Pt.3: Negative feedback and frequency response

WHAT IS NEGATIVE

This month we look into why open loop amplifiers suffer poor frequency response and we investigate how negative feedback increases the bandwidth.

By BRYAN MAHER

Analog amplifiers form such an important part of electronic circuitry, that we find them in almost every piece of equipment. Though many signals may be digitized these days, you will always find linear amplification being performed. For example, the ultimate hifi capabilities of CD players can only be realised by using high quality amplifiers in the final stages.

Though feedback concepts can be applied to almost anything — electrical, mechanical, hydraulic or others — our interest here is in electronic systems only. So let's start at the beginning, with an (allegedly) linear electronic amplifier without any feedback at all.

Frequency dependence

Last month we made the bald statement that the open loop gain of an amplifier is different at differing frequencies. Why is that so?

All open loop amplifiers have reduced gain at the high end of the frequency range and possibly at the low end too. First, let's look at the reasons for such loss of gain, then we can investigate how negative feedback improves the situation.

Fig.1 shows a simple "linear" amplifier stage. We call it "linear" simply to distinguish it from deliberately non-linear digital or logic circuits.

By "linear circuits" we mean those with an output voltage which is supposed to be an enlarged but otherwise identical copy of the input voltage waveform.

Strictly we ought to say "roughly linear" because we know that all simple amplifiers have output voltages which are a distorted larger version of their input voltages.

Last month, we told the story of an enthusiastic young lady called Krystie who had built a simple transistor amplifier to play her favourite music. But the results were disappointing. The amplifier not only exhibited excessive distortion but the frequency range was also much less than expected. In fact, the high and low frequency notes were so weak they could barely be heard.

At this point we can find out how *negative feedback* can be an almost magical cure for all amplifier ills.

Looking again at Fig.1, let us consider in detail what should be happening at each point in the circuit and what is actually happening.

First, the signal source (CD player, tape deck or whatever) generates some signal V(in), which we apply to the input, and we expect that same signal waveform V(in) to appear at the base of the transistor. And it does too — more or less.

But why isn't the signal on the base identical to the input? Well, we observe that the signal V(in) has

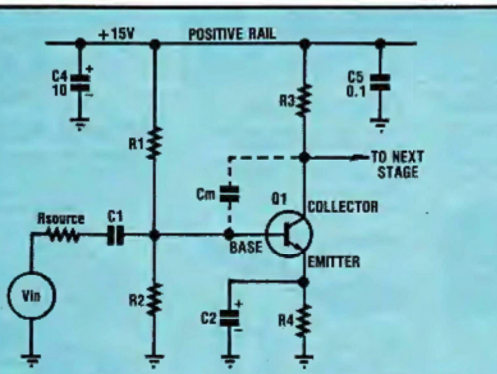

Fig.1: this simple common-emitter amplifier has no negative feedback and suffers from harmonic distortion and deficient low and high frequency response.

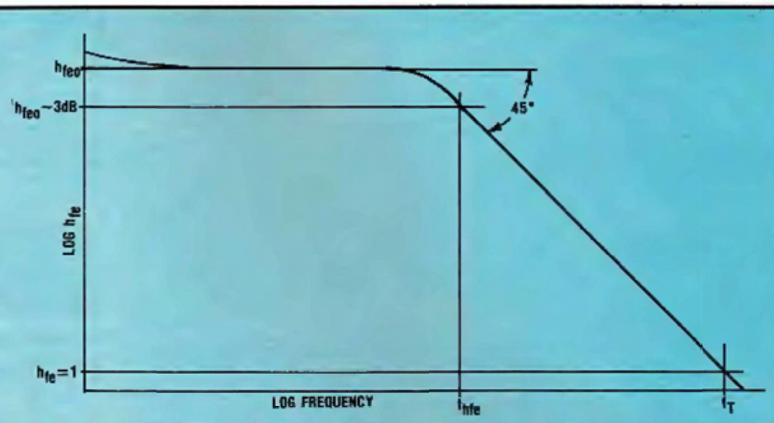

Fig.2: all transistors suffer from loss of gain at high frequencies. At a really high frequency, F_T, the gain drops to unity.

FEEDBACK?

to pass through C1 which has an appreciable impedance, so there is inevitably some signal loss in C1. The effect is worse at low frequencies. But by making C1 sufficiently large, up to 1.0μF, very little low frequency response is lost.

DC coupling?

Why not just DC-couple the whole thing? DC coupling means using no coupling capacitors at all, and indeed that is the ideal. Circuit gain remains undiminished no matter how low the frequency, right down to 0Hz (ie, DC).

But there are drawbacks to DC-coupling right from the input terminals, so common practice is to employ only one coupling capacitor, C1 right at the input. Thereafter, most feedback amplifiers are DC-coupled throughout.

Now what about high notes? Why is the simple amplifier also deficient in the top frequencies? There are two main causes and again Fig.1 gives us some answers.

Loss of h_{fe}

Last month we made some comments about the way the h_{fe} or current gain of any transistor changes at different collector currents and at different temperatures. But current gain is also different at various frequencies.

At a constant collector current and room temperature, the h_{fe} holds constant right down to DC. (It may even be a little higher at DC for the secondary reason that large steady DC currents tend to heat the transistor junction more than AC currents of the same peak value).

However, at high frequencies the value of h_{fe} decreases as shown in Fig.2, and at some very high frequency called f_T the value of h_{fe} has dropped to 1.0. On the curve Fig.2 we define another point f_{hfe}, the frequency at which the h_{fe} has fallen to 0.707 of the DC value.

If the frequencies of interest extend up to f_{hfe} or beyond, the current gain of the transistors will reduce, hence the circuit gain must be less than it is at low frequencies. In practice, we never use transistors anywhere near their upper frequency f_T; that frequency is quoted in transistor data sheets only to enable us to draw the curve Fig.2.

True, we could go out and buy different transistors having higher values of f_{hfe} and f_T, and that is a good idea for the small "front end" transistors in an amplifier. But high frequency large power transistors cost a fortune, so we have to make do with lower frequency types in power transistors.

Shunt capacitances

Though Fig.1 is all you see if you look at the physical circuit of such an amplifier first stage, there are many "stray capacitances" all around the circuit, which we show dotted. These are due to the natural effects of capacitance which exists between all separated conductors and semiconductors.

As well as the stray capacitance C_s from wiring and components to ground, there is the base-emitter capacitance C_{be} in the transistor itself. Also there is a capacitance C_m from the transistor's collector back to the base. This C_m is called the "mutual" or the "Miller" capacitance.

C_m is the most important of the stray capacitances. It is due to the collector-base capacitance C_{cb}, but as the voltage across it is the input voltage plus the stage output voltage, the capacitive current flowing is equivalent to it being a much larger capacitance C_m, where $C_m = C_{cb}(1 + A)$, where A is the stage gain. The total input capacitance C_{in} is the sum: $C_{in} = C_m + C_s + C_{be}$.

The source resistance (in parallel with R1 and R2) forms a low-pass filter with C_{in}, reducing the amplitude of high frequency signals. This happens in every gain stage of an amplifier.

But there is more. At the output side of the transistor stage we have a number of parallel paths:

(1). The equivalent collector output resistance.

(2). The collector-emitter capacitance.

(3). The collector load resistor.

(4). The input resistance and input capacitance of the following stage.

(5). The wiring stray capacitance.

Those capacitances tend to reduce the high frequency response of the stage, but the lower the parallel resultant of the resistances mentioned the less this reduction. Therefore, while higher collector

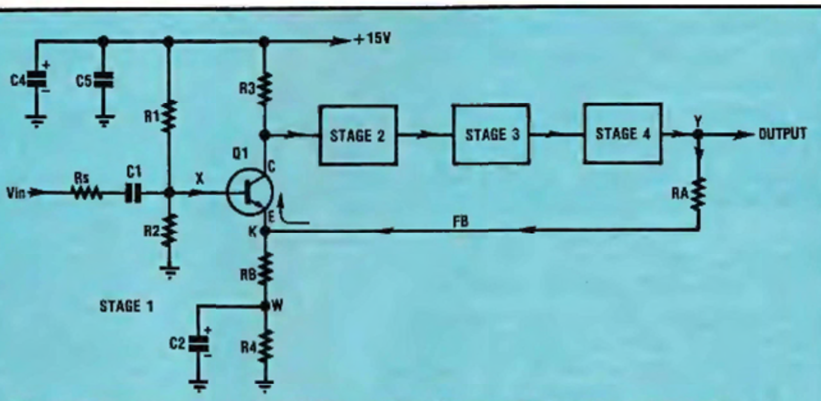

Fig.3: this four-stage amplifier, with all stages identical to Q1, has negative feedback applied via R_A from the output back to the emitter of the first stage.

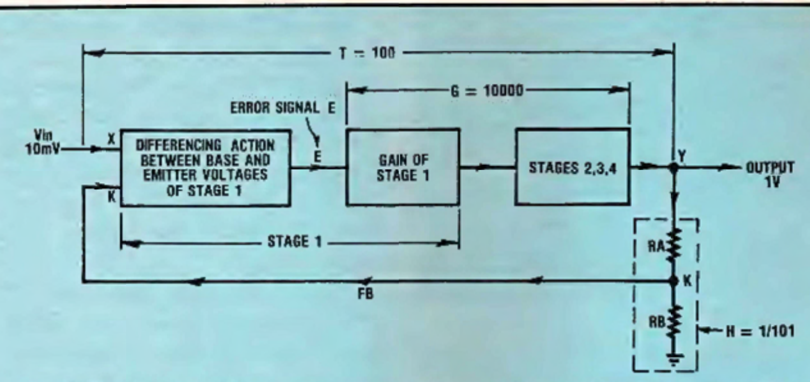

Fig.4: this is a block diagram of the four-stage amplifier shown in Fig.3. R_A and R_B set the proportion of the output voltage fed back to input K.

load resistors may result in higher low-frequency gain, they also cause the gain to drop off badly at higher frequencies.

As most amplifiers have two or more stages, the reduction of gain at higher frequencies will occur in every gain stage. The Miller capacitance will have its greatest effect in the front end high-gain stage (because stage gain A is high), while the frequency dependence of h_{fe} will be responsible for the poor high frequency response in the final high-power stage (because f_T is low).

Enter negative feedback

Fig.3 shows the outline of a four-stage amplifier where Q1 and its sundry components form stage 1. Each stage is inverting, meaning that positive-going inputs produce negative-going outputs, but four stages of phase-inversion result in overall non-inversion; ie positive-going signals at X result in a positive-going output at Y.

In previous episodes we showed some negative feedback block diagrams. Now in Fig.3 we show one possible way to actually apply the feedback voltage from the output back to the input stage. Input signals from the voltage source applied at X naturally increase the output voltage.

But you will recall that the negative feedback must be applied to the front stage in such a way that the feedback voltage reduces the output.

As you probably know, a positive-going signal at X causes an increase in Q1's collector current, which in turn produces greater voltage drop across R3, hence greater stage output at the collector. In Fig.3, resistors RA and RB form a voltage divider across the stage 4 output, from Y to ground.

Do not be confused by the presence of R4, as it is bypassed by C2. The impedance of C2 (at all audio frequencies) is much lower than the resistance of R4, hence there is little or no signal voltage drop across R4. As far as AC signal voltages are concerned, point W (junction of RB and R4) is at ground or zero potential).

Hence at the emitter K of transistor Q1 we have applied some feedback called "FB", a signal which is some fraction of the output voltage. Let us call that fraction "H". So the fraction H will be given approximately by:

$$H = RB \div (RA + RB)$$

K is the point where we have applied feedback signals to the emitter of Q1, and you will observe that we now have a "closed loop system" from Q1 collector C, through stages 2, 3 and 4, to Y, through RA, back to K, through transistor Q1 to C.

In applying negative feedback to our whole amplifier we have reduced its voltage gain. But as we will see, we have improved the amplifier's characteristics in about the same proportion as the reduction in gain.

Some names

(1). The gain of the whole amplifier (measured from the input terminal X to the output terminal Y) before any feedback is applied is called the "Open Loop Gain". This is given by the symbol "G".

(2). The gain of the whole amplifier (measured from X to Y) with feedback applied is called the "Closed Loop Gain". This has the symbol "T".

(3). Because of negative feedback, T is always smaller than G.

(4). The fraction of output used as feedback is called "H" In Fig.3, H = RB ÷ (RA + RB).

(5). The feedback signal derived from the feedback voltage divider H is (naturally enough) called "FB". FB is simply equal to H x Output.

(6). The gain around the loop (from C through stages 2, 3, & 4 to Y, through the voltage divider to K, and through the transistor gain back to C) is called the "Loop Gain".

As the Loop Gain is clearly the product of G and H, we simply refer to the Loop Gain as "GH".

(7). The input signal at X is to be amplified. So we would like the output signal at Y to be an enlarged replica of the signal at X.

(8). As the feedback signal at K is a smaller copy of the output signal at Y, by (6) above we would want the signal at K to be exactly like (but a little smaller than) the input signal at X. If it is, we have succeeded. If not, then we ask the circuit to take corrective action.

(9). The difference between the signals at X and at K is called the "Error Signal". This is given the symbol "E".

(10). It is the error signal E which is amplified by the amplifier.

(11). As the error signal is small, we

The difference between the signals at X and K is the error signal and it is this signal that is amplified by the amplifier.

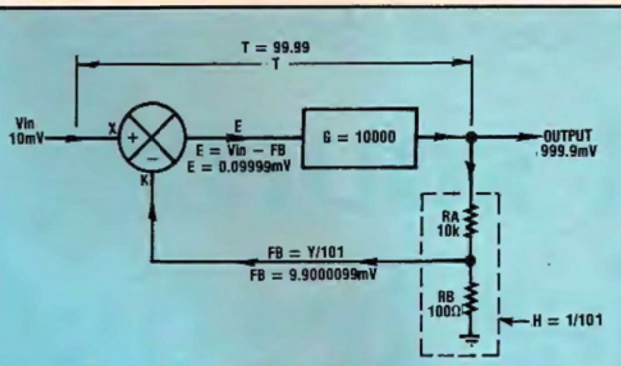

Fig.5: the signal E, which is the difference signal applied between X and K, is multiplied by 10,000 to give the output signal. Note that Vin is relatively large but signal E is extremely small.

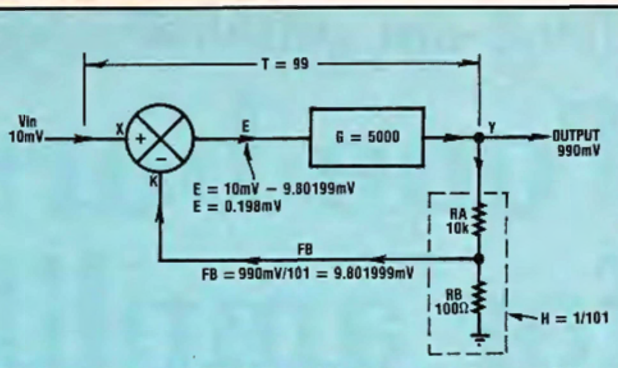

Fig.6: even though the open-loop gain of the amplifier has dropped to 5000, the negative feedback ensures that the output signal is still very close to 1 volt by making the difference signal E larger.

will need considerable gain in the amplifier.

Fig.4 is a block diagram of the circuit of Fig.3. In Fig.4 we have split the action of the first stage Q1 into its two functions:

(a) The differencing action between the input signal at X (the base) and the feedback signal FB at K (the emitter); and

(b) The action of transistor Q1 which (as in all transistors) amplifies the difference between the signals at base and emitter.

Also we have labelled Fig.4 to point up three facts:

(1). It is the error voltage E (not the input voltage at X) which is multiplied by the amplifier gain G.

(2). H is simply a fraction, set by the voltage divider ratio.

(3). FB is a signal which is some set fraction of the output signal.

So just how does all that jazz cure amplifier ills?

One thing at a time, please. Consider first the poor response to high frequencies we noted earlier.

If, for any reason, the output is not be as high as it should be, then the feedback will be smaller. This means that less will be subtracted from the input signal in forming the error signal E, so E will be a bit larger. This will increase the output to (nearly) the right signal level.

Keeping the same amplifier circuit, we redraw Fig.3 and Fig.4, grouping parts of similar function, giving us the simpler Fig.5, where we have written a possible set of conditions; ie resistances of RA and RB, together with gains, and voltages which might be measured in a typical four-stage amplifier at low and middle frequencies.

Low and middle frequencies

We have chosen the amplifier open loop gain G = 10,000, as four stages each having various gains could easily multiply up to that figure. We choose the following values as typical for such an amplifier: input voltage V(in) = 10mV; RA = 10kΩ; RB = 100Ω; G = 10,000 (at low/middle frequencies).

From these values we can calculate that:

H = RB ÷ (RA + RB)
= 100 ÷ (10,000 + 100)
= 100 ÷ 10,100
= 0.00990099.

Alternatively, 1/H = 101.

By this means we can calculate signal voltage values all around the circuit. Without all the calculation details, we have written the results on Fig.5, so if you don't want to bother with calculations, just look at the diagram. Fig.5 shows that at low and middle frequencies, the gain stages of the amplifier give an open loop gain G = 10,000.

(a) An input of 10mV gives an output of 999.9mV, so the closed loop gain T (measured from input X to output Y) is T = 100 (approx).

(b) The feedback network divides the output by 101 to give a feedback voltage FB = 9.9mV (approx), which is subtracted from V(in) to give an error signal E = 0.099mV (approx). This is multiplied by 10,000 to give an output V(out) = 999.9mV.

At high frequencies we must expect the gain G of the amplifier to be less than 10,000 (because of the shunting effect of stray capacitances, the Miller capacitance and the fall-off in h_{fe} at high frequencies as discussed earlier). At some high frequency, the open loop gain G could be down to half; ie, G = 5000. For this condition, as illustrated by Fig.6:

(c) The same input V(in) = 10mV gives an output of 990mV, so the closed loop gain (measured from X to Y) is approximately 99.

(d) The feedback divider fraction is not subject to frequency, so H still divides the new output by 101, giving a feedback voltage FB = 9.8mV (approx), which is subtracted from V(in) to give an error signal E = 0.198mV (approx), which multiplied by the reduced value G = 5000 gives the new output V(out) = 990mV.

(e) We observe (with joy) that even though the gain stages only had half gain (G = 5000 at high frequencies), we still found an overall gain only 1% down (T = 99) at high frequencies; ie, T = 990mV ÷ 10mV = 99.

Conclusion

From all the above we conclude that negative feedback automatically compensates for at least one amplifier shortcoming; the fact that the open loop gain falls at high frequencies. What about those other deficiencies, like distortion, hum and noise?.

Can it be that negative feedback will cure those ills too? That will be our topic for next month's episode. ■