Distortion in power amplifiers, Part IV: the power amplifier stages

EDN edn.com/distortion-in-power-amplifiers-part-iv-the-power-amplifier-stages/

January 23, 2008

[Part 1 offers an overview and introduction to the sources of distortion in power amplifiers. Part 2 focuses on distortion in the audio amplifier input stage. Part 3 examines distortion mechanisms in the voltage amplifier stage (VAS) of audio power amplfiers.]

The almost universal choice in semiconductor power amplifiers is for a unity gain output stage, and specifically a voltage follower. Output stages with gain are not unknown,[1] but they are not common. Most designers feel that controlling distortion while handling large currents is hard enough without trying to generate gain at the same time.

The first three parts of this series have dealt with one kind of distortion at a time, due to the monotonic transfer characteristics of small signal stages, which usually, but not invariably, work in class A.[2] Economic and thermal realities mean that most output stages are class B, and so we must now consider crossover distortion, which remains the thorniest problem in power amplifier design, and HF switchoff effects.

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

It is now also necessary to consider what kind of active device is to be used; jfets offer few if any advantages in the small current stages, but power fets are a real possibility, providing that the extra cost brings with it tangible benefit.

The class war

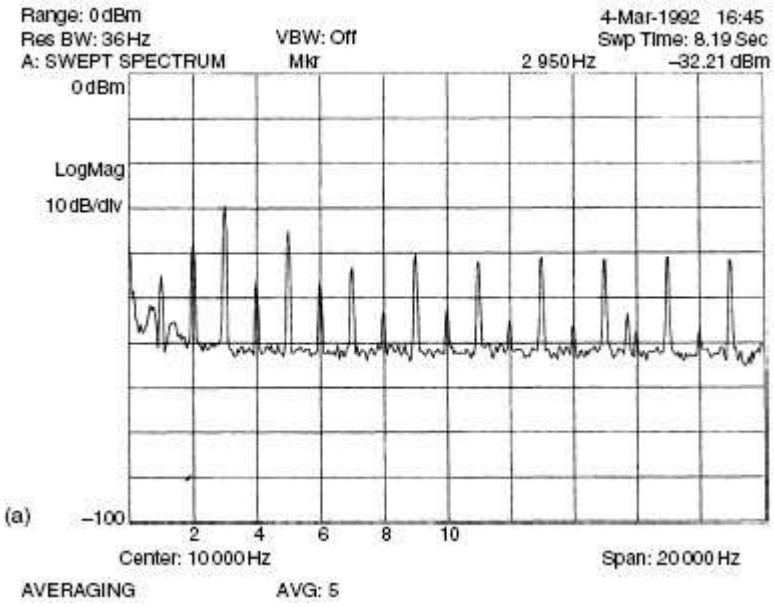
The fundamental factor in determining output stage distortion is the class of operation. Apart from its inherent inefficiency, class A is the ideal operating mode, because there can be no crossover or switchoff distortion.

However, of those designs which have been published or reviewed, it is notable that the large signal distortion produced is still significant. This looks like an opportunity lost, as of the distortion mechanisms discussed in the first part of this series, we now only have to deal with Distortion 1 (input stage), Distortion 2 (VAS), and Distortion 3 (output stage large signal nonlinearity). Distortions 4, 5, 6 and 7, as mentioned earlier, are direct results of class B operation and therefore can be thankfully disregarded in a class A design. However, class B is overwhelmingly of the greater importance, and is therefore dealt with in detail.

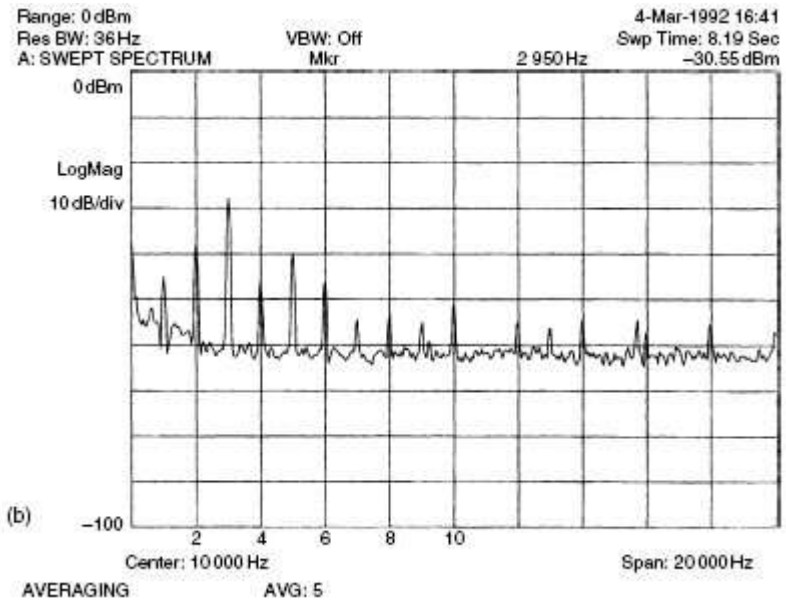
Class B is subject to misunderstanding. The statement is often made that a pair of output transistors operated without any bias are working in 'class B', and therefore 'generate severe crossover distortion'. In fact, with no bias each output device is operating for slightly less than half the time, and the question arises as to whether it would not be more accurate to call this class C and reserve class B for that condition of quiescent current which eliminates, or rather minimises, the crossover artifacts.

A further complication exists; it is not generally appreciated that moving into what is usually called class AB, by increasing the quiescent current, does not make things better. In fact, the THD reading will increase as the bias control is advanced, with what is usually known as 'g_m doubling' (i.e. a voltage gain increase caused by both devices conducting simultaneously in the centre of the output voltage range) putting edges into the distortion residual that generate high order harmonics in much the same way that underbiasing does. This important fact seems almost unknown, presumably because the g_m doubling distortion is at a relatively low level and is completely obscured in most amplifiers by other distortion mechanisms.

The phenomenon is demonstrated in *Figure 1*, which shows spectrum analysis of the distortion residuals for under biasing, optimal, and over biasing of a 150W/8 Ω amplifier at 1 kHz.


Range: 0 dBm
Res BW: 36 Hz
A: SWEPT SPECTRUM
VBW: Off
Mkr
2 950 Hz
4-Mar-1992 16:45
Swp Time: 8.19 Sec
−32.21 dBm
0 dBm
LogMag
10 dB/div
−100
2
4
6
8
10
Center: 10 000 Hz
Span: 20 000 Hz
AVERAGING
AVG: 5

(a)


Range: 0 dBm
Res BW: 36 Hz
A: SWEPT SPECTRUM
VBW: Off
Mkr
2 950 Hz
4-Mar-1992 16:41
Swp Time: 8.19 Sec
−30.55 dBm
0 dBm
LogMag
10 dB/div
−100
2
4
6
8
10
Center: 10 000 Hz
Span: 20 000 Hz
AVERAGING
AVG: 5

(b)

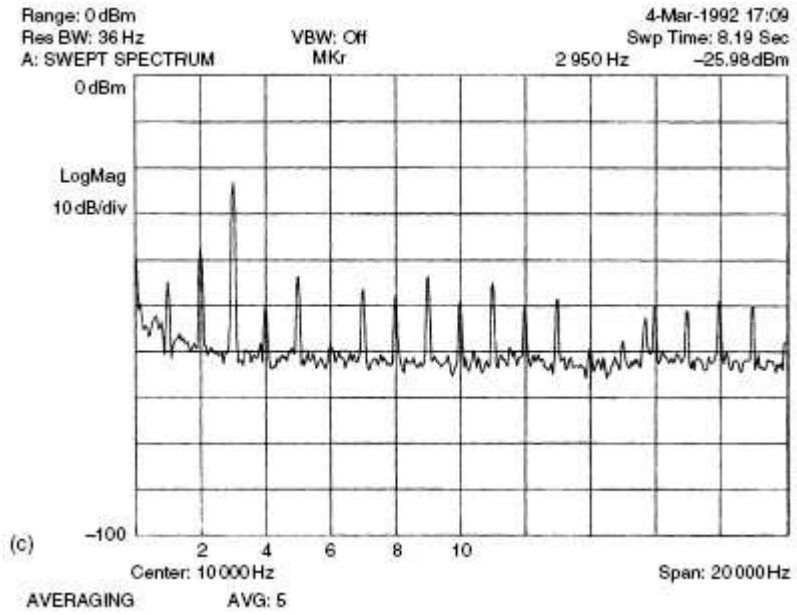


Figure 1 Spectrum analysis of class B & AB distortion residual. 1(a) Underbiased class B; 1(b) Optimal class B; 1(c) class AB.

As before, all non-linearities except the unavoidable Distortion 3 (output stage) have been effectively eliminated. The over biased case had its quiescent current increased until the g_m doubling edges in the residual had an approximately 50:50 mark/space ratio, and so was in class A about half the time which represents a rather generous amount of quiescent for class AB. Nonetheless, the higher order odd harmonics in *Figure 1(c)* are at least 10 dB greater in amplitude than those for the optimal class B case, and the third harmonic is actually higher than for the under-biased case as well. However the under biased amplifier, generating the familiar sharp spikes on the residual, has a generally greater level of high-order odd harmonics above the 5th; about 8 dB higher than the AB case.

Bearing in mind that high order odd harmonics are generally considered to be the most unpleasant, there seems to be a clear case for avoiding Class AB altogether, as it will always be less efficient and generate more high order distortion than the equivalent class B circuit, class distinction therefore seems to resolve itself into a binary choice between A or B.

It must be emphasised that these effects can only be seen in an amplifier where the other forms of distortion have been properly minimised. The r.m.s. THD reading for case **1a** was 0.00151%, for case **1b** 0.00103%, and for case **1c** 0.00153%. The tests were repeated at the 40 W power level with very similar results. The spike just below 16 kHz is interference from the test gear VDU.

This may seem complicated enough, but there are other and deeper subtleties in class B.

Distortions of the output

I have designated the distortion produced directly by output stages as Distortion 3 (see Part 1); this subdivides into three categories. Mechanism 3a describes the large signal

distortion produced by both class A and B, ultimately because of the large current swings in the active devices. In bipolars, but not fets, large collector currents reduce the beta leading to drooping gain at large output excursions. I shall use the term ‘LSN’ for large signal nonlinearity, as opposed to crossover and switchoff phenomena that cause trouble at all output levels.

The other two contributions to Distortion 3 are associated with class B only; Distortion 3b the classic crossover distortion resulting from the non-conjugate nature of the output characteristics, and is essentially frequency independent.

In contrast, Distortion 3c is switchoff distortion generated by the output devices failing to turn off quickly and cleanly at high frequencies. This mechanism is strongly frequency dependent. It is sometimes called switching distortion, but this allows room for confusion, as some writers use ‘switching distortion’ to cover crossover distortion as well. I refer specifically to charge storage turn off troubles.

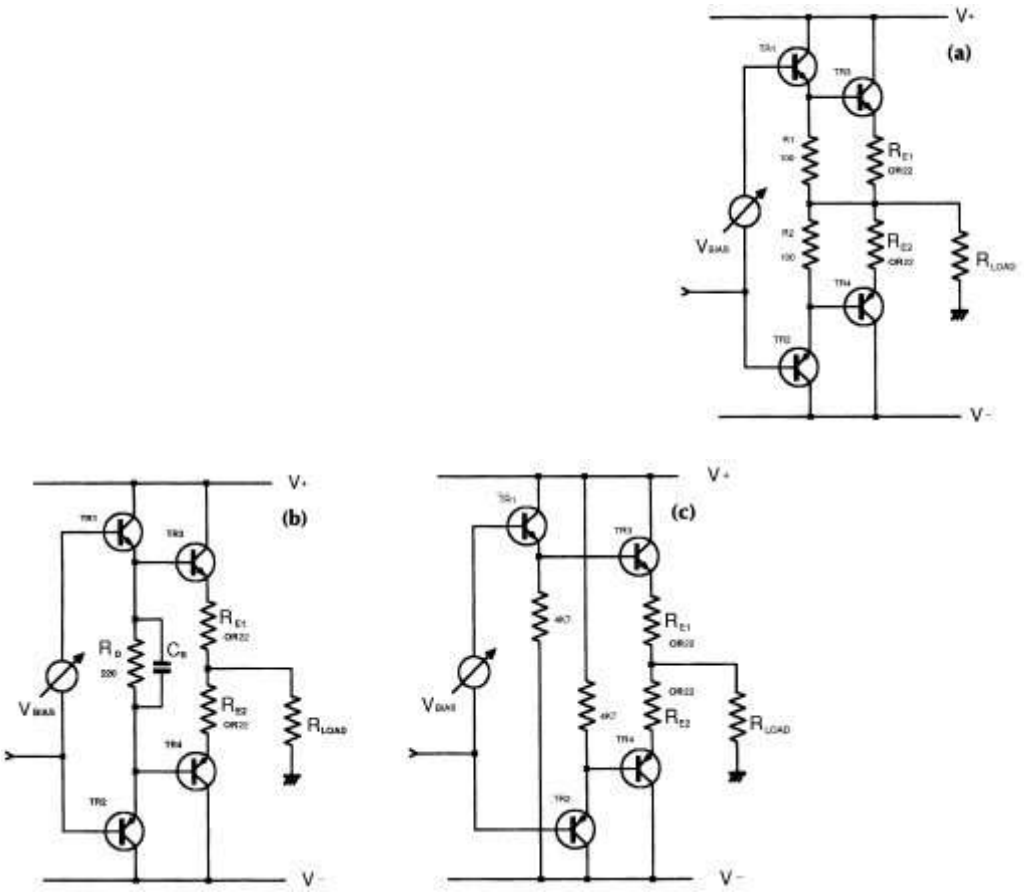


Figure 2 Three types of emitter follower output stages.

One of my aims for this series has been to show how to isolate individual distortion mechanisms. To examine output behaviour, it is perfectly practical to drive output stages open loop providing the driving source impedance is properly specified; this is difficult with a conventional amplifier, as it means the output must be driven from a frequency dependant impedance simulating that at the VAS collector with some sort of feedback mechanism incorporated to keep the drive voltage constant.

However, if the VAS is buffered from the output stage by some form of emitter follower, as described in the last part, it makes things much simpler, a straightforward low impedance source (e.g. 50 Ω) providing a good approximation of a VAS-buffered closed loop amplifier. The VAS buffer makes the system more designable by eliminating two variables ” the VAS collector impedance at LF, and the frequency at which it starts to decrease due to local feedback through C_{dom}. This markedly simplifies the study of output stage behaviour.

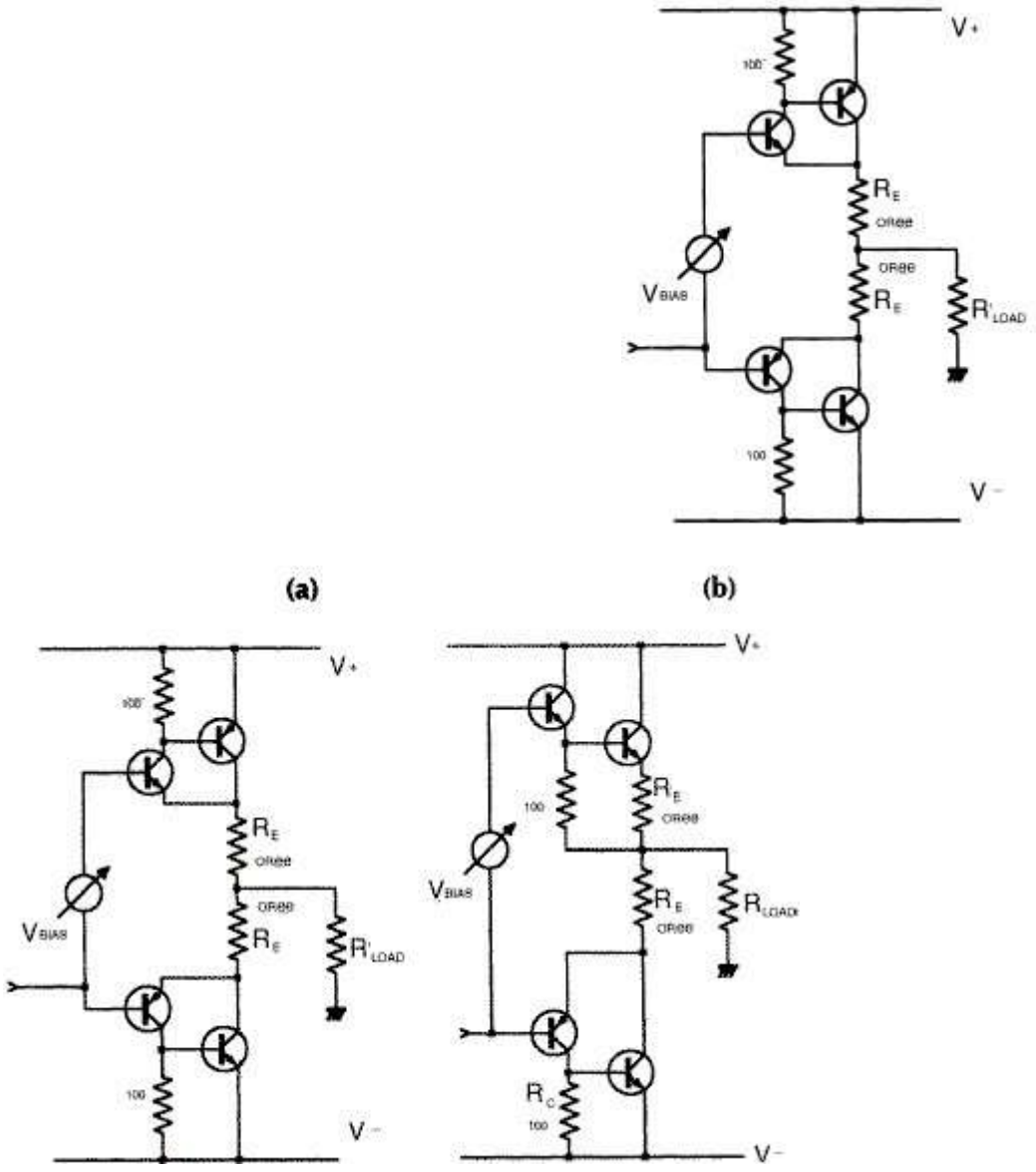


Figure 3 CFP circuit and quasi-complementary stages.

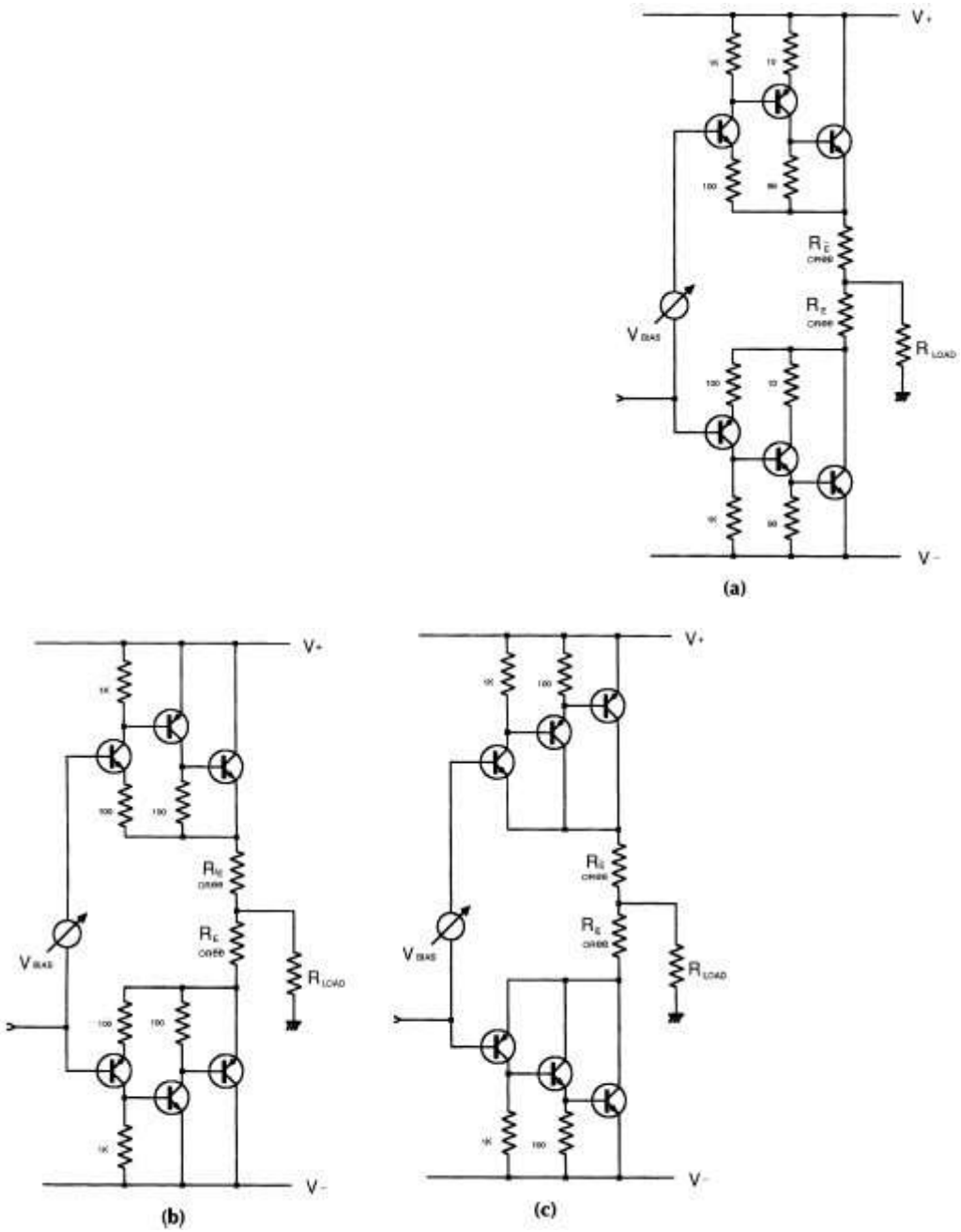


Figure 4 Three of the possible output triple configurations.

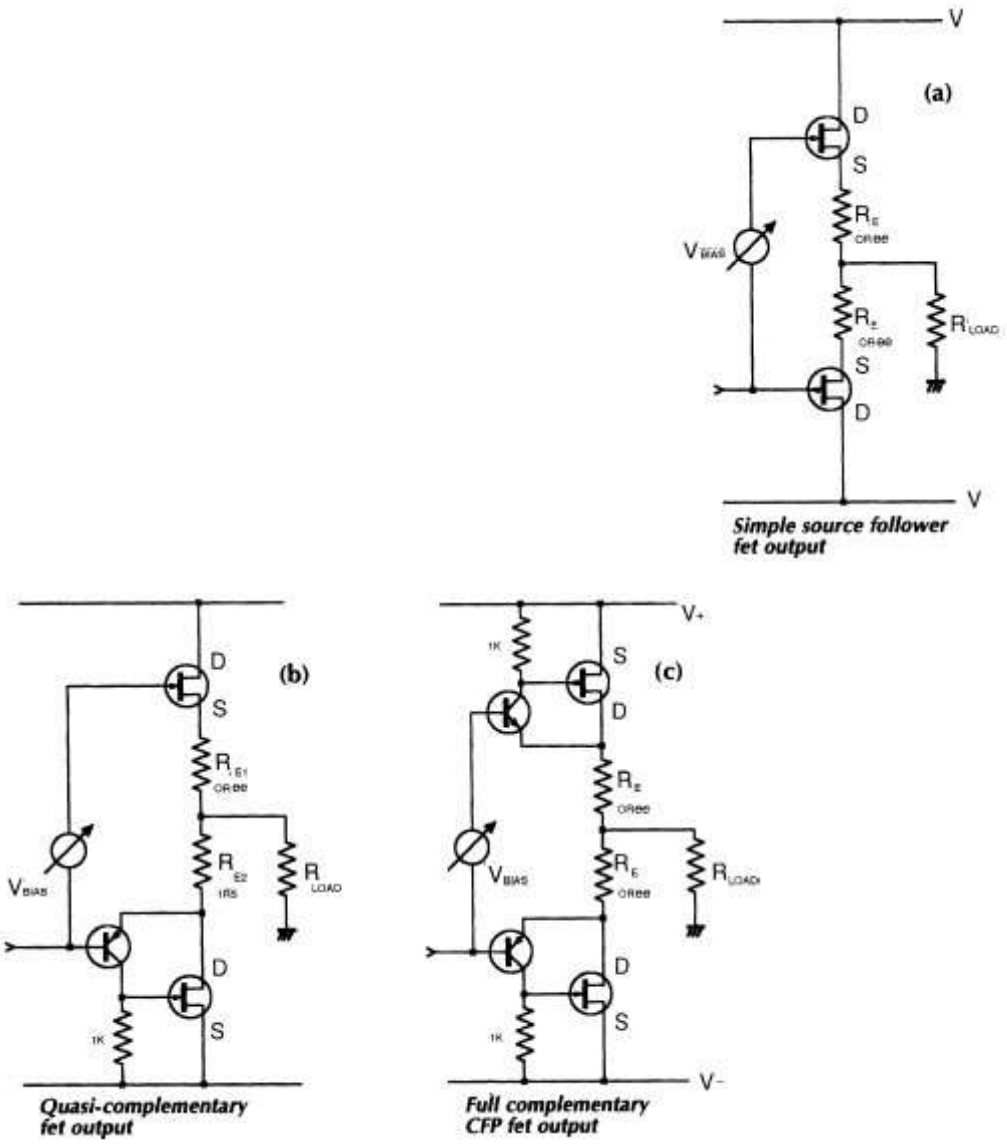


Figure 5 Three mosfet output architectures.

The large signal linearity of various kinds of open loop output stage with typical values are shown in Figures 6-15. These diagrams were all generated by spice simulation, and are plotted as incremental output gain against output voltage, with the load resistance stepped from 16 Ω to 2 Ω. The power devices are *MJ802* and *MJ4502* , which are more complementary than many transistor pairs, and minimise distracting large signal asymmetry. The quiescent current is in each case set to minimise the peak deviations of gain around the crossover point for 8Ω loading; for the moment it is assumed that you can set this accurately and keep it where you want it. The difficulties in actually doing this will be examined later.

There are at least 16 distinct configurations in straightforward output stages not including error correcting,[3] current dumping[4] or Blomley[5] types. These are as follows:

Emitter Follower	3 types	Figure 2
Complementary Feedback Pair	1 type	Figure 3

Quasicomplementary	2 types	Figure 3
Output Triples	At least 7 types	Figure 4
Power FET	3 types	Figure 5

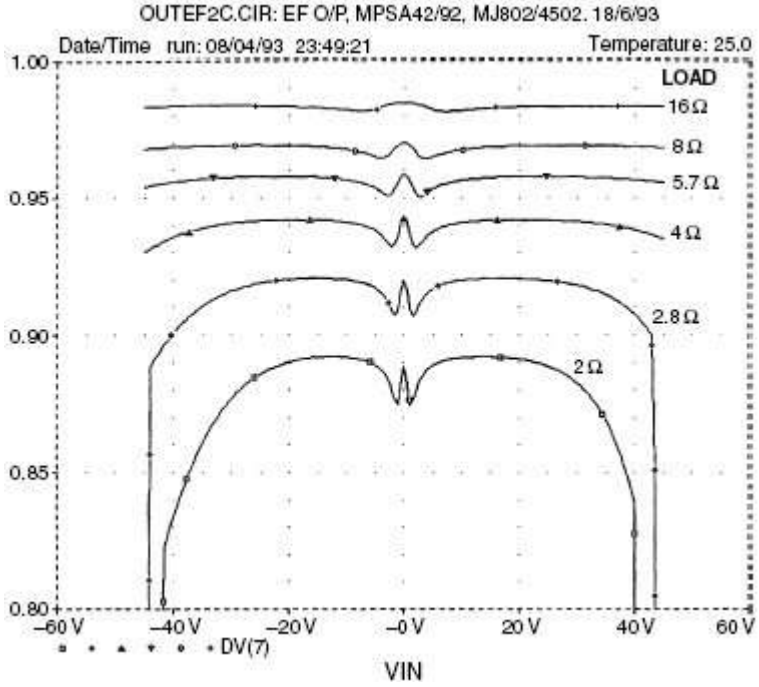
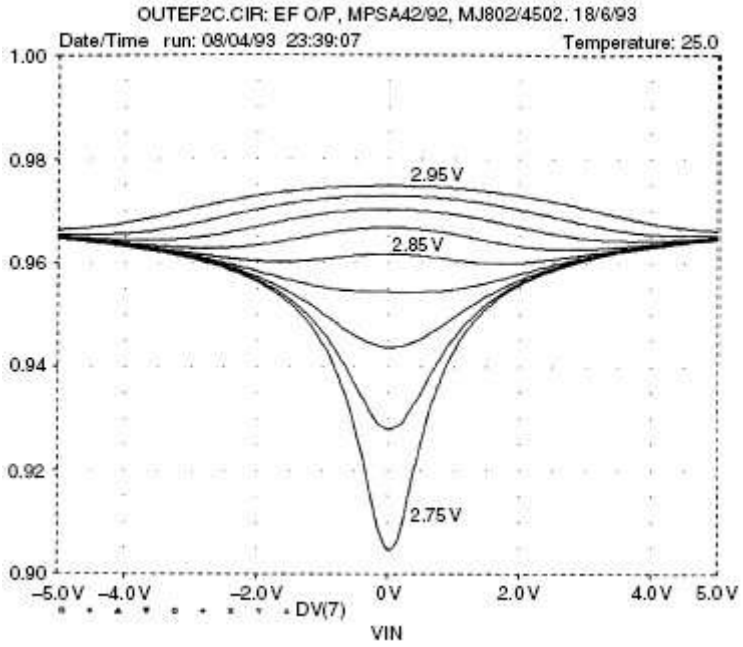


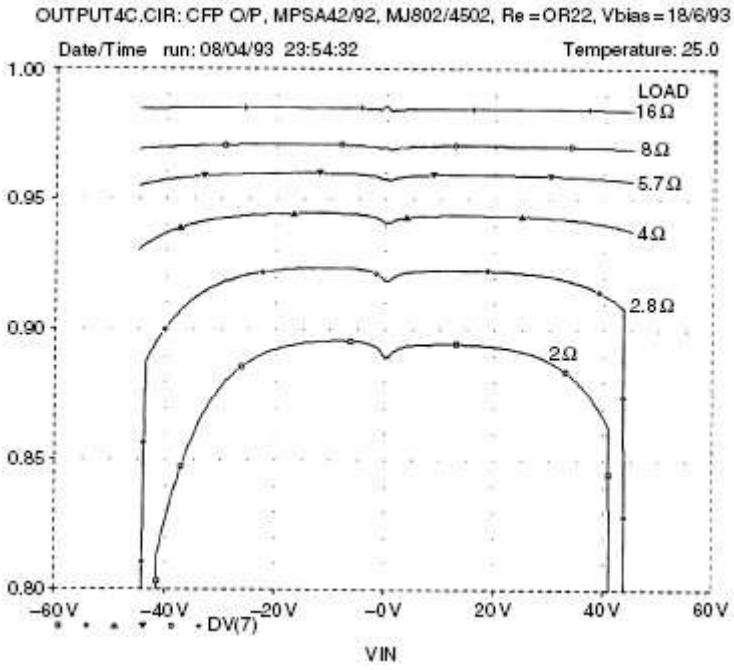
Figure 6 Emitter follower large signal gain vs output.

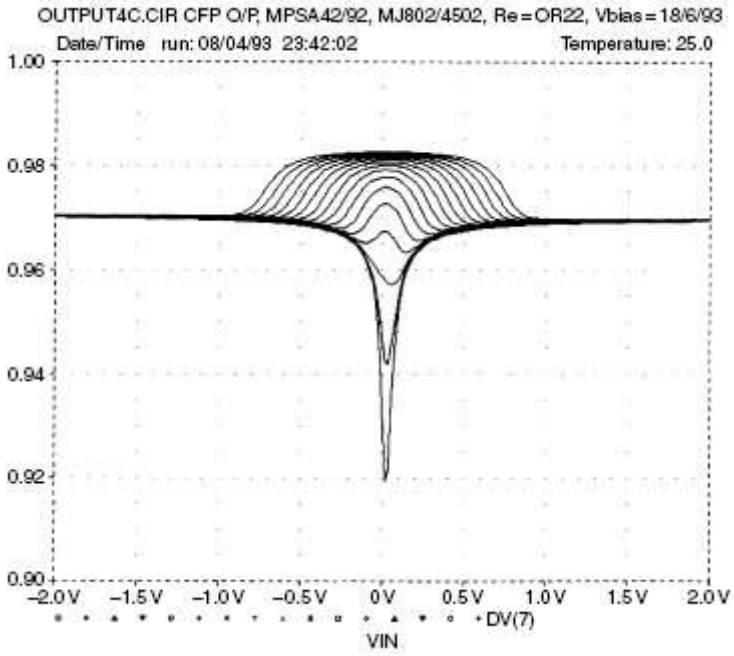
Figure 7 Emitter follower crossover region gain deviations, ±5V range.

Figure 8 Complementary feedback pair gain versus output.

Figure 9 CFP crossover region ±2V, V_{bias} as a parameter.




OUTPUT4C.CIR: CFP O/P, MPSA42/92, MJ802/4502, Re = OR22, Vbias = 18/6/93
Date/Time run: 08/04/93 23:54:32
Temperature: 25.0
LOAD
16Ω
8Ω
5.7Ω
4Ω
2.8Ω
2Ω
1.00
0.95
0.90
0.85
0.80
−60V
−40V
−20V
−0V
20V
40V
60V
DV(7)
VIN


OUTPUT4C.CIR CFP O/P, MPSA42/92, MJ802/4502, Re = OR22, Vbias = 18/6/93
Date/Time run: 08/04/93 23:42:02
Temperature: 25.0
1.00
0.98
0.96
0.94
0.92
0.90
−2.0V
−1.5V
−1.0V
−0.5V
0V
0.5V
1.0V
1.5V
2.0V
DV(7)
VIN

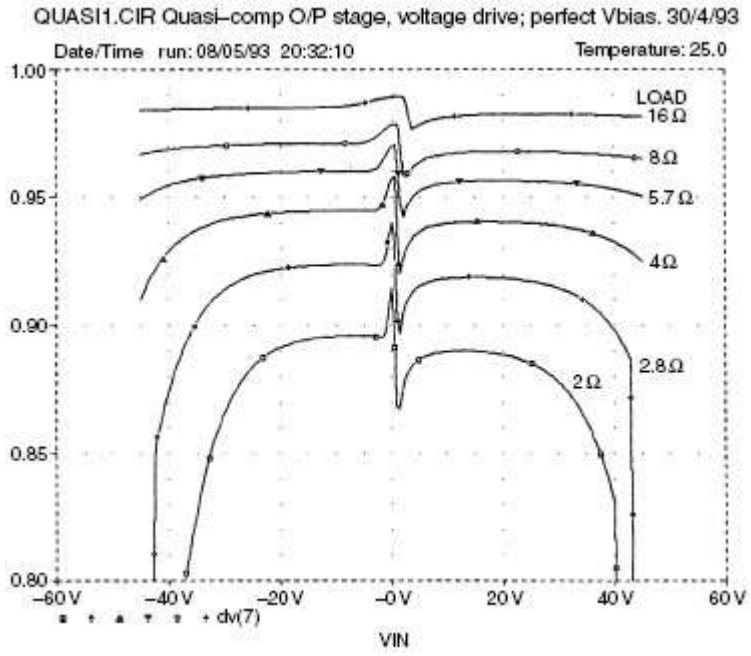
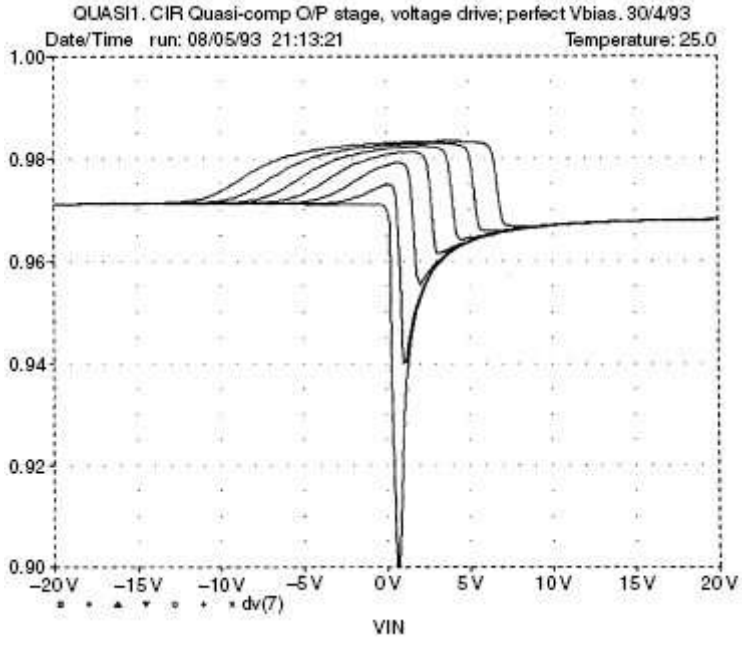
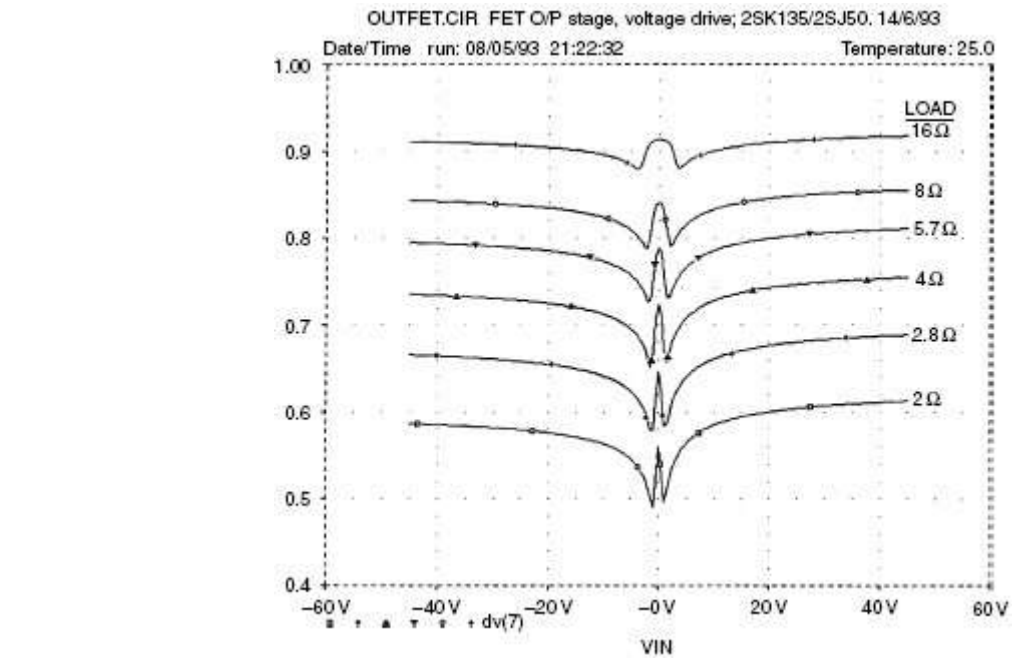


Figure 10 Quasi complementary large signal gain vs output load resistance.

Figure 11 Quasi crossover region ±20V, V_{bias} as parameter.

Figure 12 Source follower FET large signal gain vs output.





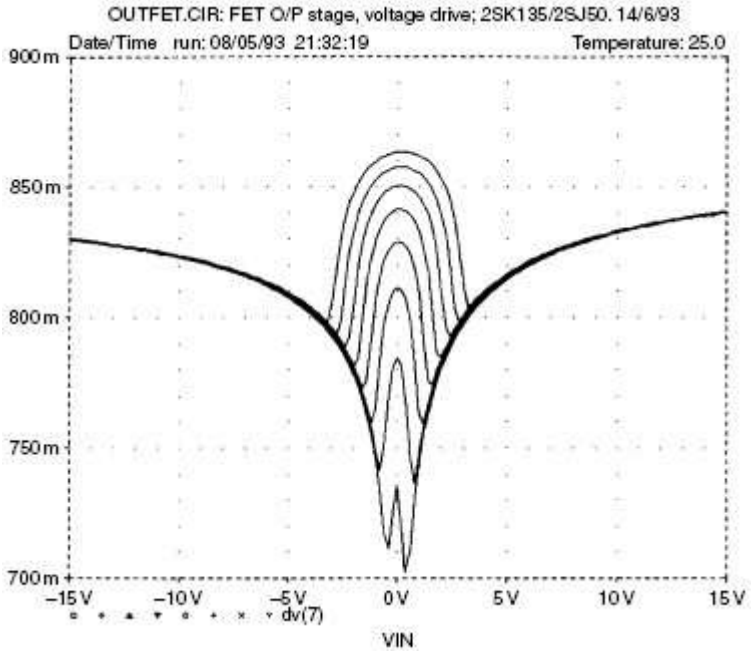


Figure 13 Source follower FET crossover region ±15 V range.

Figure 14: Complementary bipolar FET gain vs output.

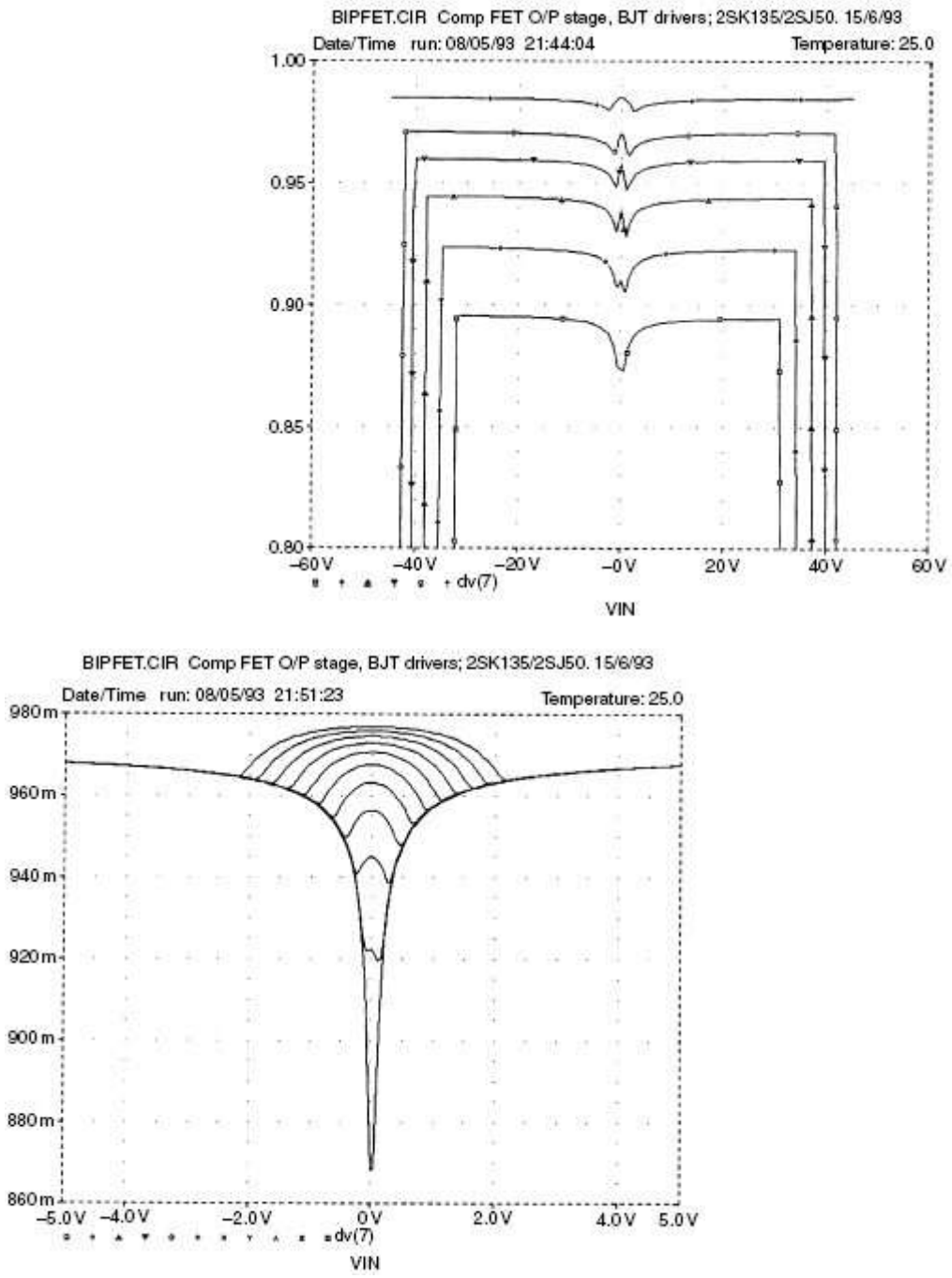


Figure 15 Complementary BJT FET crossover region ±15 V range.

The emitter follower output

Figure 2 shows three versions of the most common type of output stage; the double-emitter follower where the first follower acts as driver to the second (output) device. I have deliberately called this an emitter follower rather than a Darlington configuration, as this latter implies an integrated device with associated resistors. As for all the circuitry here, the component values are representative of real practice.

Two important attributes of this topology are:

- The input is transferred to the output via two base emitter junctions in series, with no local feedback around the stage (apart from the very local 100% voltage feedback that makes an emitter follower what it is);

- There are two dissimilar base emitter junctions between the bias voltage and the emitter resistor R_e, carrying different currents and at different temperatures. The bias generator must attempt to compensate for both at once, though it can only be thermally coupled to one. The output devices have substantial thermal inertia and thus thermal compensation represents a time average of the preceding conditions. *Figure 2(a)* shows the most prevalent version (type I) which has its driver emitter resistors connected to the output rail.

The type II configuration in *Figure 2(b)* is at first sight merely a pointless variation on type I, but in fact it has a valuable extra property. The shared driver emitter resistor R_d, with no output rail connection, allows the drivers to reverse bias the base emitter junction of the output device being turned off.

Assume that the output voltage is heading downwards through the crossover region; the current through R_{e1} has dropped to zero, but that through R_{e2} is increasing, giving a voltage drop across it, so Tr_4 base is caused to go more negative to get the output to the right voltage. This negative excursion is coupled to Tr_3 base through R_d, and with the values shown can reverse bias it by up to 0.5 V, increasing to 1.6 V with a 4 Ω load. The speed up capacitor C_s markedly improves this action, preventing the charge suckout rate being limited by the resistance of R_d.

While the type I circuit has a similar voltage drop across R_{e2}, the connection of the mid point of R_1, R_2 to the output rail prevents this from reaching Tr_3 base; instead Tr_1 base is reverse biased as the output moves negative, and since charge storage in the drivers is usually not a problem, this does little good. In the type II circuit the drivers are never reverse biased, though they do turn off. The important issue of output turn off and switching distortion is further examined in the next part of this series.

The type III topology shown in *Figure 2(c)* maintains the drivers in class A by connecting the driver emitter resistors to the opposite supply rail rather than the output rail. It is a common misconception6 that class A drivers somehow maintain better low frequency control over the output devices, but I have yet to substantiate any advantage myself. The driver dissipation is of course substantially increased, and nothing seems to be gained at LF as far as the output transistors are concerned, for in both type I and type II the drivers are still conducting at the moment the outputs turn off, and are back in conduction before the outputs turn on, which would seem to be all that matters.

Type III is equally good as type II in reverse biasing the output bases, and may give even cleaner HF turn off as the carriers are being swept from the bases by a higher resistance terminated in a higher voltage approximating constant current drive; I have yet to try this.

The large signal linearity of the three versions is virtually identical; all have the same feature of two base emitter junctions in series between input and load.

The gain/output voltage plot is shown at *Figure 6*; with BJTs the gain reduction with increasing loading is largely due to the emitter resistors. Note that the crossover region appears as a relatively smooth wobble rather than a jagged shape. Another major feature is the gain droop at high output voltages and low loads indicating that high collector currents are the fundamental cause of this.

A close up of the crossover region gain for 8 Ω loading only is shown in *Figure 7*; note that no V_{bias} setting can be found to give a constant or even monotonic gain; the double dip and central gain peak are characteristic of optimal adjustment. The region extends over about ±5V, independent of load resistance.

Complementary feedback output

The other major type of bipolar output is the complementary feedback pair (CFP) sometimes called the Sziklai Pair, *Figure 3(a)*. There seems to be only one popular configuration, though versions with gain are possible. The drivers are now placed so that they compare output voltage with that at the input. Wrapping the outputs in a local negative feedback loop promises better linearity than emitter follower versions with 100% feedback applied separately to driver and output transistors.

This topology also has better thermal stability, because the V_{be} of the output devices is inside the local feedback loop, and only the driver V_{be} affects the quiescent current. It is usually simple to keep drivers cool, and thermal feedback from them to the V_{bias} generator transistor can be much faster and mechanically simpler.

Like emitter follower outputs, the drivers are conducting whenever the outputs are, and so special arrangements to keep them in class A seem pointless. This stage, like emitter follower type I, can only reverse bias the driver bases rather than the outputs, unless extra voltage rails outside the main ones are provided.

The output gain plot is shown in *Figure 8*. Fourier analysis shows that the CFP generates less than half the large signal distortion of an emitter follower stage. (See *Table 1*) Given also the greater quiescent stability, it is hard to see why this topology is not more popular.

The crossover region is much narrower, at about ±0.3V (*Figure 9*). When under biased, this shows up on the distortion residual as narrower spikes than an emitter follower output gives.

The bad effects of g_m doubling as V_{bias} increases above optimal (here 1.296 V) can be seen in the slopes moving outwards from the centre.

Quasicomplementary outputs

The original quasicomplementary configuration7 was almost mandatory, as it was a long time before pnp silicon power transistors matched the performance of the npn versions. The standard version shown at *Figure 3(b)* is well known for poor symmetry around the crossover region, as shown at *Figure 10*. A close-up of the crossover region *(Figure 11)* reveals an unhappy hybrid of the emitter follower and CFP, as might be expected, and that no setting of bias voltage can remove the sharp edge in the gain plot.

A major improvement to symmetry may be made by using a Baxandall diode[8] as shown in *Figure 3(c)*. This stratagem yields gain plots very similar to those for the true complementary emitter follower at *Figures 6* , *7*, though in practice the crossover distortion seems rather higher. When a quasiBaxandall stage is used closed loop in an amplifier in which distortion mechanisms 1 and 2, and 4 to 7 have been properly eliminated, it is capable of better performance than is commonly believed. For example, 0.0015% (1 kHz) and 0.015% (10 kHz) at 100 W is straightforward to obtain from an amplifier with a negative feedback factor of about 34 dB at 20 kHz.

The best reason to use the quasiBaxandall approach today is to save money on output devices, as pnp power transistors remain somewhat pricier than npns. Given the tiny cost of a Baxandall diode, and the absolutely dependable improvement it gives, there seems no reason why anyone should ever use the standard quasi circuit. My experiments show that the value of R_1 in *Figure 3(c)* is not critical; making it about the same as R_c seems to work well.

Triples

With three rather than two bipolar transistors in each half of an output stage the number of circuit permutations possible leaps upwards. There are two main advantages if output triples are used correctly: better linearity at high output voltages and currents; and more stable quiescent setting as the predrivers can be arranged to handle very little power, and remain almost cold in use.

However, triples do not automatically reduce crossover distortion, and they are, as usually configured, incapable of reverse biasing the output bases to improve switch-off. *Figure 4* shows three ways to make a triple output stage; all of those shown (with the possible exception of *Figure 4(c)*, which I have just made up) have been used in commercial designs. The circuit of *4a* is the Quad *303* quasicomplementary triple. The design of triples demands care, as the possibility of local HF instability in each output half is very real.

Power FET outputs

Power mosfets are often claimed to be a solution to all amplifier problems, but they have their drawbacks: poor linearity and a high on-resistance that makes output efficiency mediocre. The high frequency response is better, implying that the second pole P2 of the amplifier response will be higher, allowing the dominant pole P1 be raised with the same stability margin, and in turn allowing more overall feedback to reduce distortion. However, the extra feedback (if it proves available in practice) is needed to correct the higher open loop distortion.

To complicate matters, the compensation cannot necessarily be lighter because the higher output resistance makes the lowering of the output pole by capacitive loading more likely.

The extended frequency response creates its own problems; the HF capabilities mean that rigorous care must be taken to prevent parasitic oscillation, as this is often promptly followed by an explosion of disconcerting violence. Fets should at least give freedom from

switchoff troubles as they do not suffer from charge storage effects.

Three types of FET output stage are shown in *Figure 5*. *Figures 12* to *15* show spice gain plots, using *2SK135/2SJ50* devices.

Most FET amplifiers use the simple source follower configuration in *Figure 5(a)*; the large signal gain plot at *Figure 12* shows that the gain for a given load is lower, (0.83 rather than 0.97 for bipolar, at 8Ω) because of low g_m. This, with the high on resistance, noticeably reduces output efficiency.

Open loop distortion is markedly higher; however large signal nonlinearity does not increase with heavier loading, there being no equivalent of 'bipolar gain droop'. The crossover region has sharper and larger gain deviations than a bipolar stage, and generally looks pretty nasty; *Figure 13* shows the difficulty of finding a 'correct' V_{bias} setting.

Figure 5(b) shows a hybrid (i.e. bipolar/FET) quasi complementary output stage.[9] The stage is intended to maximise economy rather than performance, once the decision has been made (probably for marketing reasons) to use fets, by making both output devices cheap n-channel devices; complementary mosfet pairs remain relatively rare and expensive.

The basic configuration is badly asymmetrical, the hybrid lower half having a higher and more constant gain than the source follower lower upper half. Increasing the value of R_{e2} gives a reasonable match between the gains of the two halves, but leaves a daunting crossover discontinuity.

The hybrid full complementary stage in *Figure 5(c)* was conceived[10] to maximise performance by linearising the output devices with local feedback and reducing I_q variations due to the low power dissipation of the bipolar drivers. It is highly linear, showing no gain droop at heavier loadings (*Figure 14*) and promises freedom from switchoff distortion. But, as shown, it is rather inefficient in voltage swing. The crossover region (*Figure 15*) still has some dubious sharp corners, but the total crossover gain deviation (0.96-0.97 at 8Ω) is much smaller than for the quasi hybrid (0.78-0.90) and so less high order harmonic energy is generated.

Part 5 continues looking at the output stage with a discussion including large-signal distortion, crossover distortion, switching distortion, and selecting an output stage.

References

1. Mann, R. 'The Texan 20 + 20 watt stereo amplifier'. *Practical Wireless* , May 1972, p 48.
2. Takahashi, 'Design & construction of high slew rate amplifiers', Preprint No. 1348 (A 4) for *60th AES Convention* , 1978.
3. Hawksford MO. 'Distortion correction in audio power amplifiers', *JAES* January/February 1981 p 27 (Error correction).
4. Walker, P. 'Current dumping audio amplifier', *Wireless World* , 1975 pp 60/62.

5. Blomley. ‘New approach to class B’. *Wireless World* , February 1971, p 57 and March 1971, p 127/131.
6. Otala. ‘An audio power amplifier for ultimate quality requirements’, *IEEE Trans on Audio & Electroacoustics* , December 1973, p 548.
7. Lin, H. *Electronics* , September 1956 pp 173/175.
8. Baxandall, P. ‘Symmetry in class B. Letters’, *Wireless World* , September 1969 p 416.
9. Self, D. ‘Sound mosfet design’. *Electronics & Wireless World* , September 1990, p 760.
10. Self D. ‘Mosfet audio output’, letter, *Electronics & Wireless World* , May 1989, p 524 (see also Ref. 9).

Read the whole series: