Distortion in power amplifiers, Part VII: frequency compensation and real designs

EDN edn.com/distortion-in-power-amplifiers-part-vii-frequency-compensation-and-real-designs/

February 13, 2008

[Part 1 offers an overview and introduction to the sources of distortion in audio power amplifiers. Part 2 focuses on distortion in the audio amplifier input stage. Part 3 examines distortion mechanisms in the voltage amplifier stage (VAS). Part 4 focuses on distortion in the audio power amplifier output stage. Part 5 continues the discussion of distortion in the power amplifier output stages. Part 6 looks at some of the remaining distortion mechanisms in audio power amplifiers.]

The distortion performance of an amplifier is determined not only by open loop linearity, but also the negative feedback factor applied when the loop is closed. In most practical circumstances doubling the NFB factor halves the distortion.

To date, this series has focused on basic circuit linearity. I have assumed that open loop gain falls at 6 dB/octave due to a single dominant pole, with the amount of NFB permissible at HF being set by the demands of HF stability. Because of this, the distortion residuals from a 'blameless' amplifier are comprised almost entirely of crossover artifacts due to their high frequency content. Audio amplifiers using more advanced compensation are rather rare. However, certain techniques do exist …

This series has stuck close to conventional topologies, because even commonplace circuitry has been shown to have little known aspects and interesting possibilities. This implies a two-gain-stage circuit (unity gain output stages not being counted) with most of the feedback applied globally, but smoothly transferred to the voltage amplifier stage alone as frequency increases.

Other configurations are possible; a one stage amplifier is an intriguing possibility " they are common in cmos op-amps " but is probably ill-suited to power amp impedances. See Ref. 1 for an eccentric three-stage amplifier with an open loop gain of just 52 dB (due to the dogged use of local feedback) and only 20 dB of global feedback. Most of the section below refers only to the conventional two-stage structure.

Making a pole dominant
Dominant pole compensation is the simplest kind, though its implementation involves subtlety. Simply take the lowest pole to hand (PI), and make it dominant, i.e. so much lower in frequency than the next pole P2 that the total loop gain (the open loop gain as reduced by the attenuation in the feedback network) falls below unity before enough phase shift accumulates to cause HF oscillation. With a single pole, the gain must fall at 6 dB/octave, corresponding to a constant 90° phase shift. Thus the phase margin will be 90° giving good stability. Figure 1(a) shows the traditional Miller method of making a dominant pole.

From the book "Self on Audio," by Douglas Self

Figure 1 Implementing dominant-pole compensation. (a) Miller capacitor, (b) Shunt-lag circuit (c) Output-stage Inclusive Miller compensation. (d) How to implement 2-pole compensation.


The collector pole of Tr_4 is lowered by adding the Miller capacitance C_{dom} to that which unavoidably exists as the C_{bc} of the VAS transistor. However there are other beneficial effects; C_{dom} causes 'pole splitting', in which the pole at Tr_2 collector is pushed up in frequency as P1 moves down " most desirable for stability. Simultaneously the local NFB through C_{dom} linearises the VAS.

Assuming that input stage transconductance is set to a plausible 5 mA/V, and stability considerations set the maximal 20 kHz open loop gain to 50 dB, then from the equations in Part 1, C_{dom} must be 125 pF. This is more than enough to swamp the internal capacitances of the VAS transistor, and is a realistic value.

The peak current that flows in and out of this capacitor for an output of 20 V r.m.s., 20 kHz, is 447 µA. Recalling that the input stage must sink C_{dom} current while the VAS collector load sources it, and likewise the input stage must source it while the VAS sinks it, there are four possible places in which slew rate might be limited by inadequate current
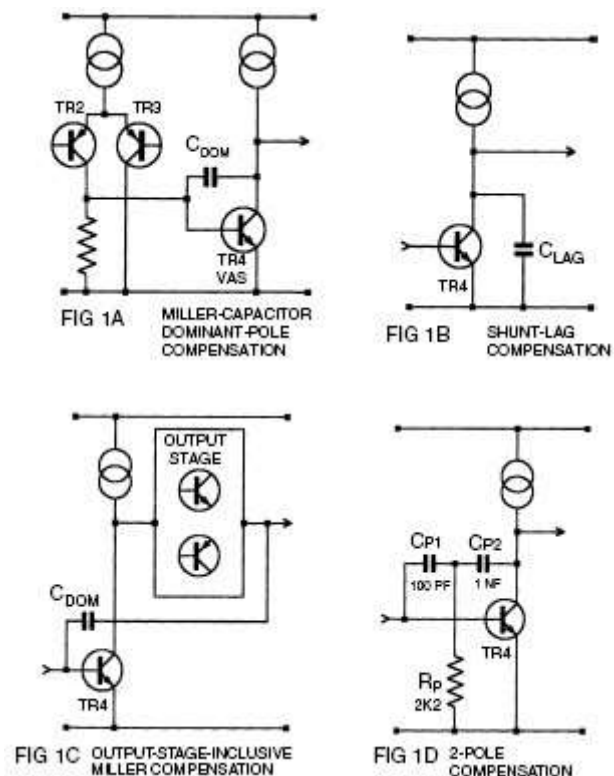
capacity. If the input stage is properly designed then the usual limiting factor is VAS current sourcing. In this example a peak current of less than 0.5 mA should be easy to deal with, and the maximum frequency for unslewed output will be comfortably above 20 kHz.

Figure 1(b) shows a much less satisfactory method, the addition of capacitance to ground from the VAS collector. This is usually called shunt lag compensation, and as Peter Baxandall aptly put it, 'The technique is in all respects sub-optimal'.[2]

We have already seen in Part 3 that loading the VAS collector resistively to ground is a very poor option for reducing LF open loop gain, and a similar argument shows that capacitive loading to ground for compensation purposes is an even worse idea. To reduce open loop gain at 20 kHz to 50 dB as before, the shunt capacitor C_{lag} must be 43.6 nF, which is a whole different order of things from 125 pF. The current flowing in C_{lag} at 20 V r.m.s., 20 kHz, is 155 mA peak, which is going to require some serious electronics to provide it. This important result can be derived by simple calculation, and I have confirmed it with Spice simulation. The input stage no longer constrains the slew rate limits, which now depend entirely on the VAS.

A VAS working under these conditions is almost certain to have poor linearity. The current variations in the stage, caused by the extra loading, produces more distortion and there is now no local NFB through a Miller capacitor to correct it. To make matters worse, the dominant pole P1 will probably need to be set to a lower frequency than for the Miller case, to maintain the same stability margins, as there is now no pole splitting to raise the pole at the input stage collector. Hence C_{lag} may have to be even larger, and require even higher peak currents. Takahashi has produced a fascinating paper on this approach,[3] showing one way of heaving about the enormous compensation currents required for good slew rates. The only thing missing is an explanation of why shunt compensation was chosen in the first place.

Including the output stage
Miller capacitor compensation elegantly solves several problems at once, and the decision to use it is not hard. However the question of whether to include the output stage in the Miller feedback loop is less easy. Such inclusion (see Figure 1(c)) presents the desirable possibility that local feedback could linearise both the VAS and the output stage, with just the input stage left out in the cold as frequency rises and global NFB falls. This idea is most attractive as it would greatly increase the feedback available to linearise a Class B output stage.

There is certainly *some* truth in this where applying C_{dom} around the output as well as the V_{as} reduced the peak 1 kHz THD from 0.05% to 0.02%.[4] However, it should be pointed out that the output stage was deliberately under biased to induce crossover spikes because, with optimal bias, the improvement was too small to be either convincing or worthwhile. Also, this demonstration used a model amplifier with TO-92 'output' transistors. In my experience this technique just does not work with real power bipolars because it induces intractable HF oscillation.

The use of local NFB to linearise the VAS demands a tight loop with minimal extra phase shift beyond that inherent in the C_{dom} dominant pole. It is permissible to insert a cascode or a small signal emitter follower into this local loop, but a sluggish output stage seems to be pushing the phase margin too far; the output stage poles are now included in the loop, which loses its dependable HF stability. Bob Widlar has stated that output stage behaviour must be well controlled up to 100 MHz for the technique to be reliable.[5] This would appear to be virtually impossible for discrete power stages with varying loads.

While I have so far not found 'Inclusive Miller compensation' to be workable myself, others may know different. If anyone can shed further light I would be most interested.

Nested feedback loops
Nested feedback is a way to apply more NFB around the output stage without increasing the global feedback factor. The output has an extra voltage gain stage bolted on, and a local feedback loop is closed around these two stages. This NFB around the composite block reduces output stage distortion and increases frequency response, to make it safe to include in the global NFB loop.

Suppose that block A_1 (Figure 2(a)) is a distortionless small signal amplifier providing all the open loop gain and so including the dominant pole.

Figure 2 The principle of nested feedback loops.

A_3 is a unity gain output stage with its own main pole at 1 MHz and distortion of 1% under given conditions: this 1 MHz pole puts a firm limit on the amount of global NFB that can be safely applied.

Figure 2(b) shows a nested feedback version; an extra gain block A_2 has been added, with local feedback around the output stage. A_3 has the modest gain of 20 dB so there is a good chance of stability when this loop is closed to bring the gain of $A_3 + A_2$ back to unity. A_2 now experiences 20 dB of NFB, bringing the distortion down to 0.1%, and raising the main pole to 10 MHz, which should allow the application of 20 dB more global NFB around the overall loop that includes A_1. We have thus decreased the distortion that exists before global NFB is applied, and simultaneously increased the amount of NFB that can be safely used, promising that the final linearity could be very good indeed. For another theoretical example see Ref. 6.
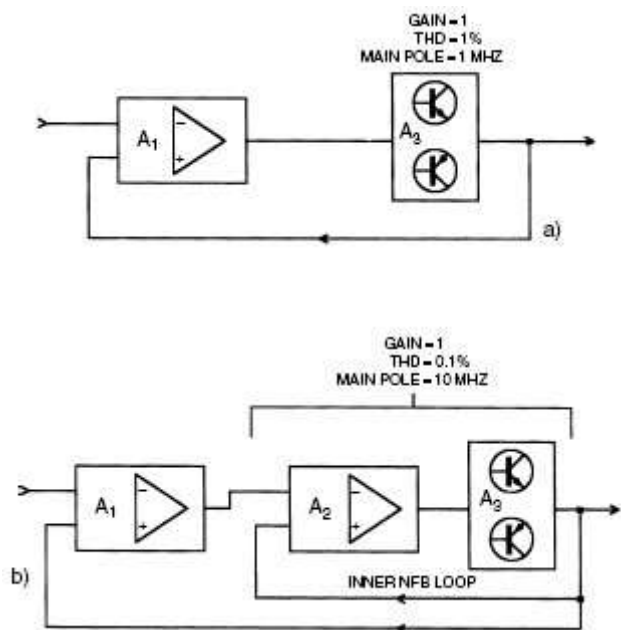
Real life examples of this technique in power amps are not easy to find, but a variation is widely used in op-amps. Many of us were long puzzled by the way that the much loved 5534 maintained such low THD up to high frequencies. Contemplation of its entrails [see NE5534 datasheet] appears to reveal a three-gain stage design with an inner Miller loop around the third stage, and an outer Miller loop around the second and third stages; global NFB is then applied externally around the whole lot. Nested Miller compensation has reached its apotheosis in cmos op-amps " the present record appears to be three nested Miller loops plus the global NFB.[7] Don't try this one at home.

Two pole compensation

Two pole compensation is a mildly obscure technique for squeezing the best performance from an op-amp,[8,9] but it has rarely been applied to power amplifiers. I know of only one example.[5] An extra HF time constant is inserted in the C_{dom} path, giving an open loop gain curve that initially falls at almost 12 dB/octave, but which gradually reverts to 6 dB/octave as frequency continues to increase. This reversion is arranged to happen well before the unity loop gain line is reached, and so stability should be the same as for the conventional dominant pole scheme, but with increased negative feedback over part of the operational frequency range. The faster gain roll off means that the maximum amount of feedback can be maintained up to a higher frequency. There is no measurable mid band peak in the closed loop response.

One should be cautious about any circuit arrangement which increases the NFB factor. Power amplifiers face loads that vary widely: it is difficult to be sure that a design will always be stable under all circumstances. This makes designers rather conservative about compensation, and I approached this technique with some trepidation. However, results were excellent with no obvious reduction in stability. Figure 7 shows the result of applying this technique to the Class B amplifier described below.

The simplest way to implement two pole compensation is shown in Figure 1(d), with typical values. C_{p1} should have the same value as it would for stable single pole compensation, and C_{p2} should be at least twice as big; R_P is usually in the region 1 k"10 k. At intermediate frequencies C_{p2} has an impedance comparable with R_p, and the resulting extra time constant causes the local feedback around the VAS to increase more rapidly with frequency, reducing the open loop gain at almost 12 dB/octave.

At HF the impedance of R_p is high compared with C_{p2}, the gain slope asymptotes back to 6 dB/octave, and then operation is the same as conventional dominant pole, with C_{dom} equal to the series capacitance combination. So long as the slope returns to 6 dB/octave before the unity loop gain crossing occurs, there seems no obvious reason why the Nyquist stability should be impaired.

Figure 3 shows a simulated open loop gain plot for realistic component values; C_{p2} should be at least twice C_{p1} so the gain falls back to the 6 dB/octave line before the unity loop gain line is crossed. The potential feedback factor has been increased by more than 20 dB from 3 kHz to 30 kHz, a region where THD tends to increase due to falling NFB. The open loop gain peak at 8 kHz looks extremely dubious, but I have so far failed to detect any resulting ill effects in the closed loop behaviour.

Figure 3 J3 Spice plot of the open-loop gain of a 2-pole compensated amplifier. The difference between the two plots shows the amount of extra NFB possible.

There is however a snag to the simple approach shown here, which reduces the linearity improvement. Two-pole compensation may decrease open loop linearity at the same time as it raises the feedback factor that strives to correct it. At HF, C_{p2} has low impedance and allows R_p to directly load the VAS collector to ground. This worsens VAS linearity as we have seen. However, if C_{p2} and R_P are correctly proportioned the overall reduction in distortion is dramatic and extremely valuable. When two pole compensation was added to Figure 4, the crossover glitches on the THD residual almost disappeared, being partially replaced by low level 2nd harmonic which almost certainly results from VAS loading. The positive slew rate will also be slightly reduced.

This looks like an attractive technique, as it can be simply applied to an existing design by adding two inexpensive components. If C_{p2} is much larger than C_{p1}, then adding/removing R_p allows instant comparison between the two kinds of compensation. Be warned that if an amplifier is prone to HF parasitics then this kind of compensation may exacerbate them.

Design example: a 50 W class B amplifier

Figure 4 shows a design example of a Class B amplifier intended for domestic audio. Despite its conventional appearance, the circuit delivers a far better distortion performance than that normally associated with the arrangement.
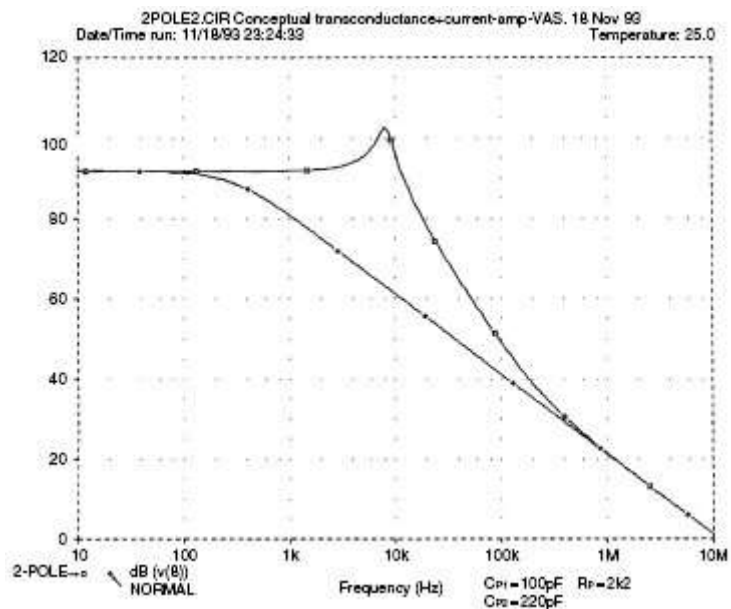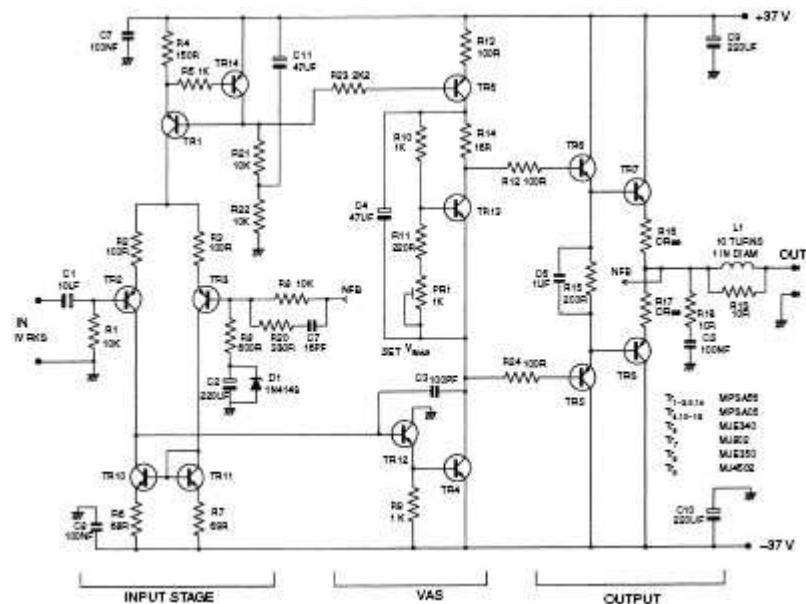
Figure 4 50 W Class B amplifier circuit diagram. Transistor numbers correspond with the generic amplifier in chapter 16.

With the supply voltages and values shown it delivers 50W/8Ω from IV r.m.s. input. In previous articles I have used the word *blameless* to describe amplifiers in which all distortion mechanisms, except the apparently unavoidable ones due to Class B, have been rendered negligible. This circuit has the potential to be *blameless* , but achieving this depends on care in cabling and layout. It does not aim to be a cookbook project; for example, overcurrent and DC offset protection are omitted.

The investigation presented in chapters 19 and 20 concluded that power fets were expensive, inefficient and non linear. Bipolars make good output devices. The best BJT configurations were the emitter follower type II, with least output switch-off distortion, and the complementary feedback pair (CFP) giving best basic linearity and quiescent stability.

I assume that domestic ambient temperatures will be benign, and the duty moderate, so that adequate quiescent stability can be attained by suitable heatsinking and thermal compensation. The configuration chosen is therefore emitter follower type II, which has the advantage of reducing switch-off nonlinearities (Distortion 3(c)) due to the action of R_{15} in reverse biasing the output base emitter junctions as they turn off. The disadvantage is that quiescent stability is worse than for the CFP output topology, as there is no local feedback loop to servo out V_{be} variations in the hot output devices.

The NFB factor was chosen as 30 dB at 20 kHz, which should give generous HF stability margins. The input stage (current source Tr_1 , Tr_{14} and differential pair $Tr_{2,3}$) is heavily degenerated by $R_2 R_3$ to delay the onset of third harmonic Distortion 1. To assist this the contribution of transistor internal re variation is minimised by using the unusually high tail current of 4mA. $Tr_{10,11}$ form a degenerated current mirror that enforces accurate balance of the $Tr_{2,3}$ collector currents, preventing second harmonic distortion. Tail source $Tr_{1,14}$ has a basic PSRR 10 dB better than the usual two diode version, though this is academic when C_{11} is fitted.

Input resistor R_1 and feedback arm R_8 are made equal and kept as low as possible consistent with a reasonably high input impedance, so that base current mismatch caused by beta variations will give a minimal DC offset. This does not affect $Tr_2 - Tr_3$ V_{be} mismatches, which appear directly at the output, but these are much smaller than the effects of I_b. Even if $Tr_{2,3}$ are high voltage types with low beta, the output offset should be within ±50mV, which should be quite adequate, and eliminates balance presets and DC servos. A low value for R_8 also gives a low value for R9, which improves the noise performance.

The value of C_2 shown (220 µF) gives an LF roll off with R_9 that is -3dB at 1.4 Hz. The aim is not an unreasonably extended sub-bass response, but to prevent an LF rise in distortion due to capacitor non linearity.

For example, 100 µF degraded the THD at 10 Hz from less than 0.0006% to 0.0011%. Band limiting should be done earlier, with non electrolytic capacitors. Protection diode D_1 prevents damage to C_2 if the amplifier suffers a fault that makes it saturate negatively; it looks unlikely but causes no measurable distortion.[10] C_7 provides some stabilising phase advance and limits the closed loop bandwidth; R_{20} prevents it upsetting Tr_3.

The VAS stage is enhanced by an emitter follower inside the Miller compensation loop, so that the local NFB which linearises the VAS is increased by augmenting total VAS beta, rather than by increasing the collector impedance by cascoding. The extra local NFB effectively eliminates VAS nonlinearity (Distortion 2).

Increasing VAS beta like this presents a much lower collector impedance than a cascode stage due to the greater local feedback. The improvement appears to make a VAS buffer to eliminate Distortion 4 (loading of VAS collector by the nonlinear input impedance of the output stage) unnecessary. C_{dom} is relatively high at 100 pF, to swamp transistor internal capacitances and circuit strays, and make the design predictable. The slew rate calculates as 40V/µsec. The VAS collector load is a standard current source, to avoid the uncertainties of bootstrapping.

Quiescent current stability
Since almost all the THD from a *blameless* amplifier is crossover, keeping the quiescent optimal is essential. Quiescent stability requires the bias generator to cancel out the V_{be} variations of four junctions in series; those of two drivers and two output devices. Bias generator Tr_{13} is the standard V_{be} multiplier, modified to make its voltage more stable against variations in the current through it. These occur because the biasing of Tr_5 does not completely reject rail variations: its output current drifts initially due to heating thus changing its V_{be}. Keeping a Class B quiescent stable is hard enough at the best of times, and so it makes sense to keep these extra factors out of the equation.

The basic V_{be} multiplier has an incremental resistance of about 20 Ω; in other words its voltage changes by 1 mV for a 50µA drift in standing current. Adding R_{14} converts this to a gently peaking characteristic that can be made perfectly flat at one chosen current; see Figure 5. Setting R_{14} to 22 Ω makes the voltage peak at 6 mA, and standing current now must deviate from this value by more than 500 µA for a 1 mV bias change.

BIASGEN3.CIR Class–B transistor bias generatory current-compensate R
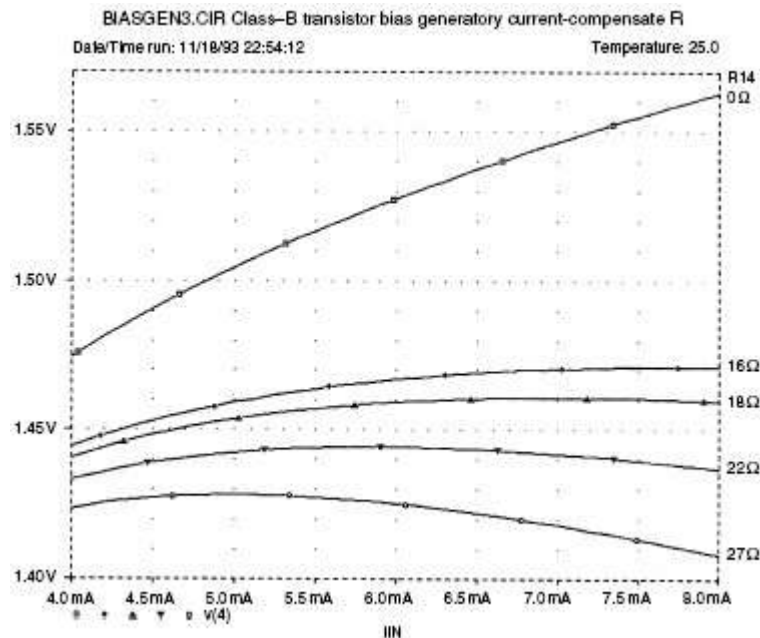Date/Time run: 11/18/93 22:54:12   Temperature: 25.0

Figure 5 Spice plot of the voltage-peaking behaviour of a current-compensated bias generator.

The R_{14} value needs to be altered if Tr_5 is run at a different current. For example, 16 Ω makes the voltage peak at 8 mA instead. If TO3 outputs are used, the bias generator should be in contact with the top or can of one of the output devices, rather than the heatsink, as this is the fastest and least attenuated source for thermal feedback.

Output stage
The output stage is a standard double emitter follower apart from the connection of R_{15} between the driver emitters without connection to the output rail. This gives quicker and cleaner switch-off of the outputs at high frequencies; this may be significant from 10 kHz upwards dependent on transistor type. Speed up capacitor C_5 improves the switch-off action. C_6, R_{18} form the Zobel network while L_1, damped by R_{19}, isolates the amplifier from load capacitance.

Figure 6 shows the 50W/8 Ω distortion performance, about 0.001% at 1 kHz, and 0.006% at 10 kHz. The measurement bandwidth makes a big difference to the appearance, because what little distortion is present is crossover derived, and so high order. It rises at 6 dB/octave, the rate at which feedback factor falls. The crossover glitches emerge from the noise, like Grendel from the marsh, as the test frequency increases above 1 kHz. There is no precipitous THD rise in the ultrasonic region, and so I suppose this might be called a high speed amplifier.
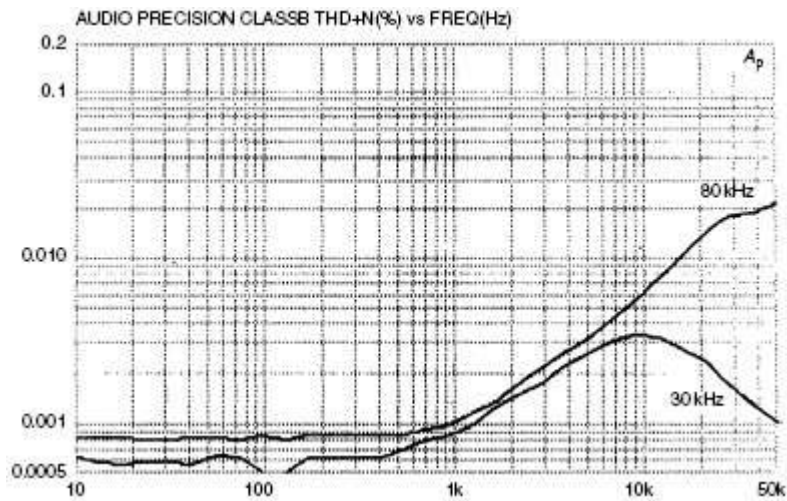
Figure 6 Class B amplifier: THD performance at 50 W/8-ohm; measurement bandwidths 30 kHz and 80 kHz.

Note that the zigzags on the LF end of the plot are measurement artifacts, apparently caused by the Audio Precision system trying to winkle distortion from visually pure white noise. Below 700 Hz the residual was pure noise with a level equivalent to approx 0.0006% at 30 kHz bandwidth. The actual THD here must be microscopic.

This performance can only be obtained if all seven of the distortion mechanisms are properly addressed; Distortions 1″4 are determined by the circuit design, but the remaining three depend critically on physical layout and grounding topology.

Figure 7 shows the startling results of applying two pole compensation to the amplifier. C_3 remains 100 pF, while C_{p2} was 220 pF and R_P 1kΩ. The extra NFB does its work extremely well, the 10 kHz THD dropping to 0.0015%, while the 1 kHz figure can only be guessed at. There were no unusual signs of instability on the bench, but I have not tried a wide range of loads.

This experimental amplifier was rebuilt with three alternative output stages: the simple quasi-complementary, the quasi-Baxandall and the CFP. The results for both single and two pole compensation are shown in Figures 8, 9, and 10. The simple quasi-complementary generates more crossover distortion, as expected, and the quasi-Baxandall version is not a lot better, due to remaining asymmetry around the crossover region. The CFP gives even lower distortion than the original EF-II output. Figure 10 shows only the result for single pole compensation; in this case the improvement with two pole was marginal and the trace is omitted for clarity.
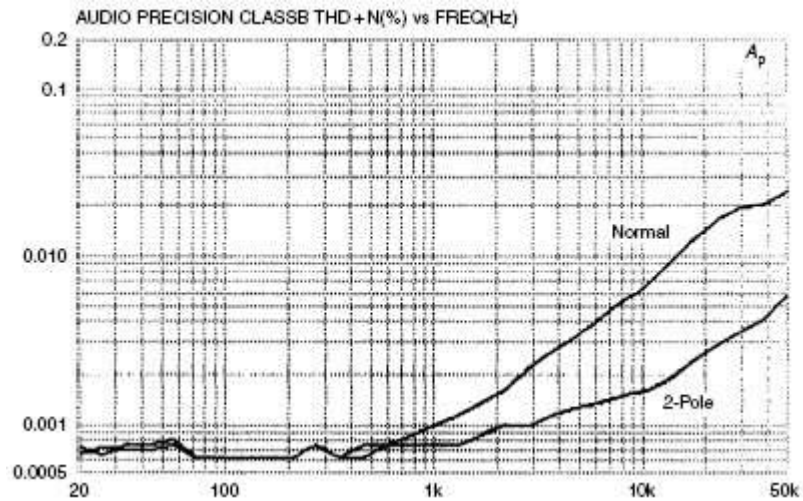
Figure 7 The dramatic THD improvement obtained by converting the Class B amplifier to two pole.
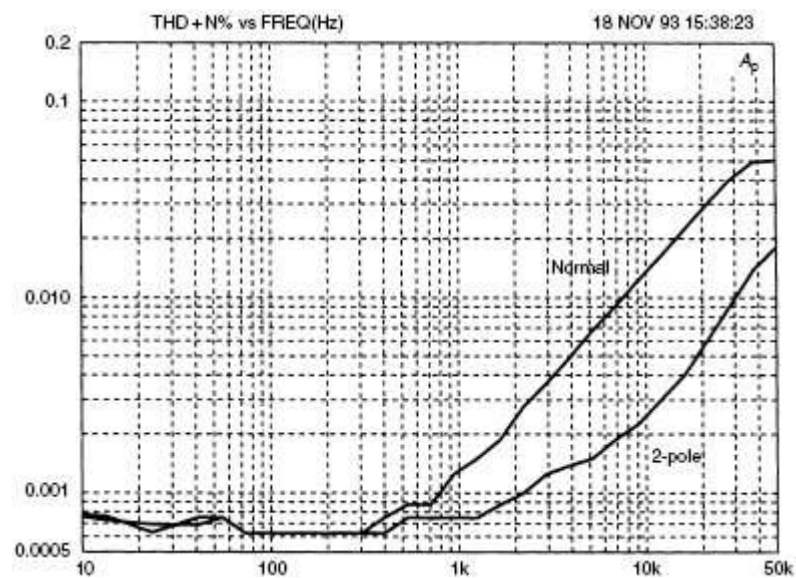

Figure 8 Class B amplifier with simple quasi-complementary output. Lower trace is for two pole compensation (80 kHz bandwidth).
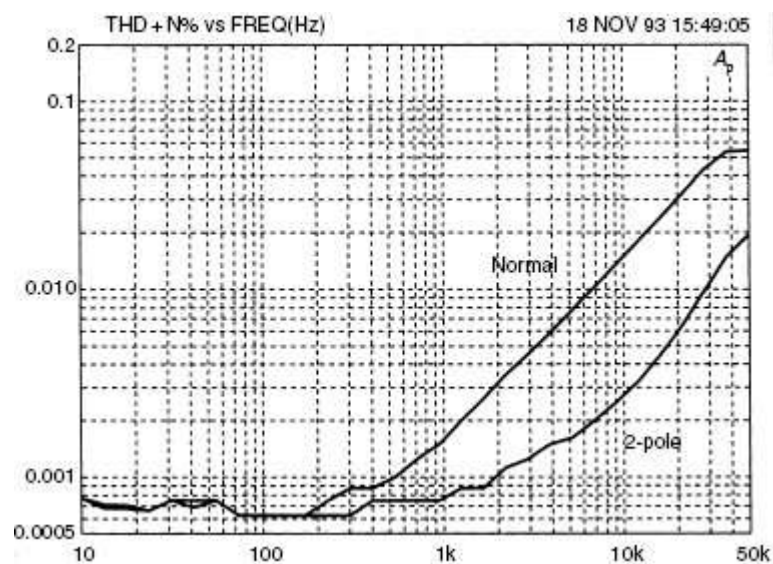
Figure 9 Class B amplifier with quasi-complementary plus Baxandall diode output. Lower trace is the two pole case (80kHz bandwidth).

Figure 10 Class B amplifier with complementary feedback pair (CFP) output stage.

The AP plots in earlier parts of this series were mostly done with an amplifier similar to Figure 4, though of higher power. Main differences were the use of a cascode-VAS with a buffer, and a CFP output to minimise distracting quiescent variations. Measurements at powers above 100 W/8Ω used a version with two paralleled output devices.
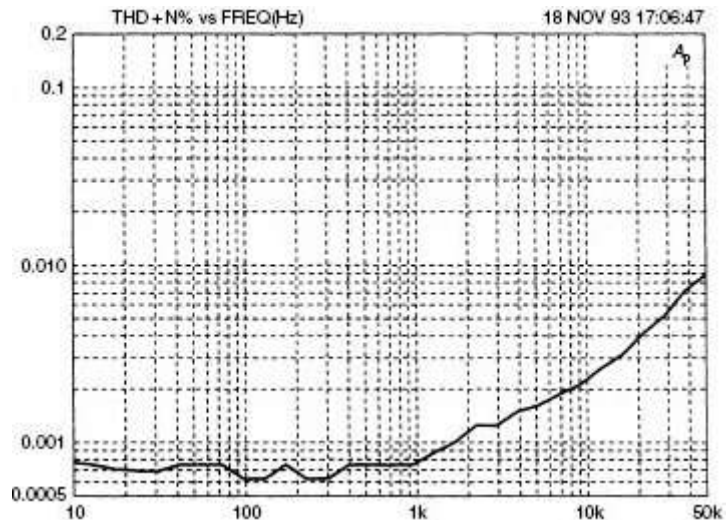
Coming up in <u>part 8</u>: Class-A amplifiers.

References

1. Otala, 'An audio power amplifier for ultimate quality requirements', *IEEE Trans on Audio & Electroacoustics* , Vol AU-21, No. 6, December 1973.
2. Baxandall, 'Audio power amplifier design. Part 4', *Wireless World* , July 1978, p 76.
3. Takahashi *et al* . 'Design & construction of high slew-rate amplifiers', *AES* 60th Convention, Preprint No. 1348 (A-4) 1978.
4. Self, 'Crossover Distortion & Compensation', Letters, *Electronics & Wireless World* , August 1992, p 657.
5. Widlar, 'A monolithic power op-amp', *IEEE Journal of Solid-State Circuits* , Vol 23, No 2, April 1988.
6. Bonello, 'Advanced negative feedback design for high performance amplifiers', *AES* 67th Convention, Preprint No. 1706 (D-5) 1980.
7. Pernici, 'A CMOS low-distortion amplifier with double-nested et al miller compensation', *IEEE Journal of Solid-State Circuits* , July 1993, p 758.
8. 'Fast compensation extends power bandwidth', Linear Brief 4, National Semiconductor Linear Applications Handbook 1991.
9. Feucht, *Handbook of Analog Circuit Design* , Academic Press 1990, p 264.
10. Self, 'An Advanced Preamplifier', *Wireless World* , November 76, p 43.

Read the whole series: