The General Principles of Power Amplifiers

How a Generic Amplifier Works

Figure 3.1 shows a very conventional power amplifier circuit; it is as standard as possible. A great deal has been written about this configuration, though the subtlety and quiet effectiveness of the topology are usually overlooked, and the explanation below therefore touches on several aspects that seem to be almost unknown. The circuit has the merit of being docile enough to be made into a functioning amplifier by someone who has only the sketchiest of notions as to how it works.

The input differential pair implements one of the few forms of distortion cancelation that can be relied upon to work reliably without adjustment – this is because the transconductance of the input pair is determined by the physics of transistor action rather than matching of ill-defined parameters such as beta; the logarithmic relation between $I_c$ and $V_{be}$ is proverbially accurate over some eight or nine decades of current variation.

The voltage signal at the voltage-amplifier stage (hereafter VAS) transistor base is typically a couple of millivolts, looking rather like a distorted triangle wave. Fortunately the voltage here is of little more than academic interest, as the circuit topology essentially consists of a transconductance amp (voltage-difference input to current output) driving into a transresistance (current-to-voltage converter) stage. In the first case the exponential $V_{be}/I_c$ law is straightened out by the differential-pair action, and in the second the global (overall) feedback factor at LF is sufficient to linearize the VAS, while at HF shunt negative feedback (hereafter NFB) through $C_{dom}$ conveniently takes over VAS linearization while the overall feedback factor is falling.

The behavior of Miller dominant-pole compensation in this stage is actually exceedingly elegant, and not at all a case of finding the most vulnerable transistor and slugging it. As frequency rises and $C_{dom}$ begins to take effect, negative feedback is no longer applied globally around the whole amplifier, which would include the higher poles, but instead is seamlessly transferred to a purely local role in linearizing the VAS. Since this stage effectively contains a single gain transistor, any amount of NFB can be applied to it without stability problems.

The amplifier operates in two regions; the LF, where open-loop (O/L) gain is substantially constant, and HF, above the dominant-pole breakpoint, where the gain is decreasing steadily at 6dB/octave.

Assuming the output stage is unity gain, three simple relationships define the gain in these two regions:

$$LFgain = g_m \cdot \beta \cdot R_c$$  \hspace{1cm} Equation 3.1
At least one of the factors that set this (beta) is not well controlled and so the LF gain of the amplifier is to a certain extent a matter of pot luck; fortunately this does not matter, so long as it is high enough to give a suitable level of NFB to eliminate LF distortion. The use of the word *eliminate* is deliberate, as will be seen later. Usually the LF gain, or HF local feedback factor, is made high by increasing the effective value of the VAS collector impedance $R_c$, either by the use of current-source collector load, or by some form of bootstrapping.

The other important relations are:

$$HF\text{gain} = \frac{8_m}{\omega \cdot C_{\text{dom}}} \quad \text{Equation 3.2}$$

Dominant-pole frequency

$$P1 = \frac{1}{C_{\text{dom}} \cdot \beta \cdot R_c} \quad \text{Equation 3.3}$$
Where:

\[ \omega = 2 \cdot \pi \cdot \text{frequency} \]

In the HF region, things are distinctly more difficult as regards distortion, for while the VAS is locally linearized, the global feedback factor available to linearize the input and output stages is falling steadily at 6dB/octave. For the time being we will assume that it is possible to define an HF gain (say, \( N \) dB at 20kHz), which will assure stability with practical loads and component variations. Note that the HF gain, and therefore both HF distortion and stability margin, are set by the simple combination of the input stage transconductance and one capacitor, and most components have no effect on it at all.

It is often said that the use of a high VAS collector impedance provides a current drive to the output devices, often with the implication that this somehow allows the stage to skip quickly and lightly over the dreaded crossover region. This is a misconception – the collector impedance falls to a few kilohms at HF, due to increasing local feedback through \( C_{\text{dom}} \), and in any case it is very doubtful if true current drive would be a good thing: calculation shows that a low-impedance voltage drive minimizes distortion due to beta-unmatched output halves\(^1\), and it certainly eliminates the effect of Distortion 4, described below.

**The Advantages of the Conventional**

It is probably not an accident that the generic configuration is by a long way the most popular, though in the uncertain world of audio technology it is unwise to be too dogmatic about this sort of thing. The generic configuration has several advantages over other approaches:

- The input pair not only provides the simplest way of making a DC-coupled amplifier with a dependably small output offset voltage, but can also (given half a chance) completely cancel the second-harmonic distortion that would be generated by a single-transistor input stage. One vital condition for this must be met; the pair must be accurately balanced by choosing the associated components so that the two collector currents are equal. (The typical component values shown in Figure 3.1 do not bring about this most desirable state of affairs.)

- The input devices work at a constant and near-equal \( V_{ce} \), giving good thermal balance.

- The input pair has virtually no voltage gain so no low-frequency pole can be generated by Miller effect in the TR2 collector-base capacitance. All the voltage gain is provided by the VAS stage, which makes for easy compensation. Feedback through \( C_{\text{dom}} \) lowers VAS input and output impedances, minimizing the effect of input-stage capacitance, and the output-stage capacitance. This is often known as pole-splitting\(^2\); the pole of the VAS is moved downwards in frequency to become the dominant pole, while the input-stage pole is pushed up in frequency.

- The VAS Miller compensation capacitance smoothly transfers NFB from a global loop that may be unstable, to the VAS local loop that cannot be. It is quite wrong to state that all the benefits of feedback are lost as the frequency increases above the dominant pole, as the VAS is still being linearized. This position of \( C_{\text{dom}} \) also swamps the rather variable \( C_{cb} \) of the VAS transistor.
The Distortion Mechanisms

My original series of articles on amplifier distortion listed seven important distortion mechanisms, all of which are applicable to any Class-B amplifier, and do not depend on particular circuit arrangements. As a result of further experimentation and further thought, I have now increased this to ten.

In the typical amplifier THD is often thought to be simply due to the Class-B nature of the output stage, which is linearized less effectively as the feedback factor falls with increasing frequency. This is, however, only true when all the removable sources of distortion have been eliminated. In the vast majority of amplifiers in production, the true situation is more complex, as the small-signal stages can generate significant distortion of their own, in at least two different ways; this distortion can easily exceed output stage distortion at high frequencies. It is particularly inelegant to allow this to occur given the freedom of design possible in the small-signal section.

If the ills that a Class-B stage is prone to are included then there are eight major distortion mechanisms. Note that this assumes that the amplifier is not overloaded in any way, and therefore is not suffering from:

- activation of any overload protection circuitry;
- overloading not affecting protection circuitry (for example, insufficient current to drive the output stage due to a VAS current source running set to too low a value);
- slew-rate limiting;
- defective or out-of-tolerance components.

It also assumes the amplifier has proper global or Nyquist stability and does not suffer from any parasitic oscillations; the latter, if of high enough frequency, cannot be seen on the average oscilloscope and tend to manifest themselves only as unexpected increases in distortion, sometimes at very specific power outputs and frequencies.

In Figure 3.2 an attempt has been made to show the distortion situation diagrammatically, indicating the location of each mechanism within the amplifier. Distortion 8 is not shown as there is no output capacitor.

The first four distortion mechanisms are inherent to any three-stage amplifier.

Distortion 1: Input Stage Distortion

This concerns nonlinearity in the input stage. If this is a carefully balanced differential pair then the distortion is typically only measurable at HF, rises at 18 dB/octave, and is almost pure third harmonic. If the input pair is unbalanced (which from published circuitry it usually is) then the HF distortion emerges from the noise floor earlier, as frequency increases, and rises at 12 dB/octave as it is mostly second harmonic.

This mechanism is dealt with in Chapter 4.
**Distortion 2: VAS Distortion**

Nonlinearity in the voltage-amplifier stage (which I call the VAS for brevity) surprisingly does not always figure in the total distortion. If it does, it remains constant until the dominant-pole frequency $P_1$ is reached, and then rises at 6 dB/octave. With the configurations discussed here it is always second harmonic.

Usually the level is very low due to linearizing negative feedback through the dominant-pole capacitor. Hence if you crank up the local VAS open-loop gain, for example by cascoding or putting more current-gain in the local VAS–$C_{\text{dom}}$ loop, and attend to Distortion 4 below, you can usually ignore VAS distortion.

This mechanism is dealt with in Chapter 5.

**Distortion 3: Output Stage Distortion**

Nonlinearity in the output stage, which is naturally the obvious source. This in a Class-B amplifier will be a complex mix of large-signal distortion and crossover effects, the latter generating a spray of high-order harmonics, and in general rising at 6 dB/octave as the amount of negative feedback decreases. Large-signal THD worsens with 4 Ω loads and worsens again at 2 Ω. The picture is complicated by dilatory switch-off in the relatively slow output devices, ominously signaled by supply current increasing in the top audio octaves.

These mechanisms are dealt with in Chapters 6 and 7.
**Distortion 4: VAS-Loading Distortion**

This is loading of the VAS by the nonlinear input impedance of the output stage. When all other distortion sources have been attended to, this is the limiting distortion factor at LF (say, below 2 kHz); it is simply cured by buffering the VAS from the output stage. Magnitude is essentially constant with frequency, though the overall effect in a complete amplifier becomes less as frequency rises and feedback through $C_{\text{dom}}$ starts to linearize the VAS.

This mechanism is dealt with in Chapter 7.

The next three distortion mechanisms are in no way inherent; they may be reduced to unmeasurable levels by simple precautions. They are what might be called topological distortions, in that they depend wholly on the arrangement of wiring and connections, and on the physical layout of the amplifier.

**Distortion 5: Rail-Decoupling Distortion**

Nonlinearity caused by large rail-decoupling capacitors feeding the distorted signals on the supply lines into the signal ground. This seems to be the reason that many amplifiers have rising THD at low frequencies. Examining one commercial amplifier kit, I found that rerouting the decoupler ground return reduced the THD at 20 Hz by a factor of 3.

This mechanism is dealt with in Chapter 7.

**Distortion 6: Induction Distortion**

This is nonlinearity caused by induction of Class-B supply currents into the output, ground, or negative-feedback lines. This was highlighted by Cherry[3] but seems to remain largely unknown; it is an insidious distortion that is hard to remove, though when you know what to look for on the THD residual it is fairly easy to identify. I suspect that a large number of commercial amplifiers suffer from this to some extent.

This mechanism is dealt with in Chapter 7.

**Distortion 7: NFB Take-Off Distortion**

This is nonlinearity resulting from taking the NFB feed from slightly the wrong place near where the power-transistor Class-B currents sum to form the output. This may well be another very prevalent defect.

This mechanism is dealt with in Chapter 7.

The next two distortion mechanisms relate to circuit components that are non-ideal or poorly chosen.

**Distortion 8: Capacitor Distortion**

In its most common manifestation this is caused by the non-ideal nature of electrolytic capacitors. It rises as frequency falls, being strongly dependent on the signal voltage across the capacitor.
The most common sources of nonlinearity are the input DC-blocking capacitor or the feedback network capacitor; the latter is more likely as it is much easier to make an input capacitor large enough to avoid the problem. It causes serious difficulties if a power amplifier is AC-coupled, i.e. has a series capacitor at the output, but this is rare these days.

It can also occur in ceramic capacitors that are nominally of the NP0/C0G type but actually have a significant voltage coefficient, when they are used to implement Miller dominant-pole compensation. This mechanism is dealt with in detail in Chapter 7.

**Distortion 9: Magnetic Distortion**

This arises when a signal at amplifier output level is passed through a ferromagnetic conductor. Ferromagnetic materials have a nonlinear relationship between the current passing through them and the magnetic flux it creates, and this induces voltages that add distortion to the signal. The effect has been found in output relays and also speaker terminals. The terminals appeared to be made of brass but were actually plated steel.

This mechanism is also dealt with in detail in Chapter 7.

**Distortion 10: Input Current Distortion**

This distortion is caused when an amplifier input is driven from a significant source impedance. The input current taken by the amplifier is nonlinear, even if the output of the amplifier is distortion free, and the resulting voltage drop in the source impedance introduces distortion.

This mechanism is purely a product of circuit design, rather than layout or component integrity, but it has been put in a category of its own because, unlike the inherent Distortions 1–4, it is a product of the interfacing between the amplifier and the circuitry upstream of it.

This mechanism is dealt with in Chapter 4.

**Distortion 11: Premature Overload Protection**

The overload protection of a power amplifier can be implemented in many ways, but without doubt the most popular method is the use of VI limiters that shunt signal current away from the inputs to the output stage. In their simplest and most common form, these come into operation relatively gradually as their set threshold is exceeded, and introduce distortion into the signal long before they close it down entirely. It is therefore essential to plan a sufficient safety margin into the output stage so that the VI limiters are never near activation in normal use. This issue is examined more closely in Chapter 17.

Other methods of overload protection that trigger and then latch the amplifier into a standby state cannot generate this distortion, but if this leads to repeated unnecessary shutdowns it will be a good deal more annoying than occasional distortion.
Nonexistent or Negligible Distortions

Having set down what might be called the Eleven Great Distortions, we must pause to put to flight a few paper tigers …

The first is common-mode distortion in the input stage, a specter that haunts the correspondence columns. Since it is fairly easy to make an amplifier with less than <0.00065% THD (1 kHz) without paying any attention at all to this issue it cannot be too serious a problem. It is perhaps a slight exaggeration to call it nonexistent, as under special circumstances it can be seen, but it is certainly unmeasurable under normal circumstances.

If the common-mode voltage on the input pair is greatly increased, then a previously negligible distortion mechanism is indeed provoked. This increase is achieved by reducing the C/L gain to between 1 and 2×; the input signal is now much larger for the same output, and the feedback signal must match it, so the input stage experiences a proportional increase in common-mode voltage.

The distortion produced by this mechanism increases as the square of the common-mode voltage, and therefore falls rapidly as the closed-loop gain is increased back to normal values. It therefore appears that the only precautions required against common-mode distortion are to ensure that the closed-loop gain is at least five times (which is no hardship, as it almost certainly is anyway) and to use a tail-current source for the input pair, which again is standard practice. This issue is dealt with in more detail in the chapter on power amplifier input stages.

The second distortion conspicuous by its absence in the list is the injection of distorted supply-rail signals directly into the amplifier circuitry. Although this putative mechanism has received a lot of attention[4], dealing with Distortion 5 above by proper grounding seems to be all that is required; once more, if triple-zero THD can be attained using simple unregulated supplies and without paying any attention to the power-supply rejection ratio (PSRR) beyond keeping the amplifier free from hum (which it reliably can be) then there seems to be no problem. There is certainly no need for regulated supply rails to get a good performance. PSRR does need careful attention if the hum/noise performance is to be of the first order, but a little RC filtering is usually all that is needed. This topic is dealt with in Chapter 9.

A third mechanism of very doubtful validity is thermal distortion, allegedly induced by parameter changes in semiconductor devices whose instantaneous power dissipation varies over a cycle. This would surely manifest itself as a distortion rise at very low frequencies, but it simply does not happen. There are several distortion mechanisms that can give a THD rise at LF, but when these are eliminated the typical distortion trace remains flat down to at least 10Hz. The worst thermal effects would be expected in Class-B output stages where dissipation varies wildly over a cycle; however, drivers and output devices have relatively large junctions with high thermal inertia. Low frequencies are of course also where the NFB factor is at its maximum. This contentious issue is dealt with at greater length in Chapter 6.

To return to our list of the unmagnificent eleven, note that only Distortion 3 is directly due to output stage nonlinearity, though Distortions 4–7 all result from the Class-B nature of the typical output stage. Distortions 8–11 can happen in any amplifier, whatever its operating class.
The THD curve for the standard amplifier is shown in Figure 3.3. As usual, distortion increases with frequency, and as we shall see later, would give grounds for suspicion if it did not. The flat part of the curve below 500 Hz represents non-frequency-sensitive distortion rather than the noise floor, which for this case is at the 0.0005% level. Above 500 Hz the distortion rises at an increasing rate, rather than a constant number of dB/octave, due to the combination of Distortions 1–4. (In this case, Distortions 5–7 have been carefully eliminated to keep things simple; this is why the distortion performance looks good already, and the significance of this should not be overlooked.) It is often written that having distortion constant across the audio band is a good thing – a most unhappy conclusion, as the only practical way to achieve this with a normal Class-B amplifier is to increase the distortion at LF, for example by allowing the VAS to distort significantly.

It should now be clear why it is hard to wring linearity out of such a snake-pit of contending distortions. A circuit-value change is likely to alter at least two of the distortion mechanisms, and probably change the O/L gain as well; in the coming chapters I shall demonstrate how each distortion mechanism can be measured and manipulated in isolation.

Open-Loop Linearity and How to Determine It

Improving something demands measuring it, and thus it is essential to examine the open-loop linearity of power-amp circuitry. This cannot be done directly, so it is necessary to measure the NFB factor and calculate open-loop distortion from closed-loop measurements. The closed-loop gain is normally set by input sensitivity requirements.

Measuring the feedback factor is at first sight difficult, as it means determining the open-loop gain. Standard methods for measuring op-amp open-loop gain involve breaking feedback loops and
manipulating C/L gains, procedures that are likely to send the average power amplifier into fits. Nonetheless the need to measure this parameter is inescapable as a typical circuit modification – e.g. changing the value of R2 changes the open-loop gain as well as the linearity, and to prevent total confusion it is essential to keep a very clear idea of whether an observed change is due to an improvement in O/L linearity or merely because the O/L gain has risen. It is wise to keep a running check on this as work proceeds, so the direct method of open-loop gain measurement shown in Figure 3.4 was evolved.

**Direct Open-Loop Gain Measurement**

The amplifier shown in Figure 3.1 is a differential amplifier, so its open-loop gain is simply the output divided by the voltage difference between the inputs. If output voltage is kept constant by providing a constant swept-frequency voltage at the positive input, then a plot of open-loop gain versus frequency is obtained by measuring the error-voltage between the inputs, and referring this to the output level. This gives an upside-down plot that rises at HF rather than falling, as the differential amplifier requires more input for the same output as frequency increases, but the method is so quick and convenient that this can be lived with. Gain is plotted in dB with respect to the chosen output level (+16 dBu in this case) and the actual gain at any frequency can be read off simply by dropping the minus sign. Figure 3.5 shows the plot for the amplifier in Figure 3.1.

The HF-region gain slope is always 6 dB/octave unless you are using something special in the way of compensation, and by the Nyquist rules must continue at this slope until it intersects the horizontal line representing the feedback factor, if the amplifier is stable. In other words, the slope is not being accelerated by other poles until the loop gain has fallen to unity, and this provides a simple way of putting a lower bound on the next pole P2; the important P2 frequency (which is usually somewhat mysterious) must be above the intersection frequency if the amplifier is seen to be stable.

![Test circuit for measuring open-loop gain directly. The accuracy with which high O/L gains can be measured depends on the test-gear CMRR](image)

**Figure 3.4:** Test circuit for measuring open-loop gain directly. The accuracy with which high O/L gains can be measured depends on the test-gear CMRR.
Given test gear with a sufficiently high common-mode rejection ratio (CMRR) balanced input, the method of Figure 3.4 is simple; just buffer the differential inputs from the cable capacitance with TL072 buffers, which place negligible loading on the circuit if normal component values are used. In particular be wary of adding stray capacitance to ground to the negative input, as this directly imperils amplifier stability by adding an extra feedback pole. Short wires from power amplifier to buffer IC can usually be unscreened as they are driven from low impedances.

The test-gear input CMRR defines the maximum open-loop gain measurable; I used an Audio Precision System-1 without any special alignment of CMRR. A calibration plot can be produced by feeding the two buffer inputs from the same signal; this will probably be found to rise at 6dB/octave, being set by the inevitable input asymmetries. This must be low enough for amplifier error signals to be above it by at least 10dB for reasonable accuracy. The calibration plot will flatten out at low frequencies, and may even show an LF rise due to imbalance of the test-gear input-blocking capacitors; this can make determination of the lowest pole P1 difficult, but this is not usually a vital parameter in itself.

Using Model Amplifiers

Distortions 1 and 2 can dominate amplifier performance and need to be studied without the manifold complications introduced by a Class-B output stage. This can be done by reducing the circuit to a model amplifier that consists of the small-signal stages alone, with a very linear Class-A emitter-follower attached to the output to allow driving the feedback network; here small signal refers to current rather than voltage, as the model amplifier should be capable of giving a full power-amp voltage swing, given sufficiently high rail voltages. From Figure 3.2 it is clear that this will allow study of Distortions 1 and 2 in isolation, and using this approach it will prove relatively easy to design a small-signal amplifier with negligible distortion across the audio band, and this is the only sure foundation on which to build a good power amplifier.

Figure 3.5: Open-loop gain versus frequency plot for Figure 3.1. Note that the curve rises as gain falls, because the amplifier error is the actual quantity measured.
A typical plot combining Distortions 1 and 2 from a model amp is shown in Figure 3.6, where it can be seen that the distortion rises with an accelerating slope, as the initial rise at 6 dB/octave from the VAS is contributed to and then dominated by the 12 dB/octave rise in distortion from an unbalanced input stage.

The model can be powered from a regulated current-limited PSU to cut down the number of variables, and a standard output level chosen for comparison of different amplifier configurations; the rails and output level used for the results in this work were ±15 V and +16 dBu. The rail voltages can be made comfortably lower than the average amplifier HT rail, so that radical bits of circuitry can be tried out without the creation of a silicon cemetery around your feet. It must be remembered that some phenomena such as input-pair distortion depend on absolute output level, rather than the proportion of the rail voltage used in the output swing, and will be increased by a mathematically predictable amount when the real voltage swings are used.

The use of such model amplifiers requires some caution, and gives no insight into BJT output stages, whose behavior is heavily influenced by the sloth and low current gain of the power devices. As a general rule, it should be possible to replace the small-signal output with a real output stage and get a stable and workable power amplifier; if not, then the model is probably dangerously unrealistic.

The Concept of the Blameless Amplifier

Here I introduce the concept of what I have chosen to call a Blameless audio power amplifier. This is an amplifier designed so that all the easily defeated distortion mechanisms have been rendered negligible. (Note that the word Blameless has been carefully chosen not to imply perfection, but merely the avoidance of known errors.) Such an amplifier gives about 0.0005% THD at 1 kHz and approximately 0.003% at 10 kHz when driving 8 Ω. This is much less THD than a Class-B

![Graph](image-url)

Figure 3.6: The distortion from a model amplifier, produced by the input pair and the voltage-amplifier stage. Note increasing slope as input pair distortion begins to add to VAS distortion.
amplifier is normally expected to produce, but the performance is repeatable, predictable, and definitely does not require large global feedback factors.

Distortion 1 cannot be totally eradicated, but its onset can be pushed well above 20kHz by the use of local feedback. Distortion 2 (VAS distortion) can be similarly suppressed by cascoding or beta-enhancement, and Distortions 4–7 can be made negligible by simple topological methods. All these measures will be detailed later. This leaves Distortion 3, which includes the intractable Class-B problems, i.e. crossover distortion (Distortion 3b) and HF switch-off difficulties (Distortion 3c). Minimizing 3b requires a Blameless amplifier to use a BJT output rather than FETs.

A Blameless Class-B amplifier essentially shows crossover distortion only, so long as the load is no heavier than 8Ω; this distortion increases with frequency as the amount of global NFB falls. At 4Ω loading an extra distortion mechanism (3a) generates significant third harmonic.

The importance of the Blameless concept is that it represents the best distortion performance obtainable from straightforward Class-B. This performance is stable and repeatable, and varies little with transistor type as it is not sensitive to variable quantities such as beta.

Blamelessness is a condition that can be defined with precision, and is therefore a standard other amplifiers can be judged against. A Blameless design represents a stable point of departure for more radical designs, such as the Trimodal concept in Chapter 10. This may be the most important use of the idea.

References