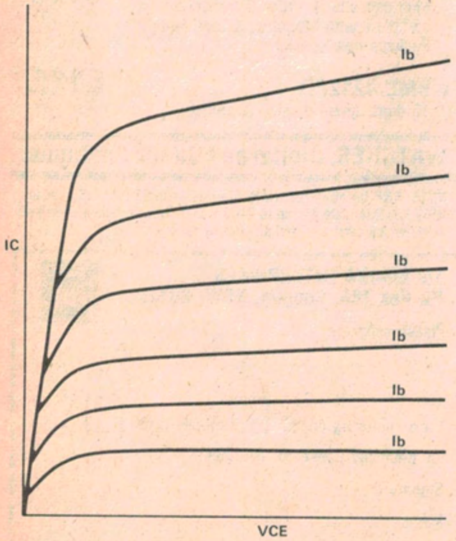

Fig. 1. Non-linear characteristics of a typical bi-polar transistor. Non-linearity causes large amounts of harmonic distortion.

The first and greatest of these is distortion. The bipolar transistor is inherently non-linear: (Fig 1.) put a signal in and out comes the original signal plus lots of harmonics. Plus intermodulation products of all these signals. Very dirty. Secondly, the transistor is very sensitive to changes in operating conditions: small shifts result in major changes in linearity, and temperature changes affect even its fundamental characteristics.

Capacitance Effects

In its simplest form, the bipolar transistor is a sandwich of three pieces of semiconductor. The centre piece is of opposite "polarity" to the outers, and acts as the control device. Application of voltages to these three pieces produces a charged region on either side of each junction. This charge behaves like a capacitance. Thus a transistor in a circuit looks like a complex circuit which includes capacitance between the electrodes (fig 2). Since a capacitor takes finite time to charge and discharge, a time delay occurs between the application of a signal to the input and its appearance at the output. If the signal is ac, the time delay is manifest as a phase delay which increases with frequency, reaching a maximum of 90°, and producing an attenuation of 6 dB per octave (Fig 3, Ref 1).

There is a very simple solution to the first problem of distortion, until we run afoul of the capacitance effects, which seems easy enough to overcome. Therein lies the trap.

Standard Design Practices

The obvious solution to the problems of non-linearity and unstable operating characteristics is negative feedback, right? Where we run into trouble is in what is called "phase margin". Remember the capacitances associated with *each* transistor in the circuit? Each produces a phase shift of 90°. At some frequency the phase will have shifted by 180° or more. This, combined with the 180° phase reversal in the negative feedback loop results in the feedback being 360° "out of phase" with the input. In other words, the feedback is positive.

If the loop gain ...... that is, the product of the forward gain of the amplifier, (which is usually less than one .... is greater than unity, we have an oscillator instead of an amplifier!

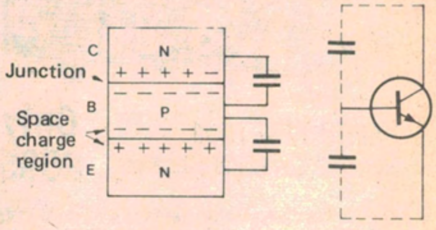

Fig. 2. Capacitance occurs across transistor junctions.

Even if this condition does not occur, but phase is *almost* 360° and/or gain is close to unity, there will be a rise in response, and ringing. Since the capacitances producing phase shift also produce a roll-off in high-frequency response, it is quite possible and desirable, to control both characteristics so that this condition doesn't occur.

This usually means putting additional lump capacitance somewhere within the loop to roll off response and bring it below unity at the 180° shift frequency. This works because attenuation occurs at a continuous rate of 6 dB/octave, but phase shift cannot exceed 90° for a single pole.

There is one major and several minor flaws with this arrangement.

Feedback Variation

If the forward gain of an amplifier is reduced, while the feedback factor (i.e: the fraction of the output signal which is fed back) remains constant,

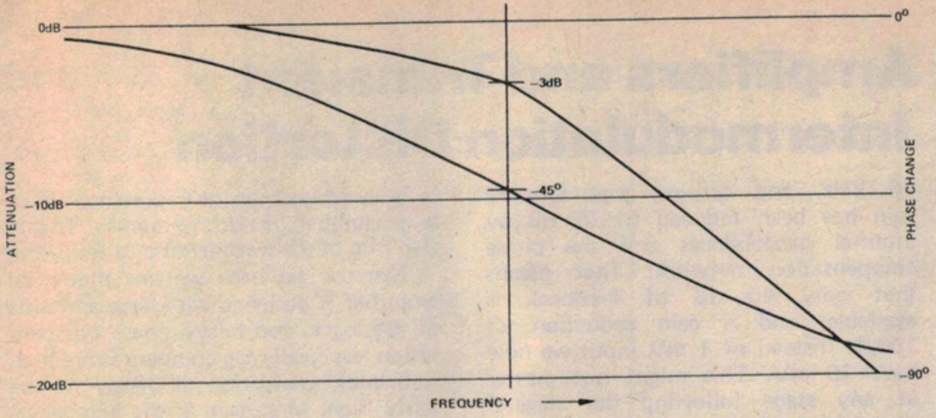

Fig. 3. Graph showing phase change and attenuation, resulting from junction capacitance, plotted against frequency.

it can be shown that the loop gains is reduced by an amount equal to the product of the forward gain reduction and the feedback factor, so that overall gain is almost unchanged. *But the actual amount of feedback is considerably reduced.*

You might say that feedback has been used up in flattening response. As a result, less feedback is available for other things, including *cleaning up distortion.* Remember, feedback does nothing unless it is actually applied. The importance of this apparently obvious fact will shortly become clear.

Figure 4 shows how negative feedback reduces distortion and flattens frequency response. A signal applied to the input really consists of an infinite series of changes in input voltage level. An instantaneous change in input is amplified and appears at the output at a larger magnitude than the input.

Suppose that an input change V, appears at the input and is amplified by a factor, A. If the next incremental change is represented as V+1 then the output should equal (V+1)x A. If a fraction of the output is fed back out of phase with the input, the input level will be reduced by an amount proportional to the signal fed back. But suppose that for a change in input of V+1, the gain of the amplifier was some other value than A, while for input V it remained A. The result is a different input/output ratio and a different ratio of input to feedback. The effective input level change now differs from the applied input.

The result is an input waveform distortion opposite to that produced by the amplifier. Since this now passes through the distorting amplifier the distortion component is cancelled out by an amount dependent on the amount of feedback.

Similarly, if gain remains constant throughout the waveform, but changes with frequency, the amount of signal available for feedback purposes changes and the effective input level changes in a complementary fashion.

Limitations

The greater the amount of feedback applied, the greater the correction of any deviations from linearity. Therefore it would appear that forward linearity, i.e. distortion and frequency response, needn't concern us too much because we can always use buckets of feedback to clean things up.

Much conventional design work seems to proceed along these lines, with impressive specifications — but poor performance as the result!

Phase Effects

Each capacitance in the circuit introduces a roll-off of 6 dB/octave, and a phase shift which is almost 90° at the point where response is down 20 dB. Two such poles introduce almost 180° and three introduce almost 270°. At some frequency there will be a shift of exactly 180° and if we feed back enough signal the amplifier will oscillate at that frequency (fig 5).

The usual remedy is to lower the turnover frequency of one of the poles by introducing *additional* capacitance in the circuit. This changes the amplitude/phase characteristic, and if done properly allows even greater feedback to be applied. But there is still a limit. Unfortunately, this roll-off often occurs within or just above the audio band.

The trouble is, that the actual feedback is reduced as frequency rises. But most transistors, especially power devices, show an *increase* in distortion with rising frequency. In addition, class B amplifiers show a rise in distortion as level is reduced. The only solution is to use whopping amounts of feedback in order to get acceptable performance at high frequencies. This also produces very impressive specs in the mid-range (which some manufacturers emphasise in their advertising). And this is where the fun and games really begin.

Overload

Suppose such an amplifier were driven with a continuous tone sufficient to drive it to full output at, say 1 kHz; distortion somewhere around .001% Suppose, further, that this was achieved with 60 dB feedback, a high figure and pretty impressive, one suggesting an open loop distortion of 1% (pretty low). Finally, suppose that full output is achieved with an input of 1 volt. This represents a gain reduction due to feedback of 1000/1, reducing the 1 volt to 1 mV. What do you suppose would happen if we removed the feedback loop without changing the input signal?

Well, now with the full 1 volt applied we have an overload factor of 1000. Anybody for distortion? — because that's about all that will come out.

Now, let's shift frequency to ▶

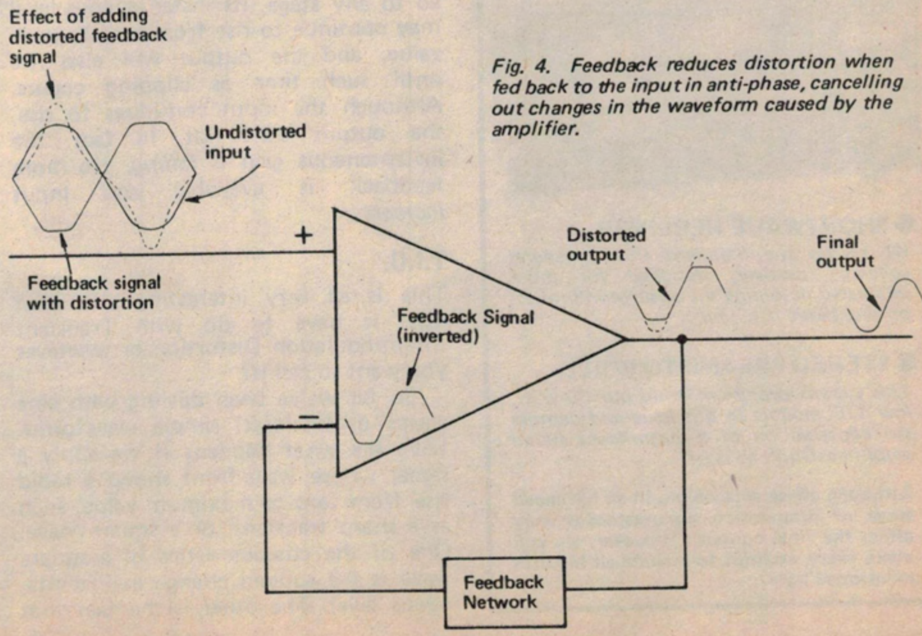

Fig. 4. Feedback reduces distortion when fed back to the input in anti-phase, cancelling out changes in the waveform caused by the amplifier.

10 kHz, and assume that forward gain has been reduced by 20 dB by internal capacitances and the phase compensation network. That means that only 40 dB of feedback is available, and a gain reduction of 100/1. Instead of 1 mV input we now have 10 mV. This might not matter at any stage following the stage(s) at which roll-off occurs, but every previous stage must be capable of handling a signal level ten times the magnitude imposed at 1 kHz. If it can't, then distortion will be introduced in these early stages. These distortion components will be passed on to the output stages at high level, possibly a higher level than they can handle without generating their own distortion and overheating.

Shifting Operating Points

We cannot assume that an input waveform, or the waveform which results from distortion is symmetrical. If it is not, then it can be shown that such a waveform contains a dc component, the presence of which shifts the operating point of the stage to which it is applied. Since such waveforms occur naturally, any good amplifier must be capable of coping with them, but if distortion adds an additional component we have another source of distortion.

High feedback amplifiers generally show low levels of distortion up to the clipping point, at which point the onset of distortion is sudden and severe. This should come as no surprise. When a waveform is presented to any stage, its instantaneous level may continue to rise from zero to peak value, and the output will also rise until such time as clipping occurs. Although the input continues to rise, the output does not. In fact, the instantaneous gain is *falling*. No more feedback is available and input *increases*.

T.I.D.

This is all very interesting but what does it have to do with Transient Intermodulation Distortion or whatever you want to call it?

So far we've been dealing with sine waves or (at least) simple waveforms. Let's see what happens if we apply a signal whose wave-front shows a rapid rise from zero to maximum value, such as a sharp transient, or a square wave. One of the characteristics of a square wave is the sudden change in instantaneous level. The other is the fact that it is produced by the presence of a large number of odd harmonics. This is also true of the wave-front of a transient.

Not too far back we saw that if an amplifier is designed with large amounts of feedback and heavy phase compensation, especially lag compensation, high frequency distortion is likely to be fairly high. Moreover, if the compensation roll-off occurs at a late stage there is a good possibility of overloading early stages.

This is only the start of our problems. Reference was made to lag compensation. This is so-called because output phase lags input. Phase lag means phase delay which also means time delay. A delay of 90° at 5 kHz represents a time delay of 1/20000 second. This is the amount of delay between the time the signal is applied to the input and its appearance at the output. This delay applies to *all* parts of the waveform. It also means that, for this period of time *no feedback of any kind* is applied to the input. In the amplifier described earlier, enormous distortion levels will occur due to overload unless the internal stages can handle this increase in input level. This might not be too much of a problem with simple wave-forms because some feedback will have appeared before the peak is reached. But the transient or square wave may have components at frequencies ten times the fundamental, and *they will* reach their peak before feedback is applied.

As if this weren't enough, should the overload cause major shifts in the operating conditions of any stages, high internal distortions can be generated even after feedback has been applied. In Fig. 6 we see a square wave with a much higher frequency superimposed. Notice how part of the higher frequency signal has been eliminated due to overload caused by the lower frequency.

Solutions

One outstanding characteristic of any amplifier exhibiting low T.I.M. must obviously be very wide frequency response. Moreover, it will exhibit wide response even without feedback. Therefore, such an amplifier will have very little phase compensation, and that which is used will be of the lead variety, producing a forward shift (2). It will also exhibit this wide response at full power. This in turn results in a high slew rate, that is, the rate at which a change in instantaneous input level will be tracked at the output. The wide open loop response implies the use of relatively small amounts of feedback. To do this, and still achieve low steady state distortion levels, requires that each stage be designed for low distortion, which means the use of very linear devices combined with local loop feedback, to produce low open loop distortion so that low overall feedback levels are required to obtain a respectable figure.

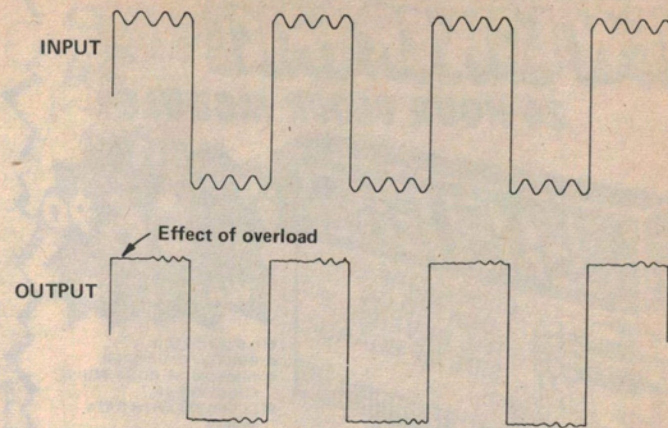

Fig. 5. Overload caused by the lower frequency square wave signal results in any higher frequency information being lost.

One more requirement: we cannot eliminate phase shift and high frequency roll-off completely, we can only minimize them. Therefore, to avoid overload in any stage, each stage must have a band-width equal to or greater than the one preceding it. This even applies to the input. If necessary, the input signal should be filtered to prevent an input rise which exceeds the amplifier's ability to track. We can only apply local feedback to an output stage if it operates in class A, or is a compound such as a Darlington. If no local feedback is available during part of a cycle, and since any really practical and useful feedback network would have to include the driver which, in turn, normally operates class A, the amount of distortion generated during a portion of the cycle would be considerable. This leaves only the voltage amplification stages, and since feedback increases bandwidth, then either some degree of roll-off must go, and/or we must aim for a very high f_t in the output stages. Practical designs usually employ series emitter resistors in these stages — sometimes with compensation.

- **See our 60 watt, low TID amplifier module project on page 45.**

REFERENCES

1. *"V-FETs for Everyone", W Parsons, ETI January 1978.*
2. *"An Audio Amplifier for Ultimate Quality Requirements", Lohstroh and Otala, IEEE Transactions on Audio and Electro-Acoustics Vol AU-21, No.6.*