Designing a low-distortion audio output stage - Part 1: Introduction, the problem with push-pull outputs

Kendall Castor-Perry - September 27, 2012

This material originally appeared as one long article in Linear Audio Volume 2, published in September 2011. Linear Audio is a book-format audio magazine published half-yearly by Jan Didden.

Introduction
These articles introduce a systematic approach to the design of very low distortion push-pull output stage for audio power amplifiers. Work on this has been going on (well, on and off, to tell the truth) since February 2nd, 1981, according to my notebooks.

The results presented here focus on algebraic analysis and simulation. But a great deal was learned from producing working hardware especially in the initial years. Since this material first appeared in Linear Audio magazine, further feedback – or, rather, error correction! – has thrown additional light on implementation issues, and I’ve referred to this in the text where it’s helpful.

Push-me, pull-you: the core problems presented by push-pull output stages
Conventional push-pull output stages (the current-delivering, unity-gain back-ends of the vast majority of linear amplifier topologies at any power level) aren't very linear. What does that mean? It means that under load (i.e., when significant current of either polarity flows at the output terminal), an error voltage is created between input and output. This error voltage isn't generally related linearly to either the output voltage or the output current, and represents a significant source of non-linearity and of the consequent distortion to signals. Cordell's excellent recent book [5] has a comprehensive and contemporary study.

So, one goal of power stage design is to create a high output-current circuit that delivers high linearity before the soothing balm of global negative feedback (NFB) is applied. Poor linearity in the output stage is typically the driving factor for the choice of how much NFB to use.

Now, I'm not a "feedback denier," but I like to see it used to make an excellent amplifier 'excellenter' in some way, not to make a marginal amplifier meet a quantitative specification that may or may not have any impact on the sound quality. So I like to begin with an analysis of circuit behaviour that shows where good linearity performance comes from, and quantifies its sensitivity to circuit parameters. This analytical rigour isn't found often enough in audio amplifier engineering discourse, in my view.

The basic linearity issue stems from the imperfect handover between the pushing and pulling halves of the circuit. At some point, this handover turns into a rout, and the current flowing in one of the halves falls to zero. This cessation of correct operation causes dynamic difficulties when the stage
needs to 'get going' again as the direction of demanded current changes. Significant distortion and HF stability issues ensue, and it's generally considered a Bad Thing if the current in one half of a push-pull output stage falls to zero.

So the second issue we must contend with is the 'switching distortion' problem. This is the bundle of dynamic effects generated by the cessation of current flow in one half of the push-pull stage.

Linearity and switching effects are bracketed together under the broadly used term "crossover distortion," though several mechanisms cause the difficulties. Margan concentrates on the subtle consequences of switching distortion [1]. It's a goal of this present work to eliminate entirely the performance-limiting behaviour that arises from these mechanisms.

Note that just setting a finite quiescent current (that's the standing current in the output devices at zero signal level) doesn't ensure that the current in one half of the push-pull stage won't fall to zero if you pull enough current out of the other side. In all 'classical' Class AB output stages, the current in one half is certain to fall to zero if you pull enough current out of the other side.

The uneasy truce between the pushing and pulling halves also results in uncertainties in circuit bias currents. The temperature dependencies of multiple devices of varying technologies and power levels make for a thermal stability challenge.

Thank goodness that the affluent consumers that purchase high quality audio amplifiers operate them in centrally-heated (or cooled) living environments. High output current linear amplifiers that need to run over the entire industrial (or even military) temperature range, need a more rigorous and disciplined approach to setting and maintaining critical biasing levels.

Comparison with other non-switching and low distortion approaches

Many non-switching linear amplifier topologies have made it into publication, patent or production. Beginning with Blomley in 1969, the 70s and 80s was a productive period of such development; certainly the golden age of the linear audio amplifier. Duncan [4] has a useful survey of these approaches.

The goal of topology modifications to produce non-switching amplifiers is to force current in the idle half of the amplifier to stay finite, rather than fall to zero. This is generally done by smoothly converting the half-output stage to a constant current source (or sink). It allows the inactive half of a push-pull output stage to make a "soft landing" into a non-zero current state, and subsequently a "soft takeoff" into the active state. Margan [1] shows a simple, elegant and empirical way of doing this, at some expense to headroom and power dissipation.

The topology presented here achieves this goal in an analytically sound way. It provides design equations that allow both the quiescent and the limiting currents to be both set precisely and sustained accurately over batch and temperature, without the need for any adjustment on the production line. We'll see soon that this structured approach to device biasing and the avoidance of switch-off can also fully correct for the open-loop linearity issues known to be caused by the varying combined transconductance of the paralleled output stage halves, as they dance through the crossover region.

These problems have also been addressed by other approaches that seek to improve linearity without a corresponding increase in the loop gain 'stick' often deployed to beat these errors into the ground. They come under the broad heading of 'error correction.' Cordell [5] gives a good overview
of his own approach; see also Duncan [4] and Didden [9].

Such circuits generally fall into two categories. In one, a signal corresponding to the error being generated by the output stage is developed, and then structurally subtracted from the signal feeding the loudspeaker. In the other, a subtle rearrangement of feedback connections creates additional internal loop gain (sometimes non-obvious) that can suppress non-linearities, while leaving the external transfer function unchanged.

What's presented in this paper isn't error correction, however. It's a structurally linear approach that creates output stage halves with very predictable transfer functions whose non-linearities cancel in a way that's not sensitive to actual operating current, leaving only the desired signal.

**An evaluation methodology: treat current and voltage separately**

A power output stage can exhibit departures from linear behaviour that are related both to the voltage it's trying to assert onto the load, and to the current that the load demands in response. Minimizing the total output error, to create a really high quality stage, requires attending to both of these error sources independently. It's hard to separate the effects when the output is just loaded with a resistor.

In the real world, significant phase shifts exist between the voltage across the load and the current flowing in it (the current is often significant when the output voltage is zero, and vice versa). This significantly affects the nature of the errors that the output stage contributes. Classical crossover distortion artefacts happen at low load currents, not low load voltages!

The now-standard technique of plotting the incremental gain over a swept output voltage (Self [2], Cordell [5]) can be adapted to give some additional insight. To check the effect of output current, independent of voltage, the input voltage is fixed at some value and a DC sweep is done on a current source injected at the output.

It's instructive to plot the partial derivative of output voltage with load current; that's the dynamic output resistance. In the ideal world, this should be independent both of the load current and of the output voltage chosen. When an audio signal is applied to the loaded output stage, the output impedance variation causes distortion, because the voltage divider between the output impedance and the load impedance has a current-dependent value.

To check the effect on linearity of output voltage swing only, the source roles are reversed, and a fixed load current is taken from the output stage while the input voltage is swept over a range. This reveals how the large signal voltage gain is being affected by the static load current - and indeed, many output stages show some quite alarming deviations when tested this way. To round out the analysis, a conventional voltage sweep into a resistive load is also done.

Another useful adaptation of this method is to change the type of sweep used. The normal approach uses a sequence of DC analyses, plotting 'timeless' operating points as a function of the setup voltages and currents. If instead a ramp over time is used, a SPICE transient analysis can show how the stage actually responds to this 'real' signal. The slew rate of the ramp can be chosen to challenge the dynamic response behaviour of the output stage. Now the behaviour of the linearity in the crossover region can be tested dynamically, for any load current slew rates. Cordell [5] demonstrates that significant dynamic crossover distortion can be created in some output stage forms by such current slew rates.

**Starting point: the long-tailed pair-based buffer**
Starting point: the long-tailed pair-based buffer

Standard power output stages aren't very precise circuits. The inherent offset voltage of base-emitter junctions (or gate-source interfaces) makes it tough to establish stable and optimum operating points. We can do a lot better than this, using local feedback techniques borrowed from precision circuit design discipline.

Consider the basic forms in figures 1 (a) and (b). Familiar from the op-amp world, these are unity gain buffers (UGB) using a very tight local feedback loop. They have a small offset voltage that’s only dependent on the Vbe difference between the two transistors. Form (a) has a long history in the electronics industry; at least as far back as the LM102 buffer introduced in 1967 (Addis [7], and see also Self [3]). Type (b) can also be thought of as a CFB pair whose input transistor has been replaced by a long-tailed pair.

![Figure 1: Two forms of the unity gain buffer (UGB) output stage.](image)

There are two main high-quality ways of supplying the collector currents to the long-tailed pair. One is to have a current source equal to half the tail current feeding the active collector, and to just short the other one to the rail. The other is to use a current mirror that forces the two currents to have some specific relationship (note: they may not be equal - see later on). Figure 1 employs ideal current mirrors made with the SPICE F primitive, where the current source senses the current in an associated voltage source.

It’s tempting to try a push-pull version of this UGB circuit. The very small input offset voltage means that we now only have to create a very small extra voltage between the input transistor bases in order to define the voltage across the sensing resistors \( \{\text{re}\} \) and hence the quiescent current. This voltage doesn’t need to track the properties or environment of the power devices, resulting in much-improved quiescent stability.
Such a configuration (in the complementary form of figure 1b, with simple resistive collector loads) was used in the 200W "PowerSlave" amplifier described in the April 1978 issue of the UK magazine ETI [6]. Figure 2 shows an adapted test circuit for simulation, derived from the published PowerSlave design (which my best friend built at the time). Ideal transistors have been used for the input devices, which are run at a tail current of 30mA as published (they get quite warm in the real amplifier). Emitter current sensing resistors are 0.22 ohms and the load resistance is 4 ohms. Input voltage is swept from -2 V to +2 V, which causes output current to vary over ~-0.5 A to +0.5 A, which is where all the action is.

As the bias voltage between input bases is stepped, the circuit transitions from a "gm-doubling" mode (both halves contributing to output transconductance, with finite quiescent current flowing) to a zero-gm 'dead-zone' mode (reduced or even zero transconductance, with zero quiescent current). Extra emitter degeneration 'softens' the edges of the gm-doubling region but doesn't change the width or magnitude. The narrow almost-linear region is very sensitive to the value of the bias voltage forced across the emitter resistors; see figure 3, where the increments in the bias total bias voltage are 20 mV per step.
Figure 3: varying the bias of the PowerSlave push-pull UGB stage.

Just as in the conventional case, the inactive half of the output stage will eventually turn off when enough current is drawn from the active half. It's easy to see (homework for you, if it's not!) that this happens at a load current of around twice the quiescent current. This circuit definitely doesn't prevent output stage switch-off, and clearly still suffers from gm-doubling.

Now, if the input long-tailed pair is built using identical, ideal transistors passing the same current, the input offset voltage is zero. But if the current is set to be different in the two devices, the long-tailed pair develops an offset voltage (proportional to absolute temperature). This offset voltage can be used to define the voltage across the output stage emitter resistors, so that a separately generated bias voltage between the two halves of the output stage is then not necessary.

An offset voltage can also be developed by inserting a small resistor at the emitter of the feedback transistor (or unbalancing the pair of degeneration resistors). Using such a mismatch-generated offset doesn't do anything to mitigate the sensitivity to the resultant bias voltage, but it does give us a way of accurately setting it.

In Part 2, we'll introduce and analyze the modification to this driver stage that for many years I've referred to as "Class i" - though I'm sure others have also used this moniker.

References
[10] Owen Jones, private communication

About the author
As well as being The Filter Wizard, Kendall's a skilled circuit designer, product engineer and all-round analog expert. He's a widely published communicator of theory and method in many electronics disciplines. His fascination with electronics and audio dates back to boyhood. He spent 21 years designing filters, precision instrumentation, signal processing equipment and music systems. Then 11 years selling and designing in analog semiconductors. He's created advanced products to chase signals across many domains, extract the information from them and do something useful with it.

These days, Kendall is with Cypress, doing system architecture, product definition and strategic application analysis for their precision analog and mixed-signal devices. And also, of course, supporting customers with all sorts of analog and digital filter designs. He's also educating and mentoring a new generation in the ways of analog and systems thinking. Striving to improve the world one dB at a time!