A Compilation of Technical Papers on Audio Amplifier Systems

Malcolm Hawksford, Emeritus Professor

Department of Computer Science and Electronic Engineering

University of Essex, Colchester UK

mjh@essex.ac.uk

INDEX

13 Relationships between Noise Shaping and Nested Differentiating Feedback Loops

14 MC LFD Preamplifier 1988

15 Current-Steering Transimpedance Amplifiers for High-Resolution Digital-to-Analogue Converters

16 Low Distortion programmable Gain Cell Using Current-Steering Cascode Topology

17 Low-Distortion Programmable Gain Cell Using Current-Steering Cascode Topology

18 Topological Enhancements of Translinear Two-Quadrant Gain Cells

19 Tandem Quadruplet VCA Topology

20 Voltage-Controlled Amplifier Systems

21 Linearization of Class-D Output Stages for High-Performance Audio Power Amplifiers

22 Dynamic Model-Based Linearization of Quantized Pulse-Width Modulation for Applications in Digital-to-Analog Conversion and Digital Power Amplifier Systems

23 Linearization of Multilevel, Multi-width Digital PWM with Applications in Digital-to-Analog Conversion

24 An Oversampled Digital PWM Linearization Technique for Digital-to-Analog Conversion

25 Digital audio power amplifier for DSD data streams

26 Switching amplifiers and feedback

27 The Essex Echo: Audio According to Hawksford, Part 1

28 The Essex Echo: Audio According to Hawksford, Part 2

Distortion Correction in Audio Power Amplifiers*

M. J. HAWKSFORD

Audio Research Group, Department of Electrical Engineering Science, University of Essex, Colchester, UK

An audio power amplifier design technique is presented which has the property of minimizing the nonlinear distortion that is generated in class A and class AB output stages.

A modified feedback technique has been identified that is particularly suited to the design of near-unity gain stages. The technique can linearize the transfer characteristic and minimize the output resistance of the output stage. Consequently it is possible to design a power amplifier that uses fairly modest overall negative feedback, yet attains minimal crossover distortion together with an adequate damping factor.

A generalized feedforward–feedback structure is presented from which a system model is derived that can compensate for both nonlinear voltage and nonlinear current transfer characteristics. From this theoretical model, several circuit examples are presented which illustrate that only circuits of modest complexity are needed to implement the distortion correction technique.

In conclusion a design philosophy is described for an audio power amplifier which is appropriate for both bipolar and FET devices, whereby only modest overall negative feedback is necessary.

0 INTRODUCTION

This paper discusses the problems of minimizing crossover distortion in class A and class AB audio power amplifiers. Traditionally output-voltage-derived negative feedback and appropriate biasing of the output transistors have been applied with varying degrees of success in an attempt to achieve acceptable linearity. However, since all transistors exhibit nonlinearity and as, in particular, the output transistors are generally operated into cutoff, successful suppression of the distortion using these techniques is limited.

There are several fundamental problems that can be encountered when using negative feedback to minimize distortion in power amplifiers:

1) Bipolar power transistors are usually of limited bandwidth (typical f_T = 1–5 MHz); thus if nondynamic behavior is required within the audio band, loop gains of only 30 dB are possible.

2) Since crossover distortion is transient in nature and of wide bandwidth, the inevitably falling high-frequency loop gain, together with the resulting loop delay, severely limits the degree of distortion suppression possible.

3) In output-voltage-derived negative feedback amplifiers the distortion which is generated by the output transistors is fed back to the input circuitry. Consequently the pre-output stages process both the desired input signal and the output stage distortion. Thus intermodulation is impaired, especially as the distortion bandwidth can significantly exceed that of the audio signal.

4) If the output resistance of the output stage is nonzero (independent of any overall feedback), the loudspeaker load is an integral component in the feedback loop. Hence if the load exhibits nonlinearity, then distortion components are again fed back to the amplifier's input stage.

A technique is described in this paper which can dramatically linearize the output device characteristics with respect to both voltage transfer and current transfer. Hence an amplifier philosophy evolves that helps to reduce the problems outlined in 1)–4).

* Presented at the 65th Convention of the Audio Engineering Society, London, 1980 February 25–28.

 0004-7554/81/010027-04$00.75

1 THEORETICAL MODEL

The principle of the distortion cancellation technique can be described by considering the generalized error feedback structure shown in Fig. 1. In this network there is error sensing feedforward as well as feedback applied around the nonlinear element N, where in the most general case the input N is unspecified. The error signal used in the system is defined as the difference between the input and the output of N. Thus if N is ideal (that is, $N = 1$), then the error signal is zero and no correction is applied. However, in all practical amplifiers N will deviate from unity, thus the error signal represents the exact distortion due to N.

1.1 Analysis

Let V_n and $N(V_n)$ be the input and output of the N network. Thus examination of the signals in Fig. 1 reveals:

$$V_{out} = N(V_n) + b\{V_n - N(V_n)\}$$

$$V_n = V_{in} + a\{V_n + N(V_n)\}.$$

Eliminating V_n,

$$V_{out} = N(V_n)\left\{(1 - b) - \frac{ab}{(1 - a)}\right\} + \frac{b}{(1 - a)} V_{in}. \quad (1)$$

If

$$(1 - a) = b \quad (2)$$

then

$$V_{out} = V_{in}. \quad (3)$$

Thus providing that stability is maintained and V_n remains finite, distortion cancellation results when Eq. (2) is enforced.

The result [Eqs. (2) and (3)] indicates that there is a continuum of solutions extending from an error feedback system through to an error feedforward system.

It is interesting to note that the input of N is unspecified. It may therefore be derived directly from V_n or indeed any other point within the structure, providing that stability is maintained. For example, by putting $a = 0$, $b = 1$, the classic feedforward system results, where if the input of N is derived from the output of the error difference amplifier, then the Quad [1], [2] feedback structure results (see dashed connection in Fig. 1).

In this paper we consider the opposite extreme where $a = 1$, $b = 0$, and the input of N is equal to V_n. This system is of the type first discussed by Llewellyn in 1941 [3] in relation to valve amplifiers and later by Cherry [4] in 1978. It will now be shown that this feedback technique is particularly relevant to the design of unity-gain follower-type output stages, where with modest circuitry a dramatic improvement in performance is possible. The theory is extended to show that linearization of devices with nonlinear current gain is also feasible.

2 CIRCUIT TOPOLOGIES FOR OUTPUT-STAGE LINEARIZATION

Power amplifiers generally use bipolar output transistors which exhibit low nonlinear current gain. Consequently when such devices are used in a complementary emitter-follower configuration, the transformed loudspeaker load as seen by the base terminals is rendered nonlinear and therefore contributes to the amplifier distortion.

If distortion correction feedback is configured to include input current sensing, it is possible to compensate for changes in current gain. Thus when combined with voltage error sensing feedback, a unity-gain stage results which can be driven from a stage with a finite-output resistance.

In Fig. 2 the schematic of a system with both voltage- and current-sensing circuitry is shown, where the system is configured to illustrate how a practical circuit (Fig. 3) may be realized.

Analysis shows that when

$$k_1 = 1 + \frac{2R_1}{R_2} \quad (4)$$

$$R_1R_3 = R_2R_4 \quad (5)$$

the voltage gain is unity even when the base currents of T_1 and T_2 are finite and V_{BE}/I_E introduces nonlinearity.

As a point of design interest, the resistor R_1 includes

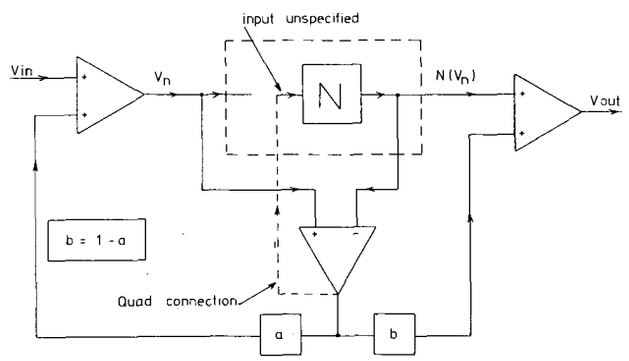


Fig. 1. Generalized feedback-feedforward structure.

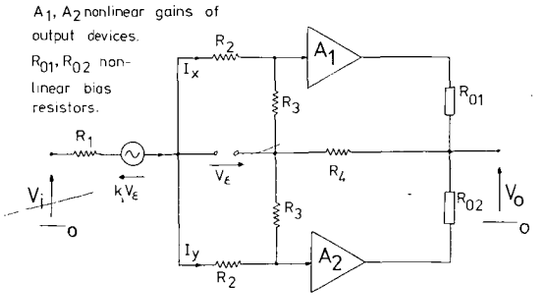


Fig. 2. Current- and voltage-error-sensing feedback.

the output resistance of the driving stage. Consequently the driving amplifier is not required to have zero output resistance.

2.1 Corollary

Since the voltage gain is unity, it follows that the output resistance of the stage is zero, even when the output resistance of the driving stage is finite. As a result, an amplifier that uses this error-correction feedback system does not in principle have to rely upon an overall output-voltage-derived negative feedback loop to achieve adequate loudspeaker damping. Also, the loudspeaker load is then effectively decoupled from the overall feedback loop, and it is this factor that prevents loudspeaker-generated distortion products from reaching the input circuitry of the power amplifier.

Three practical output stage circuits are shown in Figs. 3–5. The circuit of Fig. 3 has both voltage and current sensing and is derived from Fig. 2. However, if the output devices have adequate current gain (such as MOSFET or Darlington transistors), then current sensing is unnecessary. As a result, the much simplified circuits of Figs. 4 and 5 are illustrated to show the modest circuit requirements that are needed to realize only error-voltage sensing. The circuit of Fig. 5 is particularly attractive as the transistors T_3, T_4 form both a complementary error difference amplifier as well as "amplified diodes" for biasing the output transistors.

3 CONCLUSIONS

This paper has described an approach to power amplifier design where the nonlinear distortion generated by the output transistors is compensated by simple fast-acting local circuitry which can result in a high degree of linearity that is appropriate to class A and class AB follower-type output stages.

The technique should find favor among designers who adhere to the low-feedback school of design, as corrective feedback is only applied when distortion in the output stage is generated. If, therefore, the output stage N is designed to be as linear as possible, a fact that can be aided by parallel connection of output transistors, then only minimal error signals result.

Since output stage and loudspeaker generated distortions are in principle isolated from the input stages, these stages are required only to produce modest voltage gains, as large loop gains are not required in an attempt to produce a linear amplifier. Consequently the loop gain is low and the loop bandwidth can be high, enabling a nondynamic loop behavior well in excess of the audio bandwidth.

In practical amplifier design, the sensitivity of adjustment of the balance conditions depends largely on the quiescent bias current of the output transistors, where critical adjustment results only under extremely low biasing. It has been found that for normal bias levels, adjustment is noncritical, also that sensitivity is aided by modest overall feedback.

Several prototype circuits have been investigated where the technique has proved effective. In these amplifiers no stability problems have been encountered other than with the susceptibility to oscillation of power Darlington transistors which appear critical on layout. In fact,

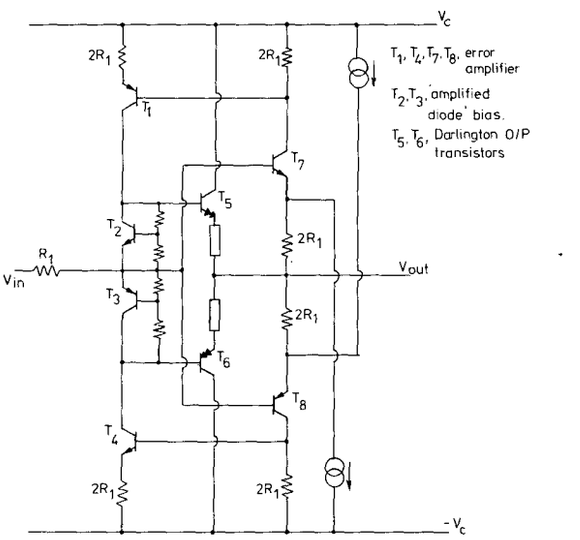


Fig. 4. Example of voltage-error-sensing circuit.

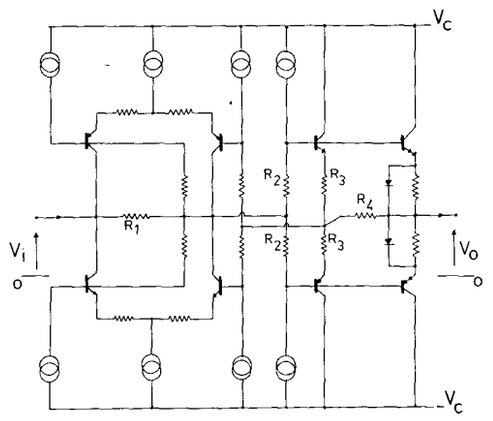


Fig. 3. Circuit schematic of current- and voltage-error-sensing output stage.

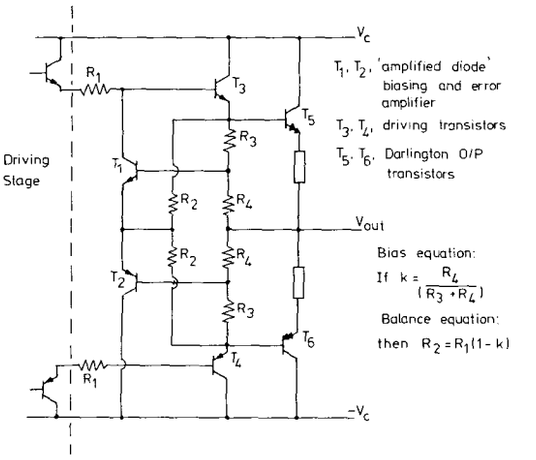


Fig. 5. Voltage error sensing circuit using amplified diodes as error amplifier.

due to the low loop gain, load-dependent instability is minimal, though standard series Zobel circuitry was employed. In practice the bandwidth of the correction circuitry is high which enables fast correction of output-stage nonlinearities. In fact, it is partly the speed of the correction loop that enables a greater suppression of distortion compared with an overall feedback system.

4 REFERENCES

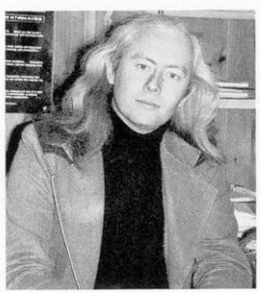
[1] P. J. Walker and M. P. Albinson, "Current Dumping Audio Amplifier," presented at the 50th Convention of the Audio Engineering Society, London, 1975 March 4–7.

[2] P. J. Walker, "Current Dumping Audio Power Amplifier," *Wireless World*, vol. 81, pp. 560–562 (1975 Dec.).

[3] F. B. Llewellyn, "Wave Translation Systems," U.S. Patent 2,245,598, 1941 June 17.

[4] E. M. Cherry, "A New Result in Negative-Feedback Theory and Its Application to Audio Power Amplifiers," *Int. J. Circuit Theory Appl.*, vol. 6, pp. 265–288 (1978 July).

THE AUTHOR



Malcolm J. Hawksford was born in Shrewsbury, England, in 1947. His professional education was at the University of Aston in Birmingham where he studied electrical engineering from 1965–68 and was subsequently awarded a first class B.Sc. degree. In 1968 he obtained a BBC Research Scholarship for three years of postgraduate study at Aston University. His research subject was the application of Delta modulation to color television systems. This work resulted in the award of a Ph.D. degree in 1972.

In 1971 he obtained a lectureship at the University of Essex in the electrical engineering science department where he has taught subjects including electromagnetic theory, audio engineering, digital communications, circuit design and television engineering. At Essex he developed an Audio Research Group where projects on amplifier design, loudspeaker crossover design, analogue-to-digital conversion and music synthesis have been undertaken.

Dr. Hawksford is a member of the Audio Engineering Society, the IEE, the Royal Television Society, and is a chartered engineer.

His hobbies include listening to music, designing audio equipment, home computing and motorcycling.

Distortion Correction Circuits for Audio Amplifiers*

M. J. HAWKSFORD

University of Essex, Department of Electrical Engineering Science, Colchester, CO4 3SQ, United Kingdom

Circuit topologies are introduced which should prove of use to the circuit designer of analog audio amplifiers. The objective is to produce circuits of modest complexity that overcome the nonlinearities inherent in single-transistor and long-tail pair circuits. This allows amplifiers with excellent linearity to be designed without resorting to overall negative feedback with high loop gains. To aid comparison of circuit nonlinear behavior, a parameter called the incremental distortion factor (IDF) is introduced and discussed.

0 INTRODUCTION

Most modern transistor amplifiers use either a single transistor or a pair of transistors in the input circuitry. It is argued that if this stage is cascaded with adequate gain, then by the expedience of overall negative feedback, the input devices will operate within the limits for small-signal operation and thus yield good overall linearity.

Often a consequence of this design philosophy is poor dynamic performance of the input circuitry, where modest input overload can result in gross distortion. There are simple circuit modifications that can be introduced: an increase in device operating current, though possibly at the expense of the noise factor; the introduction of local negative feedback (emitter degeneration) which reduces stage gain but enhances linearity and overload performance, again at the expense of the noise factor.

The aim of this paper is to introduce circuit topologies that enhance the nonlinear performance of amplifier gain cells without recourse to high overall negative feedback. It is considered by this author that the combination of high loop gain together with its inevitable dynamic performance (dominant pole) when compounded with nonlinear elements can result in poor transient distortion characteristics, especially when complex signals are being processed. Since the signals being amplified are rendered more complex due to these nonlinearities falling within a dynamic negative feedback loop, then intermodulation products result which are effectively time smeared. In the limit this must determine the ultimate resolution of an amplifier, which is its ability to transfer fine signal detail in the presence of complex signals.

The only rational methodology to minimize these attributes of nonlinear distortion is to use gain cells that are inherently linear over a wide range of their transfer characteristics and are essentially nondynamic with predictable gain characteristics. Such gain cells can then be used with amplifiers with overall negative feedback without detriment to the intermodulation performance. However, the use of linear circuitry may well render the need for high negative feedback unnecessary.

This paper investigates and catalogs examples of gain cells that generally exhibit good linearity and dynamic range. The circuits should prove of use to designers of both discrete and integrated circuitry, although some design examples which are particularly relevant to integrated-circuit fabrication are included.

In order to facilitate the comparison of various circuit topologies, a parameter called incremental distortion factor (IDF) is introduced. The IDF is related to the change in slope of the transfer characteristic with the input signal and is useful for quantifying nonlinearity under large-signal conditions.

1 PRINCIPLES OF DISTORTION CORRECTION

Three methods are identified in this section to enhance the linearity of gain cells that may already use either local or overall negative feedback within an amplifier structure. (See [1–5] for background.)

1.1 Complementary Nonlinear Stages in Cascade

If a stage has a predictable nonlinearity, then by using a nonlinear stage with a complementary transfer characteristic, overall linearity is possible (Fig. 1). This technique is, for example, used in translinear multiplier stages and in a modified form is the principle of complementary compandors.

* Manuscript received 1981 January 22.

 0004-7554/81/070503-08$00.75

1.2 Device Linearization

This method involves matching device nonlinearities as with the long-tailed pair, where the transconductance is linearized approximately by keeping r_{ep} constant over a wider range of emitter current compared with a single transistor r_e over the same current range. Thus for single transistors,

$$r_{e1} = \frac{\partial V_{be1}}{\partial I_{e1}} \tag{1}$$

$$r_{e2} = \frac{\partial V_{be2}}{\partial I_{e2}} \tag{2}$$

and for a long-tail pair of transistors,

$$r_{ep} = \frac{\partial V_{be1}}{\partial I_{e1}} + \frac{\partial V_{be2}}{\partial I_{e2}} \tag{3}$$

Comparing r_{ep} with r_{e1} or r_{e2} for a given change in emitter current, r_{ep} exhibits greater linearity.

1.3 Error Feedforward and Feedback Distortion Correction

A technique [6] that was recently reported for linearizing near unity gain output stages in analog power amplifiers uses in general a combination of error feedforward and error feedback. Fig. 2 illustrates the method in schematic form.

Analysis shows that when

$$b = (1 - a) \tag{4}$$

then

$$S_{out} = S_{in} \tag{5}$$

where a and b are constrained to values between 0 and 1.

If $a = 1$ and $b = 0$, the system becomes pure error feedback, while if $a = 0$ and $b = 1$, pure feedforward error correction results.

When the balance equation (4) is satisfied, the effects of nonlinearity in the general network N are minimized, and the output parameters S_{out} and S_{in} become linearly related. Though it is inferred that these parameters are voltages, in general they may be any suitable combination of current and voltage, such as voltage in, current out, which is of particular importance for the input stage of an audio amplifier.

Although this principle can be applied to an overall amplifier, it is recommended that the technique be restricted to single stages (which in turn can be compounded to form a complete amplifier), as this permits near nondynamic stage performance and minimizes signal distortion that can be generated in cascaded high gain, low local-feedback amplifier stages.

In practical amplifier design it is possible to compound the techniques outlined in this section to produce amplifier stages of high linearity. It is also possible, within limits, to trade off circuit complexity against performance and to choose a technique that is best suited to a particular amplifier application. In the following sections, circuit examples will be discussed to indicate how predictable amplifiers can be designed and that by the careful choice of design techniques enhanced performance results.

2 INCREMENTAL DISTORTION FACTOR (IDF)

The prime nonlinearity of a transistor which is operated with near constant collector–base voltage is defined by the exponential relationship

$$I_e = I_0 \exp\left(\frac{qV_{be}}{KT}\right) \tag{6}$$

where

I_e = emitter current
I_0 = base-emitter diode saturation current
K = Boltzman's constant
q = charge on electron
T = junction temperature (degrees Kelvin)

Some deviation from this relationship will occur, but is of little consequence here.

Thus when a transistor is used as a transconductance amplifier, nonlinear distortion will result. In order to attempt to quantify the nonlinearity, we introduce the term *incremental distortion factor* (IDF). In essence this term is a measure of the change in incremental gain of a stage to the small-signal gain. In practical circuits the IDF can most simply be expressed as a function of one or more variables. Hence by observing the variation of IDF with these parameters, an accurate measure of nonlinear performance can be made.

To explain the IDF in more detail, we proceed by analyzing first the nonlinear behavior of a simple single-transistor stage with local emitter degeneration and second the performance of a two-transistor long-tail pair. These results are also of use as a reference to allow comparison with the more elaborate gain cell topologies presented in later sections.

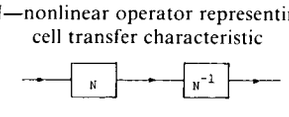


Fig. 1. Complementary linearization.

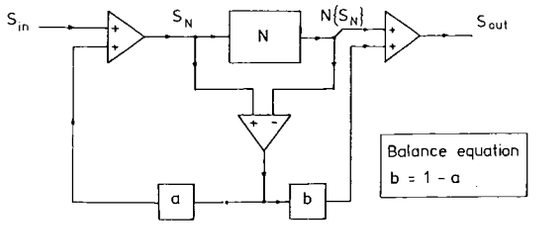


Fig. 2. Error feedforward and feedback distortion correction.

2.1 Distortion Characteristics of a Single-Transistor Cell

A single-transistor cell is shown in Fig. 3. We assume the base current to be negligible. Hence from Eq. (6),

$$V_{be} = \frac{KT}{q} \ln\left(\frac{I_e}{I_0}\right) . \quad (7)$$

Let $\alpha = KT/q$. Therefore

$$V_{be} = \alpha \ln\left[\frac{I_e}{I_0}\right] . \quad (8)$$

Applying Kirchhoff's law to the circuit shown in Fig. 3 and eliminating V_{be} [using Eq. (8)],

$$V_{in} = (i - I_x)R + \alpha \ln\left(\frac{I + i}{I_0}\right) \quad (9)$$

(bias currents I_x, I are shown in Fig. 3). In this simple example V_{in} is a function of a single variable i, that is,

$$V_{in} = f(i).$$

By differentiation we obtain

$$dV_{in} = \left(\frac{dV_{in}}{di}\right) di$$

therefore

$$dV_{in} = R\,di + \frac{\alpha}{I + i}\,di.$$

Extracting linear and nonlinear components,

$$\underbrace{dV_{in}}_{I/P \text{ voltage}} = \underbrace{\left(R + \frac{\alpha}{I}\right)di}_{\text{linear component}} - \underbrace{\left\{\frac{\alpha i}{I(I + i)}\right\}di}_{\text{nonlinear component}}. \quad (10)$$

Eq. (10) relates incremental changes in current and voltage expressed as a function of the bias current I and the present state of signal current i. It is essentially the tangent to the transfer characteristic for transconductance. For linearity, dV_{in} and di must be related by a constant multiplier. However, Eq. (10) reveals that the incremental gain is a function of i, which represents a nonlinear process. We define the IDF $N(\ldots)$ as

$$N(x) = \left[\frac{\text{nonlinear incremental gain component}}{\text{linear incremental gain component}}\right]. \quad (11)$$

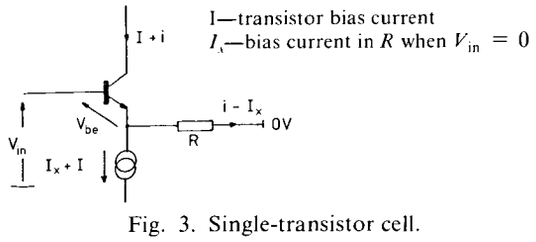


Fig. 3. Single-transistor cell.

$N(\ldots)$ is shown here to be a function of a single variable x. However, in later sections the definition is extended to functions of several variables.

Defining x, the transistor loading factor, as the ratio of signal current i to bias current I for the single-stage transistor amplifier,

$$x = \frac{i}{I} . \quad (12)$$

Hence from Eqs. (10)–(12) we obtain

$$N(x) = \frac{-x}{1 + x}\left(\frac{\alpha}{\alpha + IR}\right) . \quad (13)$$

Eq. (13) reveals that the IDF is an asymmetric function of x, as would be anticipated for a single-transistor nonlinearity. The advantage of this format is that since x is a direct measure of the signal loading of a transistor, then if large values of x result in low values of IDF, this is an expression of near linear performance. In practice x can range from −1 to +1, though usually (except under overload) x will remain well within these limits.

The main advantage of the IDF is that it permits a comparison of circuits with respect to their nonlinear performance, even when complex multiple distorting mechanisms coexist.

2.2 Distortion Characteristic of the Long-Tail Pair Cell

A treatment similar to that presented in Section 2.1 is applied here to the long-tail pair circuit shown in Fig. 4. From Kirchhoff's law,

$$V_{in} = iR + (V_{be1} - V_{be2}) .$$

Applying Eq. (8) to each transistor,

$$V_{in} = iR + \alpha \ln\left[\frac{I + i}{I - i}\right] . \quad (14)$$

Differentiating and extracting linear and nonlinear components,

$$dV_{in} = \left(R + \frac{2\alpha}{I}\right) di + \left\{\frac{2\alpha i^2}{I(I^2 - i^2)}\right\} di. \quad (15)$$

We obtain the IDF using the definition of Eq. (11):

$$N(x) = \frac{x^2}{(1 - x^2)}\left(\frac{2\alpha}{2\alpha + IR}\right) . \quad (16)$$

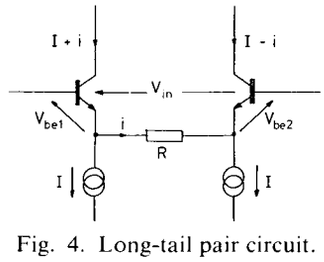


Fig. 4. Long-tail pair circuit.

Comparing Eq. (16) with Eq. (13), the differences in nonlinearity can be compared directly as a function of the transistor loading parameter x. These equations also form a reference for the circuits presented in the following sections.

3 GAIN CELL LINEARIZATION USING FEEDFORWARD ERROR CORRECTION

This section presents a series of circuit topologies that exploit error correction feedforward as outlined in Section 1. Where appropriate, the IDF is evaluated as a means of circuit comparison. All the circuits shown use bipolar transistors, though in most cases adaptation to FET devices should be feasible.

3.1 Single-Stage Feedforward Error Correction

The technique exploited in the circuit of Fig. 5 was derived from Fig. 2, where $a = 0$ and $b = 1$. Essentially when an input signal V_{in} is applied to the base of the input transistor, the resistor R_1 is used as a reference for converting V_{in} to a current. However, due to V_{be1} the voltage across R_1 is less than the input voltage. Hence by using a differential amplifier to measure the error voltage V_{be1}, a corrective current i_2 can be summed with i_1 to compensate almost exactly for the lost current. The transconductance is then almost independent of V_{be1}. Since $V_{be1} < V_{in}$, good linearity results. The main advantage of this circuit is that linearity can be achieved with only modest values of R_1, a fact that increases the transconductance of the cell, yet minimizes Johnson noise due to R_1.

The simplest method of adding the main current i_1 with the error correction current i_2 is to parallel the two collectors. However, if both collector currents of each half of the difference amplifier are used by introducing a current mirror, then either the value of R_2 can be increased, which improves linearity, or the value of R_1 can be reduced, which reduces Johnson noise and increases transconductance.

An example of a more practical amplifier is illustrated in Fig. 6, where biasing requirements and current mirror are shown.

We assume that the output signal current i_0 is derived as

$$i_0 = i_1 + \lambda i_2 \tag{17}$$

where generally λ has a value of 1 or 2 (Fig. 6 assumes $\lambda = 2$). The circuit equations are as follows:

$$V_{in} = (i_1 - I_x)R_1 + V_{be1} \tag{18}$$

$$V_{be1} = (i_2 + I_y)R_2 + (V_{be2} - V_{be3}) \tag{19}$$

$$V_{be1} = \alpha \ln\left[\frac{I_1 + i_1}{I_0}\right] . \tag{20}$$

I_x, I_y—bias currents in R_1 and R_2 when $V_{in} = 0$

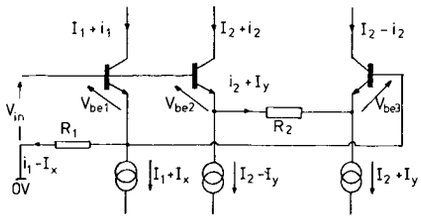


Fig. 5. Single-stage input device with feedforward error correction.

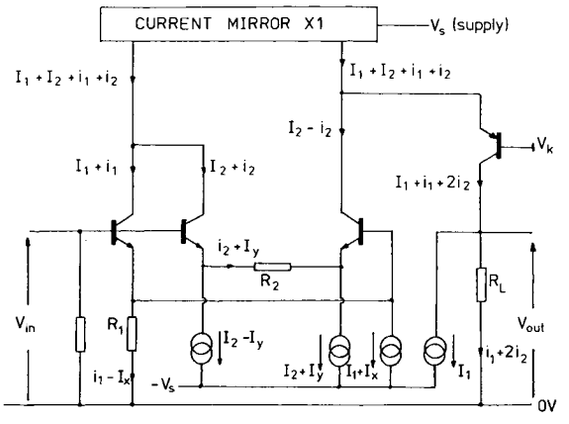


Fig. 6. Practical amplifier stage using a single input transistor with feedforward error correction.

$$V_{be2} - V_{be3} = \alpha \ln\left[\frac{I_2 + i_2}{I_2 - i_2}\right] . \tag{21}$$

Thus

$$V_{in} = (i_1 - I_x)R_1 + (i_2 + I_y)R_2 + \alpha \ln\left[\frac{I_2 + i_2}{I_2 - i_2}\right] .$$

Since

$$V_{in} = f(i_1, i_2)$$

then

$$dV_{in} = \frac{\partial V_{in}}{\partial i_1} di_1 + \frac{\partial V_{in}}{\partial i_2} di_2 .$$

Therefore

$$dV_{in} = R_1 \left[di_1 + \left(\frac{I_2R_2 + 2\alpha}{R_1I_2}\right) di_2 \right] + \frac{2\alpha}{I_2}\left[\frac{i_2^2}{I_2^2 - i_2^2}\right] di_2 .$$

By comparison with Eq. (17),

$$\lambda = \frac{I_2R_2 + 2\alpha}{I_2R_1} . \tag{22}$$

Expressing di_2 as a function of di_0, we then obtain the IDF

$$N(x, y) = \frac{2\alpha^2y^2}{\lambda I_1I_2R_1^2[(1 - x)(1 - y^2R_2/\lambda R_1) + (2\alpha/R_1I_1)(1 - y^2)]} \tag{23}$$

where

$$x = \frac{i_1}{I_1} \quad \text{and} \quad y = \frac{i_2}{I_2}.$$

Since $|y| < |x|$ and $|x| < 1$, then Eq. (23) indicates that a substantial reduction in nonlinearity is possible.

3.2 Symmetrical Long-Tail Pair with Feedforward Error Correction

The primary distortion mechanism of a single transistor is the I_e/V_{be} relationship. If a long-tail pair is chosen, then the primary distortion is reduced, as discussed in Section 2.2, where it was also shown that the nonlinearity is symmetrical about the operating point.

This section investigates the use of feedforward error correction applied to a single long-tail pair. The IDF is stated in Eq. (24), the analysis being similar to that of Section 3.1:

$$N(x, y) = \frac{4\alpha^2 y^2}{\lambda I_1 I_2 R_1^2[(1 - x^2)(1 - y^2 R_2/\lambda R_1) + (2\alpha/I_1 R_1)(1 - y^2)]} \tag{24}$$

where λ, x, y, and i_0 are as defined in Section 3.1. The circuit is given in Fig. 7.

Eq. (24) reveals that the IDF is of a lower order due to the square-law dependence on x and that the nonlinearity is symmetrical about the quiescent operating point.

This particular configuration is applicable to amplifier input stages where offset cancellation of base-emitter junctions is useful in establishing dc biasing of the complete amplifier. Note, however, that there is no requirement for accurate device matching within either the long-tail pair or the error amplifier to achieve useful linearization. (Matching is necessary for accurate dc conditions, but this is a separate problem and may not be of importance in ac-coupled stages.)

3.3 Cascaded Gain Cells to Derive Differential Output Currents

A circuit application may require a differential output current from the gain cell. Since this feature is absent from the circuits presented in Figs. 5, 6 and 7, we consider here modifications that result in differential output currents.

The principle is illustrated in the basic schematic shown in Fig. 8, where identical gain cells are compounded within a cascade topology. This technique results in a fully complementary cell with enhanced IDF due to a reduction in primary distortion by sharing the input signal between cells.

Two circuits are presented in Figs. 9 and 10, which are formed by cascading the respective circuits of Figs. 5 and 7.

3.4 Nested Feedforward Error Correction Amplifier

To conclude this section on feedforward error correction, it should be noted that the error amplifier can be nested to yield even further distortion reduction, where effectively an error amplifier is used to compensate for the main error amplifier. However, in such circuits it is likely that other sources of distortion (other than the V_{be}/I_e nonlinearity) will then be dominant. Also, such circuits become somewhat complex, and the overall improvements are likely to be small. In fact for a given total current consumption in a gain cell, increasing the error amplifier current I_2 will produce a useful reduction in distortion, since the error amplifier loading factor is reduced as a function of y^2. Further enhancement can be obtained by using the modified amplifier cells to be presented in Section 5, in particular the cell shown in Figs. 15 and 16.

4 GAIN CELL LINEARIZATION USING FEEDBACK ERROR CORRECTION

Circuit topologies similar to that of Section 3 can be designed which rely upon the error signal being fed back to the gain cell input. This corresponds to the system

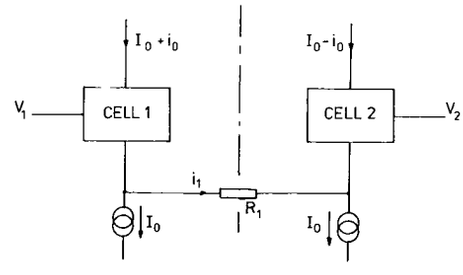


Fig. 8. Basic cascade of two identical gain cells.

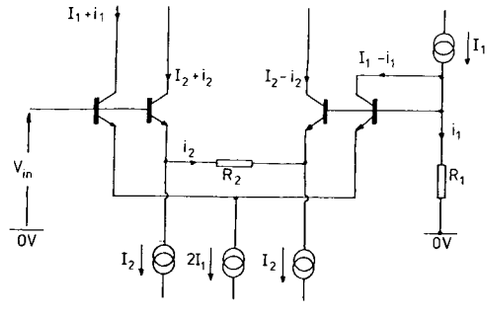


Fig. 7. Single long-tail pair with feedforward error correction.

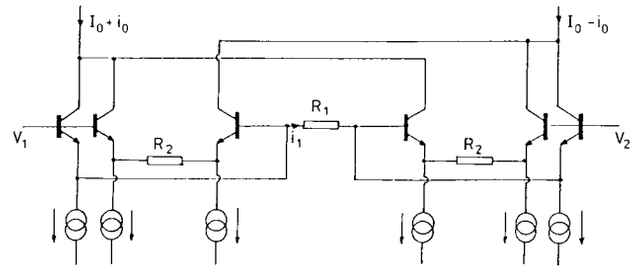


Fig. 9. Single long-tail pair with dual feedforward correction amplifiers (cascade formed from cell shown in Fig. 5).

diagram of Fig. 2, where $a = 1$ and $b = 0$. In these circuits the error signal is generally a current.

Figs. 11 and 12 show two examples which can be compared directly with the feedforward versions illustrated in Figs. 5 and 7.

Compound circuits similar to those described in Section 3.3 also can be derived by cascading gain cells with error correction feedback. These should be proven useful where differential output currents are required in fully symmetrical circuits (see Fig. 8).

The circuit equations are as follows. For the input transistor (Fig. 11),

$$V_{in} = V_{be} + (i_1 - i_2 + I_y)R_1$$

$$V_{be} = \alpha \ln\left[\frac{I_1 + i_1}{I_0}\right]$$

and for the error amplifier,

$$V_{be} = (I_x + i_2)R_2 + \alpha \ln\left[\frac{I_2 + i_2}{I_2 - i_2}\right] .$$

Let

$$R_2 = \frac{R_1 I_2 - 2\alpha}{I_2} . \tag{25}$$

Differentiating V_{in},

$$dV_{in} = R_1\, di_1 - \left[\frac{2\alpha^2 i_2^2 di_1}{I_1 I_2^3 R_1(1 + i_1/I_1)(1 - i_2^2 R_2/I_2^2 R_1)}\right]. \tag{26}$$

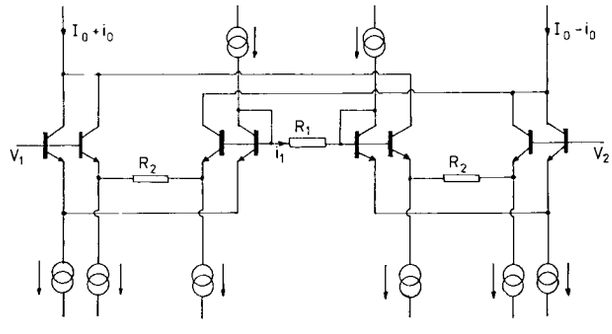


Fig. 10. Dual long-tail pair circuits with dual feedforward correction amplifiers (cascade formed from cell shown in Fig. 7).

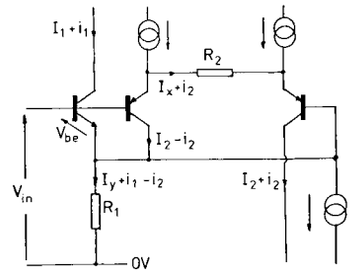


Fig. 11. Single input transistor with error correction feedback.

Therefore

$$N(x, y) = \frac{-2\alpha^2 y^2}{I_1 I_2 R_1^2(1 + x)(1 - y^2 R_2/R_1)} . \tag{27}$$

Again the loading parameters x and y are defined as in Section 3.

It is interesting to note that Eq. (25) represents a balance equation which minimizes the output current i_1 dependence on i_2 and allows R_1^{-1} to determine the transconductance exactly.

A similar analysis for the circuit in Fig. 12 gives the IDF as

$$N(x, y) = \frac{-4\alpha^2 y^2}{I_1 I_2 R_1^2(1 - x^2)(1 - y^2 R_2/R_1)} , \tag{28}$$

where the balance is again determined by Eq. (25).

These results show that the feedback circuits give virtually the same performance as their feedforward counterparts, and for practical circuits the performance should be essentially identical.

5 INDIRECT DISTORTION CANCELLATION TOPOLOGIES

A significant improvement over the standard long-tail pair can be realized by using matched transistors. In these circuits it is assumed that the I_e/V_{be} characteristics are essentially identical. As examples Figs. 13 and 14 show indirect error correction.

In Figs. 13 and 14 transistors T_1, T_3, and T_2, T_4 are matched, and since they carry the same emitter current (excluding the small base current), the base–emitter voltages are identical. Thus an error-sensing difference am-

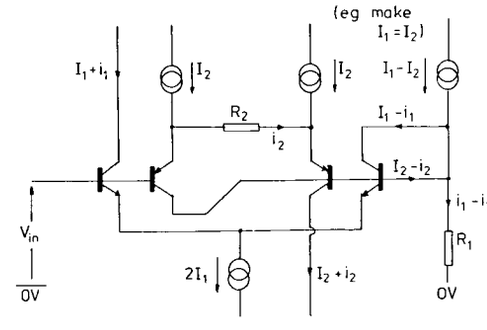


Fig. 12. Long-tail pair with error correction feedback.

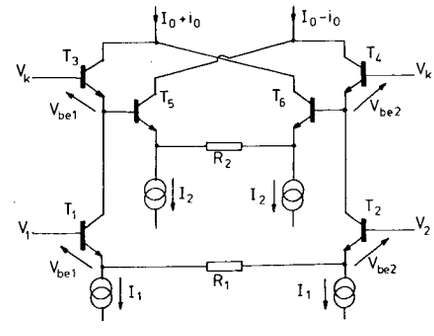


Fig. 13. Indirect error feedforward.

plifier can measure the error voltage ($V_{be1} - V_{be2}$) indirectly and compensate either by feedforward or by feedback.

The advantages of these circuits are that only a single error amplifier need be used, and transistors T_3, T_4 form a cascode configuration, which enhances bandwidth and linearity.

The IDFs for the error feedforward and error feedback circuits should compare with the circuits of Fig. 7 [Eq. (24)] and Fig. 12 [Eq. (28)], respectively, provided that there is accurate transistor matching, and base currents are neglected (i.e., high β transistors).

Finally a circuit is presented in Fig. 15 which combines the advantages of error feedforward with indirect error sensing to minimize nonlinearities.

We assume that all transistors are matched in terms of I_c/V_{bc} nonlinearity and collector–base current gain β. The values of currents and voltages are shown in Fig. 15.

$$V_{in} = (V_1 - V_2) = (V_{be1} - V_{be2}) + (V_{be3} - V_{be4}) + i_1 R$$

where

$$(V_{be1} - V_{be2}) = \alpha \ln\left[\frac{I_1 - ki_1}{I_1 + ki_1}\right]$$

and

$$(V_{be3} - V_{be4}) = \alpha \ln\left[\frac{I_1 + i_1}{I_1 - i_1}\right] .$$

Differentiating V_{in} and substituting for base-emitter voltages,

$$dV_{in} = R\,di_1 + \frac{2\alpha(1 - k^2)x^2\,di_1}{I_1(1 - x^2)(1 - k^2x^2)} \qquad (29)$$

where

$$x = \frac{i_1}{I_1} \qquad (30)$$

and

$$k = \frac{\beta - 1}{\beta + 1} . \qquad (31)$$

Therefore

$$N(x) = \frac{2\alpha(1 - k^2)x^2}{I_1 R\{(1 - x^2)(1 - k^2x^2)\}} . \qquad (32)$$

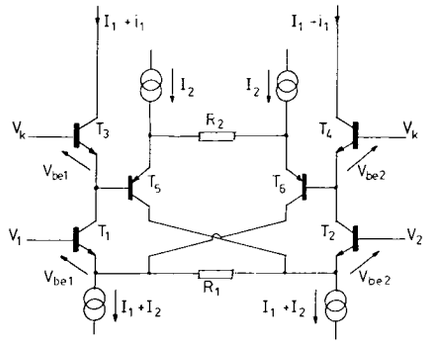


Fig. 14. Indirect error feedback.

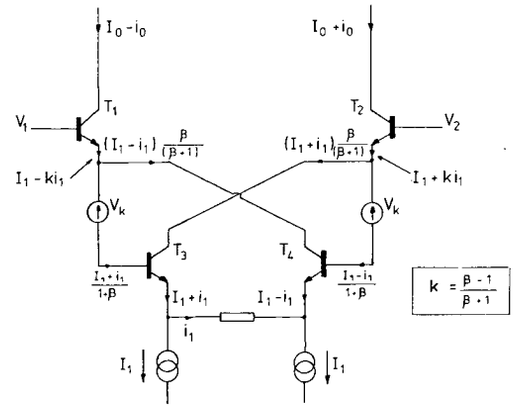


Fig. 15. Modified error feedforward with indirect V_{bc} compensation.

This circuit topology reveals that if transistor matching is achieved, the nonlinearities are mainly dependent upon transistor β.

In order to obtain an adequate dynamic range without transistor saturation, constant offset voltages V_k are required (as shown in Fig. 15). However, in situations where the input signal is controlled and small, V_k can be set to zero. Such an application is to use this circuit as the error amplifier for a single long-tail pair, as shown in Fig. 7. This compound circuit is illustrated in Fig. 16.

Defining λ, x, and y as in Section 3.2, then if

$$\lambda = \frac{R_2}{R_1} \qquad (33)$$

we have

$$dV_{in} = R_1\,di_0 + \frac{4\alpha^2(1 - k^2)y^2\,di_0}{[I_1(1 - x^2)\{I_2R_2(1 - y^2)(1 - k^2y^2) + 2\alpha(1 - k^2)y^2\} + 2\alpha\lambda I_2(1 - y^2)(1 - k^2y^2)]} . \qquad (34)$$

Therefore

$$N(x, y) = \frac{4\alpha^2(1 - k^2)y^2}{\lambda I_1 I_2 R_1^2[(1 - y^2)(1 - k^2y^2)(1 + 2\alpha/I_1R_1 - x^2) + (2\alpha/\lambda I_2R_1)(1 - k^2)y^2(1 - x^2)]} . \qquad (35)$$

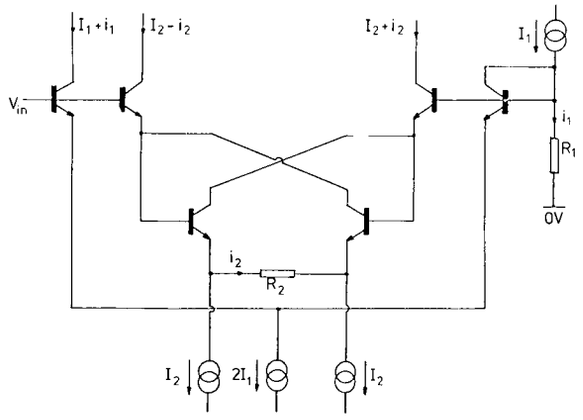


Fig. 16. Single long-tail pair using an error feedforward correction with indirect V_{be} compensation.

Hence comparing Eqs. (35) and (24), there is a further reduction of distortion of approximately $(1 - k^2)$, where k is just less than 1.

6 CONCLUSIONS

A series of circuit topologies were presented which are suitable for the input and subsequent stages of audio amplifiers. The aim has been to show that partial linearization can be achieved without recourse to excessive negative feedback. The circuits require the implementation of a balance condition which is essentially noncritical, provided that only moderate cell transconductance is required.

As an aid to circuit comparison the parameter IDF was introduced which readily expressed nonlinearities as a function of amplifier current loading factors. These expressions can be approximated still further by letting terms of the form $(1 - x^2) \rightarrow 1$ with the assumption of modest loading factors.

It is hoped that some of the circuits will prove useful to the designers of audio amplifiers and allow enhanced performance by minimizing both nonlinearity and loop-gain requirements, which have a strong correlation with transient distortion phenomena.

7 REFERENCES

[1] A. M. Sandman, "Reducing Distortion by Error Add-on," *Wireless World*, vol. 79, p. 32 (1973 Jan.).

[2] D. Bollen, "Distortion Reducer," *Wireless World*, vol. 79, pp. 54–57 (1973 Feb.).

[3] A. M. Sandman, "Reducing Amplifier Distortion," *Wireless World*, vol. 80, pp. 367–371 (1974 Oct.).

[4] P. J. Walker and M. P. Albinson, "Current Dumping Audio Amplifier," presented at the 50th Convention of the Audio Engineering Society, London, England, 1975 March 4–7.

[5] J. Vanderkooy and S. P. Lipshitz, "Feedforward Error Correction in Power Amplifiers," *J. Audio Eng. Soc.*, vol. 28, pp. 2–16 (1980 Jan./Feb.).

[6] M. J. Hawksford, "Distortion Correction in Audio Power Amplifiers," *J. Audio Eng. Soc.*, vol. 29, pp. 27–30 (1981 Jan./Feb.).

Dr. Hawksford's biography was published in the January/February issue.

Fuzzy Distortion in Analog Amplifiers: A Limit to Information Transmission?*

M. J. HAWKSFORD

University of Essex, Department of Electrical Engineering Science, Wivenhoe Park, Colchester, Essex, United Kingdom,

A theoretical model is introduced that attempts to emulate a low-level distortion mechanism inherent in bipolar junction transistor amplifiers and, as a consequence, suggests a low-level bound to the transmission of fine signal detail. The model gives positive support to the low-feedback school of design and proposes circuit techniques for maximizing signal transparency. The design principles have particular relevance to low-level signal stages, but should also find an association with all classes of amplifiers.

0 INTRODUCTION

The last decade has seen substantial debate concerning the relationship between objective and subjective assessment of amplifiers. Measurements have frequently been performed with often impressive results [1], yet on extended audition significant audible differences can still be perceptible.

Various investigations have cited, for example, the levels of harmonic distortion as a measure of excellence, where emphasis has been directed to the distribution and relative weights of the harmonic structure. Conclusions have been drawn suggesting that low-order harmonics exhibiting a smooth rolloff in amplitude with frequency [2], [3] are a useful indicator of an amplifier's performance. However, when on this basis the levels of distortion are critically compared, it is generally difficult to assert a high correlation between objective and subjective results. In fact auditioning of amplifier performance suggests that the absolute level of harmonic distortion is, within limits, only a second-order interest, as highlighted during valve/transistor comparison.

A second indicator of potential excellence depends on the assessment of transient intermodulation distortion (TID) [4], [5], a distortion that is prevalent in slow high-loop-gain feedback amplifiers. However, design criteria have been established [6], [7] which minimize the onset of TID. Clearly, TID is only part of the distortion repertoire and is probably of minimal consequence once the probability of its occurrence is low.

Primary and secondary crossover distortion, though predominant in power amplifier circuits, also occur in certain low-level operational amplifiers that use class AB output stages. However, although this nonlinear mechanism can lead to significant signal impairment, there are now a variety of design techniques [8]–[10] that successfully minimize the error signal.

A direct consequence of amplifier nonlinearity and signal interaction is partial rectification, which produces a dynamic shift in the quiescent bias state. If an amplifier incorporates energy storage elements (such as ac coupling and by-pass capacitors), then the error signal is filtered and exhibits "overhang," which is dominant in the lower midrange and bass frequency bands. Amplifiers should therefore minimize energy storage components and be designed to be near aperiodic within the audio band. Research has shown that an asymmetric pulse test is a sensitive method of assessment [11], [12].

Where amplifiers are operated at high signal levels, other mechanisms of dynamic distortion become significant. Nonlinear delay modulation (NLDM) of the

* This paper was the basis of a lecture to the British Section in 1982 October (see *JAES*, vol. 31, no. 3, pp. 164 and 166 (1983 March)). Manuscript received 1982 October 11; revised 1982 November 22.

signal will occur due to the dynamic variation of transistor parameters with signal: Modulation of collector–base capacitance with collector–base voltage, the shift of small-signal bandwidth with collector current, and general parametric changes when devices are thermally exercised are all contributory factors. However, after reviewing the many conventional forms of nonlinearity it is apparent that certain areas of subjective assessment still elude a satisfactory explanation, and it is unclear as to an optimum design strategy. Specifically the area of greatest concern is that of subjective clarity or what may be usefully described as signal transparency: the ability to resolve fine signal detail, especially in the presence of complex high-level signal components. There appears to be a distinction between distortion mechanisms that "color" the signal, thus adding their own character, and distortions that corrupt fine signal detail.

This paper addresses what is believed to be both a significant and a neglected factor of amplifier performance where two basic clues have emerged: first, that amplifiers using low or distributed feedback often audition with higher rank, even though they may exhibit higher levels of error signal, and second, that low-level amplifier stages appear particularly susceptible to signal impairment. A primitive theory is proposed and a design strategy presented as a means of performance optimization.

In preparing the work presented in this paper, a literature survey revealed an embryonic idea first published by West [13] in 1978. However, the idea was not developed to any extent, and its significance with respect to amplifier design was not established in depth. A later discussion by Curtis [14] dismissed the theory as a cause of "transistor sound." The author considers this dismissal somewhat premature and attempts in this paper to extend the theory in more detail, with respect both to the charge-control model of a transistor and to the application of the derived theory to amplifier design.

1 FUZZY NONLINEARITY: THE THOUGHT EXPERIMENT

Classical circuit theory represents current as a continuous function that flows smoothly and can be considered to have infinite precision within an uncorrelated random bound. This viewpoint is taken from a macroscopic stance of electromagnetism where the individual electrical fields of electrons merge to a nongranular continuum that allows near infinite precision in the transmission of information. Account is of course taken of the behavior of partial randomness of electrons, and this is introduced through linear noise analysis where the noise is seen as the limiting factor on low-level signal resolution. In fact basic calculations on the numerousness of electrons would suggest this to be perfectly reasonable and of little consequence to the audio circuit designer. We speculate here that this may well be an invalid assumption which disguises the true limit to the ultimate resolution of a low-noise amplifier stage.

Transistor operation depends in part on the transfer of charge from signal source to device, a theory first proposed by Beaufoy and Sparkes [15]. Essentially the theory shows that the level of collector current in a bipolar junction transistor (BJT) is a linear function of the local stored charge in the base region. The theory also proposes that the continual base current of a BJT provides a "top up" charge to compensate for recombination resulting from a finite carrier lifetime within the base. In equilibrium the rate of recombination is just balanced by the base current to maintain a constant average charge, which in turn determines the collector current.

However, in this paper we shall not be concerned directly with the mechanics of device operation, only a consequence of those mechanisms, namely, the level of charge transfer required in the amplification process. The probable importance of charge levels can be established by the following thought experiment.

In this discussion we shall evaluate the approximate levels of charge that are transferred to the base of a transistor under low-level signal excitation. Fig. 1 shows a basic zero feedback amplifier stage interfaced to a moving-coil transducer with source resistance r_c, where the input impedance of the amplifier is derived directly from the hybrid-π equivalent circuit of a transistor. In Fig. 1 $r_{bb'}$ is the base bulk resistance, $r_{b'e}$ the dc input resistance (modeling small-signal recombination), and $C_{b'e}$ the base region capacitance storing the charge q_b which controls the collector current.

A value of the base storage capacitor can be estimated directly from a knowledge of f_β, the 3-dB bandwidth of h_{fe}, which is the collector–base current gain, assuming a first-order response,

$$C_{b'e} = \frac{1}{2\pi r_{b'e} f_\beta} \Big| V_{CB} \rightarrow \text{constant} , \tag{1}$$

this expression is derived from the observation that the reactance of $C_{b'e}$ is equal to $r_{b'e}$ at the frequency f_β, it also follows that $C_{b'e} \propto I_e$ (emitter current).

Let us further our argument by considering the output voltage v_i of a moving-coil cartridge,

$$v_i = \frac{f}{f_n} V_n \sin (2\pi f t) \tag{2}$$

where V_n is the nominal cartridge output amplitude at a normalized frequency f_n, typically 1 kHz. If the dynamic range of the system is DR, then the minimum

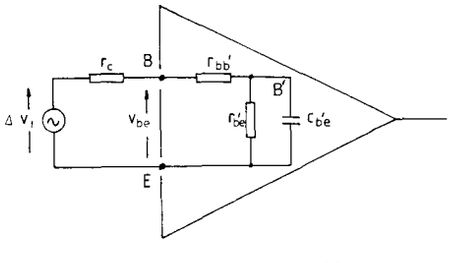


Fig. 1. Basic transistor amplifier stage.

resolvable signal level Δv_i is

$$\Delta v_i = \frac{f}{f_n}\frac{V_n}{DR}\sin(2\pi ft) . \quad (3)$$

In defining DR we refer to the smallest resolvable change in signal level that can exist without nonlinear corruption from complex high-level signal components. Ideally this change in signal level should be below the noise floor.

We next assign a minimum resolvable time period τ_m estimated by direct reference to the sampling theorem, which conveniently relates τ_m to the audible bandwidth f_a,

$$\tau_m \cong \frac{1}{2f_a} . \quad (4)$$

Assuming a sinusoidal input signal, an expression for the control base charge $q_b(t)$ for an input signal Δv_i is derived as

$$q_b(t) = \frac{1}{2\pi r_{b'e} f_\beta}\frac{f}{f_n}\frac{V_n}{DR}\sin(2\pi ft) \quad (5)$$

where

$$r_c + r_{bb'} << \left| r_{b'e} \,//\, \frac{1}{2\pi f C_{b'e}} \right| .$$

Hence the change in control charge that occurs over a time τ_m is

$$\Delta q_{b,\min} = q_b\left(t + \frac{\tau_m}{2}\right) - q_b\left(t - \frac{\tau_m}{2}\right) \quad (6)$$

where, aligning the difference equation to maximize Δq_b (in this sense our estimation is optimistically high) and assuming $\sin(\pi f\tau_m) \cong \pi f\tau_m$, we have

$$\Delta q_{b,\min} = \frac{V_n}{2r_{b'e}DR}\frac{f^2}{f_n f_a f_\beta} . \quad (7)$$

We note from standard transistor theory that

$$r_{b'e} = (1 + h_{fe})r_e \quad (8)$$

$$r_e = \frac{0.025}{I_e} , \ (I_e \text{ in amperes}) \quad (9)$$

$$f_T \simeq (1 + h_{fe})f_\beta . \quad (10)$$

Hence

$$\Delta q_{b,\min} = \frac{20V_nI_e}{DR}\frac{f^2}{f_n f_a f_T} .$$

To estimate typical values of changes in the base charge consider the following data base:

V_n	$= 200\ \mu V$	medium output moving-coil cartridge,
I_e	$= 10^{-3}$ A	transistor emitter bias current,
f_T	$= 50$ MHz	bandwidth to unity h_{fe},
DR	$= 10^4$	80-dB dynamic range,
f_n	$= 1$ kHz	normalizing frequency for cartridge,
f_a	$= 20$ kHz	audible bandwidth,
e	$= 1.96 \times 10^{-19}$ C	charge on electron,
h_{fe}	$= 500$	small-signal collector–base current gain (V_{CE} constant)

whereby the minimum change in base charge is evaluated as

$$\Delta q_{b,\min} \simeq (2 \times 10^{-6} f^2)e \quad \text{[coulombs]} \quad (11)$$

Eq. (11) shows a remarkably low level of average charge transfer that occurs for small signals observed over the minimum resolvable time period (here assumed to be 25 μs).

It is also instructive to estimate the change in the number of electrons transferred into the base region through recombination over the minimum time period τ_m, due only to the minimum signal component Δv_i.

If we assume $r_{b'e}$ to be the dominant input resistance of the transistor, the base input current Δi_i associated with Δv_i is

$$\Delta i_i = \frac{f}{f_n}\frac{V_n}{r_{b'e}DR}\sin(2\pi ft) . \quad (12)$$

The charge Δq_r transferred from source to input due to recombination in time τ_m is calculated by integration,

$$\Delta q_r = \int_{t-\tau_m/2}^{t+\tau_m/2} \Delta i_i \, dt . \quad (13)$$

Aligning the integration window to maximize Δq_r and again assuming that $\pi f\tau_m$ is small,

$$\Delta q_r = \frac{20V_nI_e}{(1 + h_{fe})DR}\frac{f}{f_n f_a} . \quad (14)$$

Using the same data base,

$$\Delta q_r = 0.2fe \quad \text{[coulombs]}. \quad (15)$$

Eqs. (11) and (15) show that low-level signals in transistor stages are associated with an extremely small transfer of charge into the base of the input transistor. The basic analysis indicates that within τ_m the signal amplitude generally has greater effect on the charge transferred for recombination than that charge having direct control of the collector current (according to charge control theory). Nevertheless both calculations yield results of only a few electrons.

We therefore propose a theory that partial signal quantization is the fundamental process that sets an inherent bound to signal transparency through a transistor stage. Both Eqs. (11) and (15) support the probable existence of significant granularity where Eq. (11) suggests a form of amplitude quantization and Eq. (15) an association with $1/f$ noise.

It is also proposed that signal interaction with inherent

nonlinearities in transistors, together with even small levels of interference from power supplies, neighboring circuitry, or undesired signal coupling (such as poor ground line design), can easily corrupt such minute signals and that such corruption should be interpreted as modifications to these low charge levels.

We conclude this preliminary discussion by giving in Table 1 typical levels of charge transferred to the base of a transistor within the minimum time period $\tau_m = 25$ μs against various signal levels to illustrate the potential dynamic range available. The example already cited in this section is used as a data base.

2 FUZZY MODELS

In this section we build upon the observations made of quantization and the relative magnitudes of low-level signals by introducing a basic model of the distortion process. It is emphasized that although the model is primitive, it is a natural extension of our thought experiment.

The proposed model is to be classed as "fuzzy" and the resulting distortion as fuzzy distortion due to its strong stochastic association. We commence by establishing two distinct groups of nonlinearity.

1) *Deterministic nonlinearity.* Classic system nonlinearity can be envisaged using a continuous model incorporating static or dynamic transfer characteristics. The main attribute of this broad distortion classification is repeatability where, assuming no time-dependent system parameters, the same error waveform will result under repeated tests. We note in particular that when measuring such distortion a degree of signal averaging is often used to suppress random events.

2) *Fuzzy nonlinearity.* A distortion process that results in an error signal with a strong stochastic element that does not include any uniform sampling function is defined here as fuzzy distortion. Such distortion will not exhibit exact error waveform replication under repeated tests. We note in particular that when measuring such distortion, any signal averaging will tend to mask the error waveform.

We proceed by further reference to the charge control model of a BJT [15] and attempt to produce a primitive model that matches the input impedance characteristic of a BJT transistor (see Fig. 1), exhibits the correct frequency response when observing h_{fe}, introduces a degree of charge quantization, and maintains the proper static relationship between base and collector currents.

The proposed model is illustrated in Fig. 2(b) and is configured so that it replaces directly the standard hybrid-π circuit shown in Fig. 2(a). A simplified notation is illustrated in Fig. 2(c).

Table 1. Typical levels of charge transferred.

	Charge Transfer in 25 μs	
	Δq_t	$\Delta q_{b,\,min}$
Bias current of 1/500 mA	$2.56 \times 10^8 e$	—
Input signal of 200 μV	$2 \times 10^6 e$	$2 \times 10^4 e$
Input signal 80 dB below 200 μV at 1 kHz	$200e$	$2e$

The model consists of an integrator to convert input signal current to charge, cascaded with a uniform quantizer with an associated dither source $n(t)$ to scatter the quanta. The integrator and quantizer are enclosed within a negative-feedback loop, which together emulate the process of recombination and quantization of the stored base charge. The quantized base–emitter voltage $V_{b'e}$, which is proportional to the stored base charge, is converted to collector current by a transconductance stage with mutual conductance g_m. From standard transistor theory,

$$g_m = \frac{h_{fe}}{(1 + h_{fe})r_e} \tag{16}$$

$$r_e = \frac{\partial V_{BE}}{\partial I_E} = \frac{kT}{eI_e} \tag{17}$$

where k is Boltzmann's constant, T the junction temperature (kelvins), e the charge on an electron, and I_e the emitter bias current.

The model shown in Fig. 2(b) has a strong resemblance to certain classes of analog-to-digital encoder, in particular feedback (pulse-code modulation) and multilevel delta sigma modulation (DSM) [16], [17]. Since these encoding schemes combine integration and quantization within a feedback loop, they form useful vehicles for comparison. A major distinction between

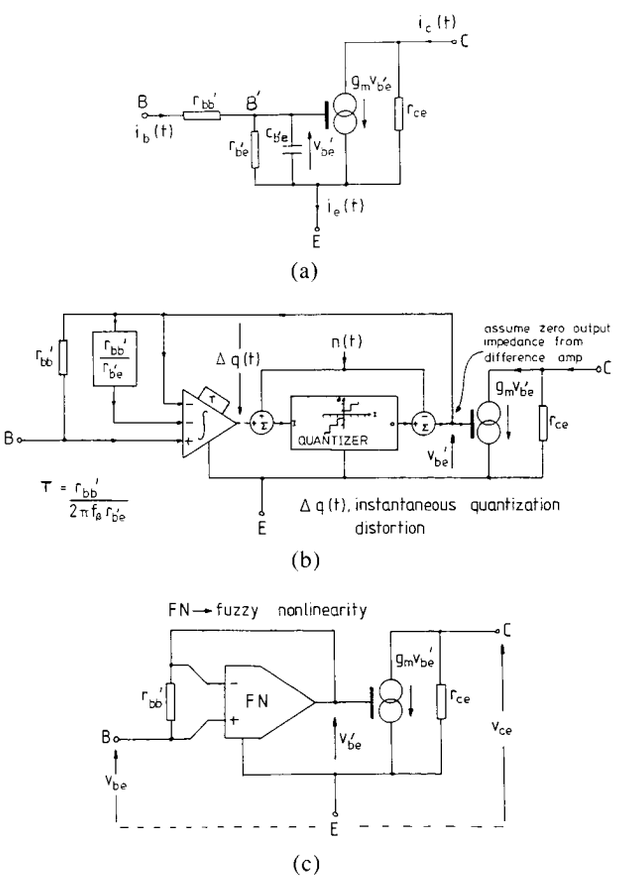


Fig. 2. Basic hybrid-π equivalent circuit of a BJT. (a) Standard circuit. (b) Circuit with modification to incorporate charge quantization. (c) Simplified functional presentation, including charge quantization.

the fuzzy model and digital encoders is that the former excludes a uniform sampling process. However, a random sampling function is permissible where the mean sampling frequency corresponds to the mean rate of recombination within the base of the transistor, which is determined by the base bias current (that is, a base bias current of 2 μA corresponds to a mean sampling rate of $\cong 10^{13}$ Hz). We note also from Eq. (2) that the mean sampling rate will undergo frequency modulation due to the instantaneous change in recombination current with change in base–emitter voltage.

We estimate the approximate frequency characteristic of the distortion spectra for the model of Fig. 2(b) by assuming the loop to be essentially linear and by representing the quantization distortion as a sinusoidal error signal added within the loop where, for purposes of analysis,

$$q(t) = Qe^{j2\pi ft} . \tag{18}$$

Thus the collector error current $I_{q,c}e^{j2\pi ft}$ follows as

$$I_{q,c} = \frac{jQg_m f/f_\beta}{\left[1 + \frac{r_{b'e}}{r_c + r_{bb'}}\right]\left[1 + \frac{jf}{[1 + r_{b'e}/(r_c + r_{bb'})]f_\beta}\right]} \tag{19}$$

where r_c is the source resistance between base and emitter, as shown in Fig. 1.

From Eq. (19) we infer the basic form of error spectrum, which is illustrated in Fig. 3. Note the effect a low source impedance has on the break frequency in the approximate error spectrum.

The error spectrum shown in Fig. 3 compares with the general trend of pulse-code-modulation-type systems [16], [17] where quantization is dominant at high frequencies. The curve ignores other forms of random noise, such as the noise associated with $r_{bb'}$. Thus in general this effect will be at or below the device noise level.

The results show that the source resistance plays a dominant role in shaping the error spectrum where optimum performance is obtained when r_c is minimized. This compares favorably with the more common noise model of a transistor where the noise sources are represented as equivalent input noise voltage and current generators. An interesting by-product of the model structure is that it includes a mechanism that modifies the output noise as a function of source resistance, this being achieved by modifying the feedback factor around the internal feedback loop. Essentially when $r_c = 0$, maximum feedback is applied, and when $r_c = \infty$, minimum feedback results. Examination of the model schematic illustrated in Fig. 2(b) should clarify this operation.

In this section we have established a modification to the basic hybrid-π model of a transistor which includes the effect of charge quantization. We now proceed to examine some implications of these observations.

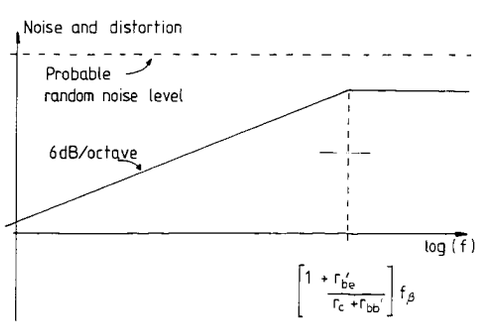


Fig. 3. Approximate error spectrum of collector current due only to quantization effects.

3 IMPLICATIONS OF FUZZY DISTORTION IN AMPLIFIER DESIGN

If we accept that a low-level nonlinear mechanism exists in transistors which has a different nature from deterministic nonlinearity, then we can make some basic observations as to the correct global strategy toward amplifier design.

Where a transistor operates with very-low-level signals that approach the noise floor, the artifacts of charge quantization will generate significant fuzzy distortion. The method to minimize this effect can be summarized as follows:

1) It is essential that low-noise devices be used that exhibit low $r_{bb'}$ and low $1/f$ (recombination noise). The device should be chosen so as to maximize $C_{b'e}$. Thus large integrated arrays of transistors where many matched devices are paralleled should prove the best choice (such as the LM394).

2) Operate transistors so that $C_{b'e}$ is maximized. From Eq. (1) this infers a substantial level of emitter bias current which in turn will lower the device input impedance.

3) Eq. (1) infers that $C_{b'e}$ is an inverse function of $r_{b'e}$ (for given f_β). Thus a device should be chosen with a low value of h_{fe} [see Eq. (8)].

4) Selection of f_β is more complex. A low level of f_β will increase $C_{b'e}$, but at the expense of lowering the break frequency in the distortion spectra (see Fig. 3). It is suggested that f_β should be sensibly in excess of 20 kHz.

5) Design the transistor stage so as to maximize the device loading factor [18]. This will maximize the changes in charge for a given signal.

6) Minimize resistance in the input mesh of a transistor. This will reduce low-frequency fuzzy distortion. For the input stage of Fig. 1 this implies a low value of r_c. The effect of r_c can be observed by reference to Eq. (19) and the error spectrum in Fig. 3.

Consider by way of example a low-level disk preamplifier stage for use with a low-output moving coil cartridge. In Fig. 4 a moving-coil cartridge with source resistance r_c and generator signal $e_c(t)$ is interfaced to

a disk amplifier which has an input resistance r_{in} and a voltage gain A_v.

The classical viewpoint would not expect r_{in} to play an important role other than providing an optimum load for the cartridge. (This may affect the frequency response, for example, when coupled with the generator source inductance.) However, for low-output-impedance moving-coil cartridges this generally has minimal effect. Indeed in selecting an "optimum" load resistance, it is normal to use an input shunt resistor.

However, fuzzy nonlinearity suggests that the level of i_{in} is of fundamental importance, and that this current must be maximized and flow into the base of the input transistor. A shunt input resistance is not an acceptable solution, as current will by-pass the transistor.

It therefore follows that the input signal must be considered in terms of both input voltage and input current. It is the input signal power that is fundamental.

3.1 Corollary 1

If we accept the notion of maximizing the signal power that flows into the base of the input transistor, then a transducer for an analog disk system must be selected such that

1) It converts a relatively high proportion of platter rotational energy into mechanical signal energy, as seen at the cantilever of the cartridge.

2) It exhibits a high mechanical-to-electrical power conversion.

It is possibly in these areas of performance where many moving-coil cartridges offer a significant performance advantage.

3.2 Corollary 2

In selecting a matching transformer/input circuit topology, the aim must be to maximize the flow of signal power into the base of the input transistor.

The proposal to maximize the input power is open to some debate. However, if it is realized that we wish both to maximize input signal current to the base of each transistor and to minimize source resistance r_c, then the notion of power maximization is a reasonable target.

Corollary 2 has profound ramifications in the choice circuit topology. Consider the classical amplifier configuration shown in Fig. 5. The circuit shows an input signal generator e_c with source impedance r_c. Again the amplifier has an input impedance r_{in} and voltage gain A_v, but a negative-feedback loop is included where the feedback factor is B with a Thévenin source impedance (seen by the inverting input) of r_f.

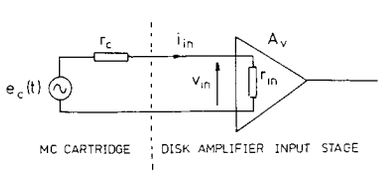


Fig. 4. Moving-coil cartridge–amplifier interface.

We proceed by calculating the instantaneous input signal power $p_{in}(t)$ to the differential input of the amplifier,

$$p_{in}(t) = \left[\frac{e_c(t)}{(r_c + r_f) + r_{in}(1 + AB)}\right]^2 r_{in} . \quad (20)$$

Differentiating $p_{in}(t)$ ωrt r_{in},

$$\frac{\partial p_{in}(t)}{\partial r_{in}} = \left[\frac{e_c(t)}{(r_c + r_f) + r_{in}(1 + AB)}\right]^2 \times \left[1 - \frac{2r_{in}(1 + AB)}{(r_c + r_f) + r_{in}(1 + AB)}\right] \quad (21)$$

and setting $\partial p_{in}(t)/\partial r_{in} = 0$ to maximize the input power, the optimum r_{in} (for maximum input power) follows as

$$r_{in}\Big|_{opt} = \frac{r_c + r_f}{1 + AB} . \quad (22)$$

This gives the maximum input power as

$$p_{in}(t)\Big|_{max} = \frac{e_c^2(t)}{4(r_c + r_f)(1 + AB)} . \quad (23)$$

Eq. (23) shows the need to minimize all extraneous resistances within the input signal mesh, which is also a requirement for good noise design (that is, minimize r_f). [See also Eq. (19) and the discussion in Section 2 concerning input mesh resistance and fuzzy distortion.]

However, a more fundamental observation shows the maximum power flow to be an inverse function of the feedback parameter. Thus although classical feedback theory would suggest an improvement by operating the device well into its linear region of operation, it in fact forces the signal to within a relatively few quanta, thus exaggerating any effects of quantization.

To illustrate the process further, consider the combined systems of Figs. 2(c) and 5, as shown in Fig. 6.

Any amplification which follows the quantization process must by necessity amplify the quantized signal, together with additional random noise sources. The effect of negative feedback on a purely linear system will reduce the levels of additional noise resources that are injected within the feedback loop by a factor of $(1 + AB)$. However, this process is not true of a loop that includes quantization. In fact in this system the feedback will again reduce the additive noise, but it will only

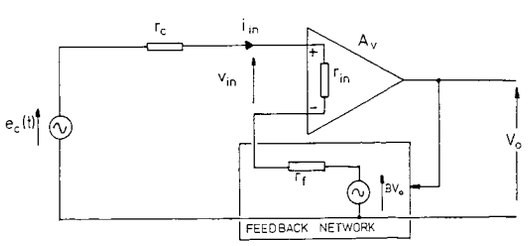


Fig. 5. Classical feedback amplifier structure.

partially reduce the effects of quantization. Thus fuzzy distortion components will be partly exposed by negative feedback, together with a complex process of intermodulation between signal additive noise and quantization distortion, which in general must include time smearing due to limitations in loop bandwidth.

4 MINIMIZATION OF FUZZY NONLINEARITY

We conclude our discussion on fuzzy distortion by suggesting a design method and basic circuit topologies that in principle meet the requirements of both high-level and low-level nonlinearities. In particular we emphasize low-level signal stages as these are potentially more susceptible to fuzzy nonlinearity.

Following the design aims discussed in Section 3, we must choose a low-noise transistor with a low value of collector–base current gain. This device should be operated at a collector current commensurate with noise considerations such that (ideally) the input impedance between base and emitter matches the source impedance of the transducer or presents an optimum load to the transducer. Provided the source signal is of suitable magnitude, the signal should be coupled directly to the base–emitter junction (assuming that high-level distortion will not be problematic), and preferably no ac coupling component should be used.

Coupled with this requirement, the input transistor should ideally use no feedback (local or overall), since Eq. (23) indicates a reduction in signal power. If the transducer output is too great, resulting in a high level of deterministic distortion, then a step-down transformer should be selected to permit using a zero feedback input stage. Ideally the input impedance should be designed to match the transformed source impedance of the transducer. This process will not change (in principle) the level of power extracted from the source (assuming a power match), but it will minimize high-level distortion and eliminate a loss of input power through the use of negative feedback. In many instances it will not be practical to design for a power match as high operating currents or many parallel devices may be necessary, though investigation into the LM394-type device should be encouraged.

In general an amplifier system will include several cascaded transistor stages within the signal path. Potentially each stage is a cause of low-level distortion, but as with noise design, the first transistor should be the dominant offender. To guarantee this ideal we must arrange for a progressive increase in input signal power as we proceed along the cascade, as well as attempting to minimize the number of series-connected transistors in the signal path.

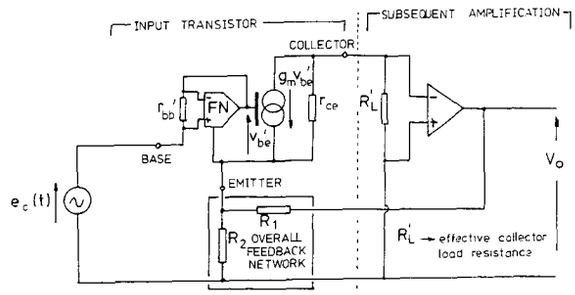


Fig. 6. Simplified feedback amplifier with quantizer model of Fig. 2(c).

In order to control deterministic distortion as signal levels are amplified, a degree of negative feedback will become mandatory, which will consist generally of a combination of distributed and multiple-feedback loops. However, in selecting the topology for the feedback structure, an increasing input-signal-power progression should be observed.

To examine this design strategy, consider $N + 1$ cascaded transistor stages, as shown in Fig. 7. Stages $1 \rightarrow N$ use distributed feedback, while the input stage 0 is optimized for fuzzy nonlinearity by using zero feedback. A single feedback loop encloses stages $1 \rightarrow N$, thus modeling a typical amplifier.

Let

$A_0, \ldots, A_N$ = amplifier gains
$B_0, \ldots, B_N$ = feedback factors[1]
$e_0, \ldots, e_N$ = amplifier input signals
} see Fig. 7

$r_0, \ldots, r_N$ = amplifier input resistances
$p_0, \ldots, p_N$ = input signal powers to amplifiers
r_c = transducer source impedance.

From Eq. (20) we calculate the rth-stage input signal power p_r. For stages $r = 1, \ldots, N$. (Assume that the source resistance is small compared with r_r.)

$$p_r = \left[\frac{e_r}{1 + A_r B_r}\right]^2 \frac{1}{r_r} \tag{24}$$

and for stage $r = 0$,

$$p_0 = \left(\frac{e_0}{r_c + r_0}\right)^2 r_0 \ . \tag{25}$$

4.1 Design Criterion

To minimize signal degradation caused by fuzzy nonlinearity in a cascade of transistor stages,

$$p_r = G_{fr} p_{r-1} \tag{26}$$

where ideally the interstage power gain $G_{fr} > 1$. We calculate the voltage gain relating e_r and e_{r-1}, for $r = 1$,

$$\frac{e_1}{e_0} = \frac{A_0}{\left\{1 + B_0 \prod_{p=1}^{N} [A_p/(1 + A_p B_p)]\right\}} \tag{27}$$

and for $r = 2, \ldots, N$,

$$\frac{e_r}{e_{r-1}} = \frac{A_{r-1}}{1 + B_r A_r} \ . \tag{28}$$

Hence we establish the constraints on the choice of feedback parameters by reference to Eqs. (24)–(28).

[1] The feedback networks are assumed to exhibit zero Thévenin source impedances.

If

$$\delta_{fl} = \left\{1 + B_0 \prod_{p=1}^{N} \frac{A_p}{1 + B_p A_p}\right\}\left\{\frac{\sqrt{G_{fl} r_0 r_1}}{r_c + r_0}\right\} \quad (29)$$

$$\delta_{fr} = \sqrt{\frac{G_{fr} r_r}{r_{r-1}}}\Bigg|_{r=2,\ldots,N} \quad (30)$$

we have

$$A_{r-1} = \delta_{fr}(1 + B_r A_r)\Big|_{r=1,\ldots,N} \quad (31)$$

where we define $\delta_{fr}|_{r=1,\ldots,N}$ as the set of fuzzy gain parameters of the amplifier system.

Examination of Eqs. (29)–(31) reveals the design criterion that will ensure a progressive power increase along the cascade of transistor stages (noting that calculated power levels refer to the input power to each transistor, not the associated circuitry).

In practice there will be a limit to the input power to a transistor that will be dependent on the acceptable levels of deterministic distortion. We note that for a bipolar transistor which adheres to the form of Eq. (17) the fractional error component of emitter current is independent of I_{EO} for a given V_{BE} (where the subscript EO infers quiescent values),

$$V_{BE} = V_{BEO} + \Delta V_{BE}$$

which corresponds to $I_E = I_{EO} + \Delta I_E$. Then

$$\frac{\Delta I_E}{I_{EO}} = e^{(q\Delta V_{BE}/KT)} - 1 . \quad (32)$$

Since the base–emitter voltage of a transistor is directly dependent upon the input power and input resistance, the input resistance should be minimized to reduce high-level distortion, for a given power level.

It is constructive to reflect upon a common circuit arrangement where a discrete transistor stage is cascaded with a BJT operational amplifier with local feedback. We will assume for simplicity that there is no overall feedback and proceed by suggesting typical circuit parameters:

Discrete stage		
input impedance	1 kΩ	(transistor)
voltage gain	20	
Operational amplifier stage		
input impedance	1 MΩ	(operational amplifier)
closed-loop gain	20	
open-loop gain	1000	(conservative estimate)

Hence from Eq. (31) $\delta_{fr} \cong 0.4$, whereby the power gain follows from Eq. (30) as 1.6×10^{-4}.

This result shows a substantial reduction of input signal power presented to the second stage, the consequence being that any quantization effects will be significantly increased.

This circuit arrangement has often been used for disk preamplifiers and is a good illustration of a potential hazard of using high-gain high-input-impedance operational amplifiers.

We conclude this section by suggesting how it is possible to use a combination of low distributed feedback with feedforward error correction as a compromise to the distortion dichotomy existing between deterministic and fuzzy nonlinearities.

Germane to the design strategy is the selection of a distributed feedback system where the appropriate gains and feedback factors are calculated according to our now established fuzzy nonlinearity criterion. In so doing we accept that deterministic nonlinearity inherent in devices will potentially increase. However, by using nested feedforward error correction we can partially compensate the deterministic error signals and achieve acceptable linearity with high loading factors, even when local negative feedback is low.

In Fig. 8 we illustrate a two-stage feedforward amplifier where the error due to base–emitter nonlinearity in T_1 is partially corrected by the differential amplifier formed by T_2 and T_3. Further error-correction stages can be used to compensate for T_2 and T_3 nonlinearity using a nested configuration. The performance of such stages as a function of loading factor was considered in a previous paper [18].

The dominant advantages of this approach is that only very modest local negative feedback need be applied via R_1 and that the high-level distortion is partially compensated by the error amplifier. Such a technique allows good signal power coupling to T_1, yet permits an acceptable high-level distortion characteristic. It therefore follows that T_1 is exercised over a wide range of its operating characteristic while retaining good overall linearity.

The example just discussed illustrates how ideas of fuzzy nonlinearity could influence amplifier design. A second area of application concerns the construction and layout of circuits. Once the very small signal levels are appreciated and the point of view of "counting electrons" is taken, such factors as metal–metal con-

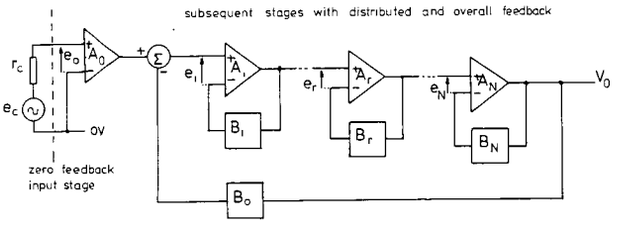


Fig. 7. Basic multiple-loop feedback amplifier topology.

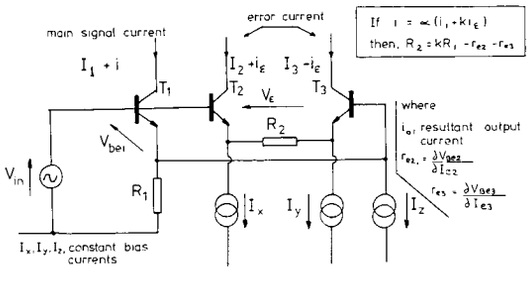


Fig. 8. Basic feedforward error-correction stage.

tacts, interference from adjacent circuits, and the displacement of charge through the dielectric of a capacitor require careful attention. These secondary factors will not be discussed in this paper, but they are influential in setting potential limits to signal transparency.

5 CONCLUSIONS

This paper has speculated on the existence of low-level nonlinearity inherent within BJT devices due to the quantization of charge carriers and has drawn attention to the relative magnitude of low-level signals. A model was introduced which forms a vehicle for comparison between an analog device and a class of digital modulation. This comparison is useful in that it is possible to speculate upon the nature and characteristics of the distortion.

Consideration of the mechanism of fuzzy distortion drew attention to the role of the input signal current at the base of a transistor and the need to maximize its value. This led directly to the usefulness of input signal power as a parameter in establishing levels of fuzzy distortion. A target design objective suggested the need to maximize this power flow, where the flow must be directly into the base–emitter junction and not into an external shunt resistor.

The role of negative feedback was then debated, where it was shown that the input signal power was an inverse function of amplifier loop gain. It was therefore concluded that levels of feedback should be minimized and that extraneous resistance within the input mesh should also be minimized. However, it was noted that the role of negative feedback offers a contribution to high-level signal distortion, but that its application must be considered with great care from the viewpoint of fuzzy nonlinearity.

A brief discussion was presented where the bounds on the selection of feedback factor and forward amplification were established. Finally a circuit technique using low levels of distributed feedback with feedforward error correction was introduced as a means of circumventing the distortion dichotomy, thus allowing both good low-level and high-level distortion characteristics, that is, the dynamic range.

It is satisfying to see some of the design objectives compatible with established design techniques which are used to minimize the artifacts of TID and also as support to the low-feedback school of design, in particular since there are now several good-quality amplifiers which adhere in part to these design objectives and also have excellent subjective ratings.

Finally it must be emphasized that the ideas presented here are the extension of a thought experiment into the approximate nature and behavior of low-level signals in amplifiers. Clearly such parameters as the physical size of transistors and the relative amounts of total charge stored in the base region are of importance. However, such considerations put in doubt the application of BJT operational amplifiers with their high open-loop gains, very high differential input impedance due to the low collector bias currents in the input transistors, and low real estate. If such devices exhibit low-level quantum effects, they are not suitable for use in high-quality audio amplifiers where precision of control of fine signal detail is mandatory. In fact, applying the thought experiment discussed in Section 1, the implication of Eq. (23), and the example of the BJT operational amplifier in Section 4, the potential consequences should at least be of concern to the circuit designer.

6 REFERENCES

[1] N. Keywood, "Amplifier Review," *Hi-Fi News*, vol. 27, pp. 37–45 (1982 Aug.).

[2] J. Hiraga, "Amplifier Harmonic Distortion Spectrum Analysis," *Hi-Fi News*, vol. 22, pp. 41–45 (1977 Mar.).

[3] C. Ray, "Negative Feedback and Non-linearity," *Wireless World*, vol. 84, pp. 47–50 (1978 Oct.).

[4] M. Otala, "Transient Distortion in Transistorized Audio Power Amplifiers," *IEEE Trans. Audio Electroacoust.*, vol. AU-18, pp. 234–239 (1970 Sept.).

[5] M. Otala, "Circuit Design Modifications for Minimizing Transient Intermodulation Distortion in Audio Amplifiers," *J. Audio Eng. Soc.*, vol. 20, pp. 396–399 (1972 June).

[6] P. Garde, "Transient Distortion in Feedback Amplifiers," *J. Audio Eng. Soc.*, pp. 314–322 (1978 May).

[7] W. M. Leach, "An Amplifier Input Stage Design Criterion for the Suppression of Dynamic Distortion," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 29, pp. 249–251 (1981 Apr.).

[8] J. Vanderkooy and S. P. Lipshitz, "Feedforward Error Correction in Power Amplifiers," *J. Audio Eng. Soc.*, vol. 28, pp. 2–16 (1980 Jan./Feb.).

[9] M. J. Hawksford, "Distortion Correction in Audio Power Amplifiers," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 29, pp. 27–30 (1981 Jan./Feb.).

[10] S. Takahashi and S. Tanaka, "Design and Construction of a Feedforward Error-Correction Amplifier," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 29, pp. 31–37 (1981 Jan./Feb.).

[11] Y. Hirata, M. Ueki, T. Kasuga, and T. Kitamura, "Nonlinear Distortion Measurement Using Composite Pulse Waveform," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 29, pp. 243–248 (1981 Apr.).

[12] Y. Hirata, "Quantifying Amplifier Sound," *Wireless World*, vol. 87, pp. 49–52 (1981 Oct.).

[13] R. West, "The Great Amplifier Debate, 2-Transistor Sound," *Hi-Fi News*, vol. 23, p. 79 (1978 Jan.).

[14] S. Curtis, "Amplifier Noise and Clipping," *Hi-Fi News*, vol. 23, pp. 81–85 (1978 June).

[15] R. Beaufoy and J. J. Sparkes, "The Junction Transistor as a Charge Control Device," *ATE J.*, vol. 13, pp. 310–327 (1957 Oct.).

[16] R. Steele, *Delta Modulation Systems* (Pentech Press, London, 1975).

[17] M. J. Hawksford, "Unified Theory of Digital Modulation," *Proc. IEE*, vol. 121, pp. 109–115 (1974 Feb.).

[18] M. J. Hawksford, "Distortion Correction Circuits for Audio Amplifiers," *J. Audio Eng. Soc.*, vol. 29, pp. 503–510 (1981 July/Aug.).

THE AUTHOR

Malcolm Hawksford was educated at the University of Aston in Birmingham, England, from 1965 to 1971. In 1968 he obtained a first class honors degree in electrical engineering, and that same year was awarded a BBC research scholarship to investigate the application of deltamodulation to color television. In 1972 he obtained a Ph.D. degree. In 1971 Dr. Hawksford became a lecturer at the University of Essex, England, in the Department of Electrical Engineering Science. During his time at Essex he actively pursued research projects within the field of audio engineering, where projects on power amplifier design, loudspeaker crossover design, analog-to-digital conversion, and music synthesis have been undertaken. He has presented papers at conventions of the Audio Engineering Society. He is currently a member of the AES, IEE, and RTS.

Optimization of the Amplified-Diode Bias Circuit for Audio Amplifiers*

M. J. HAWKSFORD

University of Essex, Department of Electrical Engineering Science, Colchester, Essex, UK

An economic enhancement to the conventional "amplified diode" bias circuit is presented for use in power amplifier circuit topologies which do not allow precise, temperature invariant control of the operating current of the bias circuit. In essence, the modification minimizes the sensitivity of the derived bias voltage to changes in operating current without compromising the desirable temperature tracking properties when thermally bonded to the complementary follower output cell.

0 BACKGROUND

The output stage of power amplifiers and some operational amplifiers requires circuitry to allow precision control of the output bias current. The classical approach is to use a bias network that consists of either a series of diodes [1] or a transistor with local feedback in a circuit called an amplified diode [2], [3]. In Fig. 1 we illustrate the basic output-cell topology for the complementary follower configuration.

The circuit objective is to produce a dc offset V_B between the base connections of the output devices to compensate for the ON bias voltage that is required to establish the output-device bias quiescent current I_Q.

Although in practice emitter resistors R_e are used as local series feedback elements to help stabilize changes in I_Q with changes in device temperature and circuit parameters, their use only degenerates performance in other areas by increasing the output resistance of the stage and by making the output resistance a nonlinear function of output current (especially in class AB stages).

Consequently, as is well known, the use of a bias network that in principle can track changes in output-device temperature is necessary to control I_Q within reasonable bounds and thus allows low or zero values of R_e to be employed. Such networks are generally of the type shown in Fig. 1, where the bias devices (diodes or transistor) should be in close thermal contact with the output cell for good temperature tracking.

However, one area of bias network design that has been given little attention is the variability of V_B with changes in operating current I (see Fig. 1). We define a function $S_I^{V_B}$, which is a measure of this dependency,

$$S_I^{V_B} = \frac{\partial V_B}{\partial I} . \qquad (1)$$

In many amplifier circuits the quiescent value of I can change as a function of temperature, where in general the trend is for a positive temperature coefficient, that is, I increases with temperature. To some extent this can be compensated by a suitable choice of feedback structure to the input stage, but this may well compromise other areas of performance and prove impractical to implement in a low-feedback amplifier.

This communication addresses the optimization of $S_I^{V_B}$ and suggests a simple modification to the design of the amplified diode which allows $S_I^{V_B}$ to be zero or indeed negative, thus reducing the tendency for increased output device bias current I_Q with temperature due to changes within the input stage of the amplifier.

1 THE MODIFIED AMPLIFIED DIODE

The modification to the basic amplified diode that enables optimization of $S_I^{V_B}$ is shown in Fig. 2 where

* Manuscript received 1983 Aug.

I_E is the emitter current and V_{BE} the base emitter voltage.

The addition to the circuit is the resistor R_3. To investigate the operation of the modified circuit, we calculate the parameter $S_I^{V_B}$. We proceed by choosing resistors R_1 and R_2 such that the current in R_1 is

$$I_{R1} = \sqrt{I_C I_B} = \frac{I_C}{\sqrt{\beta}} \tag{2}$$

where β is the current gain, I_C the collector current, and I_B the base current of T_1. Thus as an example, if $\beta = 100$, then $I_{R1} \cong 10I_B$ and $I_C \cong 10I_{R1}$. This will therefore realize good bias stability within the amplified diode. Consequently we may assume (for high B) that

$$I_C \cong I \ . \tag{3}$$

Hence,

$$(V_B + IR_3)\frac{R_1}{R_1 + R_2} = V_{BE}$$

and thus,

$$V_B = \left(1 + \frac{R_2}{R_1}\right)V_{BE} - IR_3 \ . \tag{4}$$

Differentiating V_B with respect to I to determine $S_I^{V_B}$, we have

$$S_I^{V_B} = \left(1 + \frac{R_2}{R_1}\right)\frac{\partial V_{BE}}{\partial I} - R_3 \ . \tag{5}$$

Since $I_E \cong I$, then from the diode equation,

$$I_E = I_S e^{qV_{BE}/KT} \cong I$$

where I_S is the transistor saturation current, q is the charge on an electron, K is Boltzmann's constant, and T is the junction temperature.

Differentiating, $\partial V_{BE}/\partial I \cong KT/qI$, and thus,

$$S_I = \left(1 + \frac{R_2}{R_1}\right)\frac{KT}{qI} - R_3 \ . \tag{6}$$

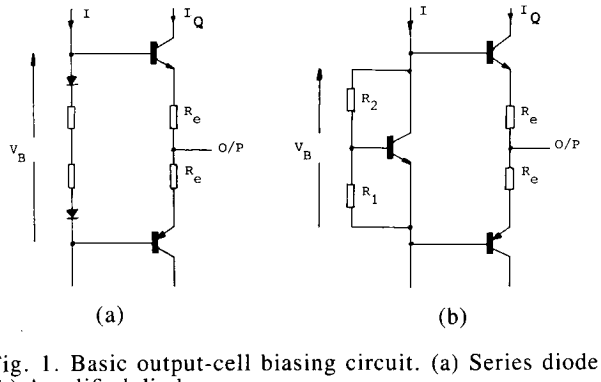


Fig. 1. Basic output-cell biasing circuit. (a) Series diode. (b) Amplified diode.

At room temperature $KT/q \cong 0.025$ V, and

$$S_I^{V_B} = \frac{0.025}{I}\left(1 + \frac{R_2}{R_1}\right) - R_3 \ . \tag{7}$$

Eq. (7) allows R_3 to be selected to minimize $S_I^{V_B}$. Hence for zero S_I, where optimum $R_3 = R_{3\,\mathrm{opt}}$,

$$R_{3\,\mathrm{opt}} = \frac{0.025}{I}\left(1 + \frac{R_2}{R_1}\right) \ . \tag{8}$$

Under this condition the bias voltage V_B is to a good first-order approximation independent of I. However, in practice it is suggested that $R_3 > R_{3\,\mathrm{opt}}$, thus implying a negative S_I. This will counteract tendencies for thermal runaway with increasing ambient temperature. Also, as $R_{3\,\mathrm{opt}}$ is temperature dependent, the value of R_3 should be calculated at the maximum device temperature. Thus for lower temperatures S_I will be slightly negative.

In circuit applications requiring a bias voltage V_B of several V_{BE}, such as where Darlington output devices are used, two transistors may be used as shown in Fig. 3.

2 CONCLUSIONS

This communication has discussed a modification to the basic amplified–diode circuit which will minimize the dependency of the bias voltage V_B on the magnitude of the amplified-diode operating current. The modification is simple, yet has proved to be extremely effective in operation. It is important on two counts. First, it will minimize changes in output-cell bias current due to changes in the driving circuit, which will generally be temperature dependent. Also in circuits that use a differential drive current to the amplified diode, it will

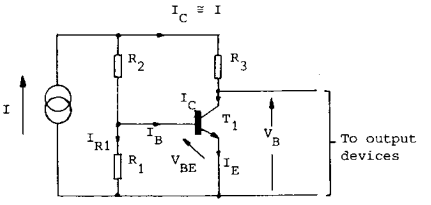


Fig. 2. Modified amplified-diode circuit.

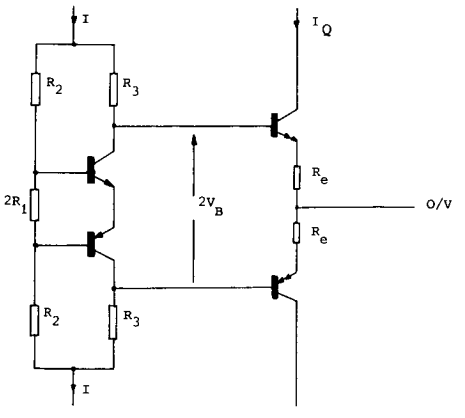


Fig. 3. Two-transistor amplified-diode circuit for use with Darlington output transistor.

minimize common-mode current variations with this drive configuration. Finally it should be noted that the modification in no way compromises the performance of the amplified diode with respect to thermal tracking of the output devices since for constant I, the voltage IR_3 is almost independent of temperature where, from Eq. (4),

$$\frac{\partial V_B}{\partial T} = \left(1 + \frac{R_2}{R_1}\right) \frac{\partial V_{BE}}{\partial T} . \quad (9)$$

3 REFERENCES

[1] P. Horowitz and W. Hill, *The Art of Electronics* (Cambridge University Press, London, 1980), pp. 75–77.

[2] M. J. Hawksford, "Distortion Correction in Audio Power Amplifiers," *J. Audio Eng. Soc. (Engineering Reports)*, vol. 29, pp. 27–30 (1981 Jan./Feb.).

[3] E. M. Cherry, "Feedback, Sensitivity, and Stability of Audio Power Amplifiers," *J. Audio Eng. Soc.*, vol. 30, pp. 282–294 (1982 May).

THE AUTHOR

Malcolm Hawksford was educated at the University of Aston in Birmingham, England, from 1965 to 1971. In 1968 he obtained a first class honors degree in electrical engineering, and that same year was awarded a BBC research scholarship to investigate the application of deltamodulation to color television. In 1972 he obtained a Ph.D. degree. In 1971 Dr. Hawksford became a lecturer at the University of Essex, England, in the Department of Electrical Engineering Science. During his time at Essex he has actively pursued research projects within the field of audio engineering, where projects on power amplifier design, loudspeaker crossover design, analog-to-digital conversion, and music synthesis have been undertaken. He has presented papers at conventions of the Audio Engineering Society. He is currently a member of the AES, IEE, and RTS.

Reduction of Transistor Slope Impedance Dependent Distortion in Large-Signal Amplifiers*

MALCOLM HAWKSFORD

University of Essex, Department of Electronic Systems Engineering, Colchester CO4 35Q, UK

0 INTRODUCTION

The static characteristics of a bipolar transistor reveal that, under large-signal excitation, there are sources of significant nonlinearity. In an earlier paper [1] consideration was given to the I_E/V_{BE} nonlinearity, where a family of techniques was presented to attempt local correction of this error mechanism. However, the collector–emitter and collector–base slope impedance of transistors also result in significant distortion, where under large-signal conditions they can become a dominant source of error [2].

The static characteristics show only part of the problem; a more detailed investigation reveals capacitive components which are dependent upon voltage and current levels. Consequently under finite-signal excitation, modulation of the complex slope impedances results in dynamic distortion. It will be shown that the level of error that results from slope distortion is not strongly influenced by negative feedback once certain loop parameters are established. Also, because of the frequency and level dependency of slope distortion, the overall error will contain components of both linear and nonlinear distortion that are inevitably linked to individual device characteristics. It is therefore anticipated that a change of transistor could, in principle, lead to a perceptible change in subjective performance, even when the basic dc parameters are similar.

In this paper consideration is given to a class of voltage amplifiers employing a transconductance gain cell g_m, a gain-defining resistor R_g, and a unity-gain isolation amplifier, together with an overall negative-feedback loop. This structure is typical of most voltage and power amplifiers. However, although it is more usual to focus attention on input stage and output stage distortion, we shall consider in isolation the distortion due only to slope impedance modulation and assume other distortions are controlled to an adequate performance level. It will be demonstrated that significant distortion results from slope modulation, and a design methodology is presented to virtually eliminate its effect, even when the slope parameters are both indeterminate and nonlinear and when signals are of substantial level.

We commence our study by investigating the role of negative feedback as a tool for the reduction of slope distortion and to show that although effective, in isolation, it is not an efficient procedure.

1 NEGATIVE FEEDBACK AND THE SUPPRESSION OF SLOPE IMPEDANCE DEPENDENT DISTORTION

Consider the elementary amplifier shown in Fig. 1, where the principal loop elements are transconductance g_m, gain-defining resistor R_g, and feedback factor k. The nonideality of the transconductance cell is represented by an output impedance Z_n, where ideally $Z_n = \infty$, but in practice is finite and signal dependent. (Any linear resistive component of Z_n is assumed isolated and lumped with R_g.) In general, Z_n is a composite of the slope parameters of the output transistors in the transconductance cell. It can also include a reflection of any load presented to the amplifier. However, we assume here a perfect unity-gain buffer amplifier to isolate the slope distortion of the transconductance cell.

Although Z_n is signal dependent, our analysis will assume small-signal linearity so that performance sensitivity to Z_n can be established. However, the circuit topologies presented in Sec. 3 are not so restricted and can suppress the nonlinearity due to Z_n modulation.

For a target closed-loop gain γ there is a continuum of k and R_g for a given g_m, where the target closed-loop gain γ for $Z_n = \infty$ is defined,

$$\gamma = \frac{g_m R_g}{1 + k g_m R_g} . \tag{1}$$

Hence for a given k, g_m, and γ, R_g is expressed as

$$R_g = \frac{\gamma}{g_m(1 - \gamma k)} \tag{2}$$

* Manuscript received 1987 June 22.

where, for $0 \leq k \leq 1/\gamma$, then $\gamma/g_m \leq R_g \leq \infty$.

The actual closed-loop gain A, for finite Z_n, is

$$A = \frac{g_m Z_n R_g}{Z_n + R_g + k g_m Z_n R_g} \tag{3}$$

and eliminating R_g defined by Eq. (2) for selected target gain γ and transconductance g_m,

$$A = \frac{g_m Z_n}{1 + Z_n g_m/\gamma} . \tag{4}$$

This result demonstrates that the dependence of the transfer function A on Z_n is independent of the selection of feedback factor k, provided the condition of Eq. (2) is satisfied to set the target gain γ.

The error contribution due to Z_n can be estimated by evaluation of the transfer error function [3], [4] E defined by

$$E = \frac{A}{\gamma} - 1 \tag{5}$$

where E represents the ratio of error signal to primary signal and can be visualized according to Fig. 2.

Substituting A from Eq. (4) into Eq. (5),

$$E = \frac{-\gamma}{\gamma + g_m Z_n} . \tag{6}$$

In practice $g_m Z_n >> \gamma$ for a well-behaved amplifier, whereby

$$E \simeq \frac{-\gamma}{g_m Z_n} . \tag{7}$$

The results of Eqs. (6) and (7) reveal that to reduce the dependence on slope distortion, the product $\{g_m Z_n\}$ must increase. However, it is important to observe that Z_n reduces with increasing frequency due to device capacitance and that g_m also reduces with frequency due to closed-loop stability requirements, so that there are fundamental constraints on the effectiveness of slope distortion reduction using overall negative feedback, particularly at high frequency.

As an aside we are assuming g_m to be linear. In practice a reduction of R_g places a heavier current demand on g_m; thus a greater distortion contribution from g_m is to be anticipated for a given output [1]. Also, in power amplifier circuits, the output stage will exhibit distortion under load, a factor not considered in the present discussion. However, the independence of E on k and R_g for a given γ and g_m is true for distortion resulting only from Z_n, and when considered in isolation, it is an interesting example of a distortion that is not reduced by moving from a zero-feedback to a negative-feedback topology, especially as the choice of R_g is often the principal distinction between low-feedback and high-feedback designs [5].

In the next section the common-emitter amplifier is examined as a transconductance cell and current mirror, and an estimate is made of the output impedance Z_n for a range of circuit conditions.

2 OUTPUT IMPEDANCE OF COMMON-EMITTER AMPLIFIER

The common-emitter amplifier is shown in Fig. 3 in both single-ended and complementary formats. In this section the output impedance of the common-emitter amplifier is analyzed in terms of the small-signal parameters for a range of source resistances R_s and emitter resistances R_E. For analytical convenience, the base and emitter bulk resistances are assumed lumped with R_s and R_E, respectively.

Fig. 4 illustrates a small-signal transistor model of the common-emitter cell, where z_{ce} and z_{cb} represent collector–emitter and collector–base slope impedances, respectively, and h_{fe} is the collector–base current gain.

The output impedance Z_c observed at the collector of the common-emitter cell is given by

$$Z_c = \frac{v_o}{\alpha i_o} = \frac{1}{\alpha}\left\{ z_{ce} + R_E + \frac{z_{ce}}{z_{be}} [R_E + R_s(1 - \alpha)] (1 + h_{fe}) \right\} \tag{8}$$

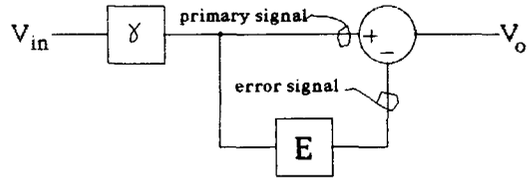


Fig. 2. Transfer error function model of voltage amplifier in Fig. 1.

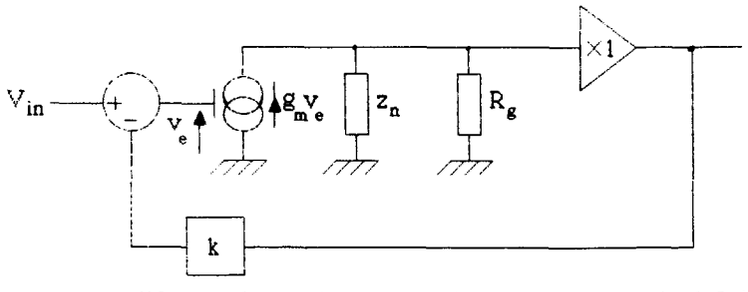


Fig. 1. Elementary amplifier topology using transconductance cell and gain-defining resistor.

where the collector/emitter current division factor is

$$\alpha = 1 + \frac{z_{be}z_{ce} + R_E\lambda}{z_{be}z_{cb} + R_s\lambda} \quad (9)$$

and

$$\lambda = (1 + h_{fe})z_{ce} + z_{cb} + z_{be} \quad (10)$$

or, alteratively, eliminating α,

$$Z_c = \frac{(z_{ce} + R_E)(z_{be}z_{cb} + R_s\lambda) + (1 + h_{fe})z_{ce}(R_Ez_{cb} - R_sz_{ce})}{z_{be}(z_{cb} + z_{ce}) + \lambda(R_s + R_E)} . \quad (11)$$

The expressions for Z_c reveal significant complexity, which is compounded by the signal dependence of the small-signal parameter set $\{z_{ce}, z_{cb}, z_{be}, h_{fe}\}$.

To simplify the results, consider a family of approximations for Z_c for specific cases of R_s and R_E, so that the dominant contributors to the output impedance can be determined.

1) *Case 1*: $R_s = 0$, $R_E = 0$.

Eq. (11) reduces to

$$Z_c \simeq \frac{z_{ce}z_{cb}}{z_{ce} + z_{cb}} \quad (12)$$

that is, Z_c is parallel combination of z_{ce} and z_{cb}.

2) *Case 2*: $R_s = 0$, $R_E >> z_{be}/(1 + h_{fe})$.

Eq. (10) approximates to $\lambda = (1 + h_{fe})z_{ce}$ and the denominator of Eq. (11) reveals $\lambda R_E >> z_{be}(z_{cb} + z_{ce})$. Hence,

$$Z_c \approx z_{cb} . \quad (13)$$

This case is typical of the current source and grounded-base amplifier as used in the cascode configuration.

3) *Case 3*: $R_s >> z_{be}$, $R_E = 0$.

From Eq. (11),

$$Z_c \simeq \frac{z_{cb}}{\frac{z_{cb}}{z_{ce}} + \frac{z_{be} + (1 + h_{fe})R_s}{z_{be} + R_s}} \quad (14)$$

where, for $R_s >> z_{be}$, Z_c is z_{ce} in parallel with $z_{cb}/(1 + h_{fe})$ and represents the worst-case output impedance condition.

4) *Case 4*: $R_s >> z_{be}$, $R_E >> z_{be}/(1 + h_{fe})$.

Applying inequalities to Eq. (11), and noting $z_{be} << z_{ce}, z_{cb}$,

$$Z_c \approx \left[\frac{z_{ce}\,z_{cb}}{(1 + h_{fe})z_{ce} + z_{cb}}\right]\left[\frac{R_s + (1 + h_{fe})R_E}{R_s + R_E}\right] + \frac{R_sR_E}{R_s + R_E} . \quad (15)$$

In selecting a circuit topology it should be noted that $z_{cb} > z_{ce}$; thus the grounded-base stage as used in the cascode will offer superior results in terms of output impedance. Nevertheless, z_{cb} is still signal dependent and represents a significant distortion mechanism where large signals are encountered, especially as z_{cb} falls with frequency. Such distortion is demonstrated in Sec. 5.

In Sec. 4 a new form of distortion correction is proposed that reduces output impedance dependence on both z_{ce} and z_{cb} even when nonlinear, and results in lower overall distortion that is virtually frequency independent.

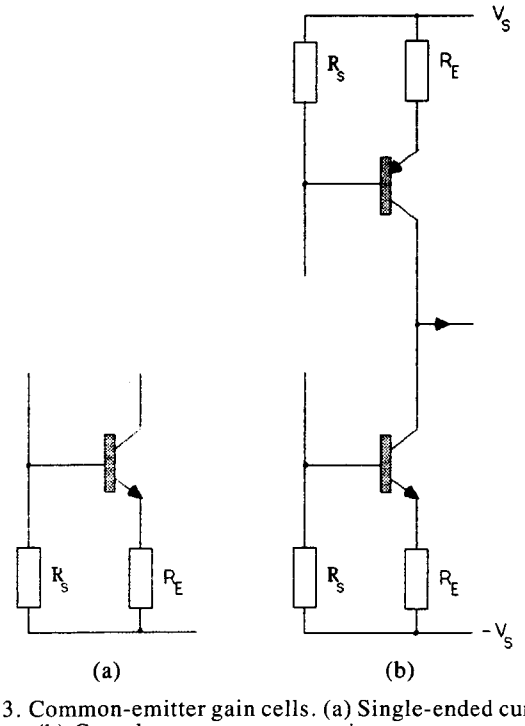


Fig. 3. Common-emitter gain cells. (a) Single-ended current mirror. (b) Complementary current mirror.

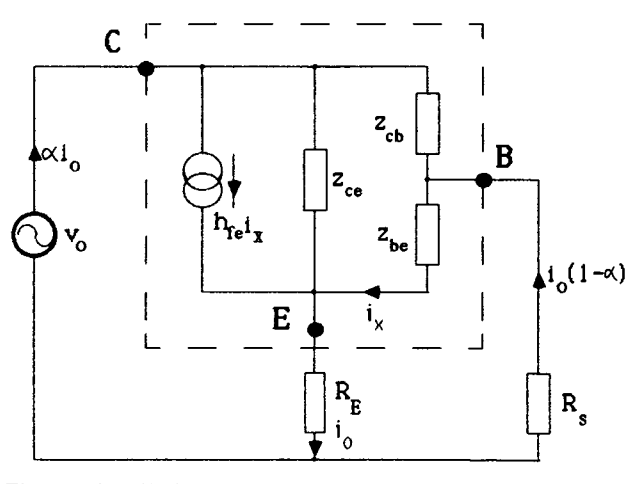


Fig. 4. Small-signal model of common-emitter amplifier showing slope impedances z_{ce} and z_{cb}.

3 REDUCTION OF NONLINEAR SLOPE IMPEDANCE DEPENDENT DISTORTION

The output impedances of the grounded-base and common-emitter amplifier cells are bounded by the device slope impedances z_{cb} and z_{ce}, respectively, as demonstrated by cases 2 and 3 in Sec. 2. However, an examination of Eq. (8) reveals that the factor α in the denominator restricts the output impedance. If a modified circuit topology could be realized such that the base current is summed with the collector current but without incurring an extra load on the collector, then the expression for collector output impedance would become

$$Z_{cu} = \frac{v_o}{\alpha i_o + (1 - \alpha) i_o} = \frac{v_o}{i_o} .$$

Hence from Eqs. (8)–(10) an upper bound on Z_{cu} is established where

$$Z_{cu} = R_E + z_{ce} + \frac{(1 + h_{fe}) z_{ce}(z_{cb} R_E - z_{ce} R_s)}{z_{be}\, z_{cb} + R_s \lambda} . \quad (16)$$

An examination of Eq. (16) reveals that, with typical component values and transistor parameters, a substantial increase in collector impedance is possible and that this is achieved even when z_{ce} and z_{cb} are dynamic. However, this result is an upper bound that assumes that all the base current is returned to the collector. In practical topologies this is compromised by a small margin, so that lower values should be anticipated.

Two circuit approaches have been identified to meet the requirement of base and collector current summation without direct connection to the collector. These are based on a local feedforward and feedback strategy, respectively, and can be used independently or compounded to give further enhancement.

3.1 Feedforward Topology

The feedforward topology is a derivative of the Darlington transistor that is occasionally employed in power amplifier current mirrors [6], [7]. In Fig. 5 two circuit examples are presented which yield similar performance. In each circuit the base current of the output device is returned to the emitter via the emitter–collector of the driver stage. Consequently the advantages of the Darlington are retained, yet with an enhanced output impedance realized by removing the respective currents in z_{ce} and z_{cb} from the output branch of the complementary stage. It should be noted that the collector–emitter voltage variation of the drivers is small, with only the output collectors swinging the full range of output voltage. The conventional Darlington connection of parallel collectors compromises this ideal, with the driver stage adding a degree of slope distortion under large-signal excitation. It is, however, important to note that a small fraction of output transistor base current is not returned to the emitter and is dependent on the ratio of R_E to transistor output impedance as seen at the emitter of the output device. This fractional loss of current will lower the bound suggested by Eq. (16), although there is still substantial advantage.

3.2 Feedback Topology

The conventional cascode as illustrated in Fig. 6(a) offers an output impedance approaching z_{cb}, which is a significant improvement over the common-emitter stage as $z_{cb} > z_{ce}$. A simple modification to the basic circuit can return the base current of the grounded-base stage to the emitter of the common-emitter stage. Consequently signal current flowing in both z_{ce} and z_{cb} now form local loops which do not include the output branch. The new topology is shown in Fig. 6(b), while in Fig. 6(c) the basic current paths are illustrated which apply even when z_{ce} and z_{cb} are nonlinear. Again, it is only the output device whose collector is required to swing over the full output voltage; thus the common-emitter stage offers a minimal slope distortion contribution.

In circuit applications where the common-emitter stages operate at a high bias current to improve I_E/V_{BE} linearity, a bypass current I_x [see Fig. 6(b)] can lower the operating current of the common-base stage. This technique both reduces output device power dissipation and aids a further increase in the slope impedances, while circuit symmetry ensures that noise in I_x does not flow in the output branch. As a practical detail, experimentation has revealed the desirability of ac bypassing of the base bias resistance of the grounded-base stages [see capacitors C in Fig. 6(b)]. This both enhances circuit operation and eliminates any tendency toward high-frequency oscillation due to the positive-feedback loop formed by the base–emitter connections.

3.3 Compound Feedback/Feedforward Topologies for z_{ce}, z_{cb} Reduction

The methods based on feedforward and feedback addition of the output device base current can be compounded to offer further performance advantage. There are many possible topologies offering minor variations, though each uses the same basic concept. It is not intended to analyze each variant, though a family of topologies is presented in Fig. 7 to stimulate development.

4 NOISE CONTRIBUTION OF GROUNDED-BASE STAGE WITH BASE CURRENT SUMMATION

In this section brief consideration is given to the contribution of noise from the common-base stage in the cascode for the two basic topologies shown in Fig. 8.

In both cases let $\overline{i_{cn}^2}$ be the mean square noise current in the collector of the common-emitter stage and let the common-base stage have respective noise voltage and noise current sources $\overline{e_n^2}$ and $\overline{i_n^2}$.

It is clear that because the common-emitter stage offers a relatively high output impedance at the collector, the equivalent voltage noise generator of the common-base stage yields a negligible contribution to the output

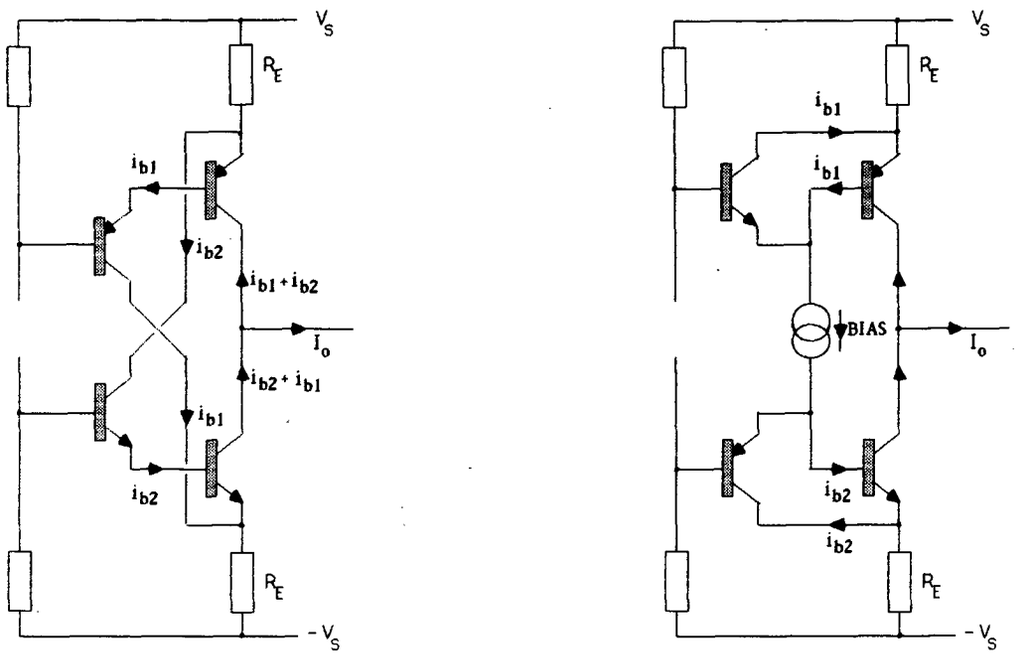


Fig. 5. Two examples of feedforward addition of output stage base currents using a two-stage topology. (Observe base current paths i_{b1} and i_{b2}.)

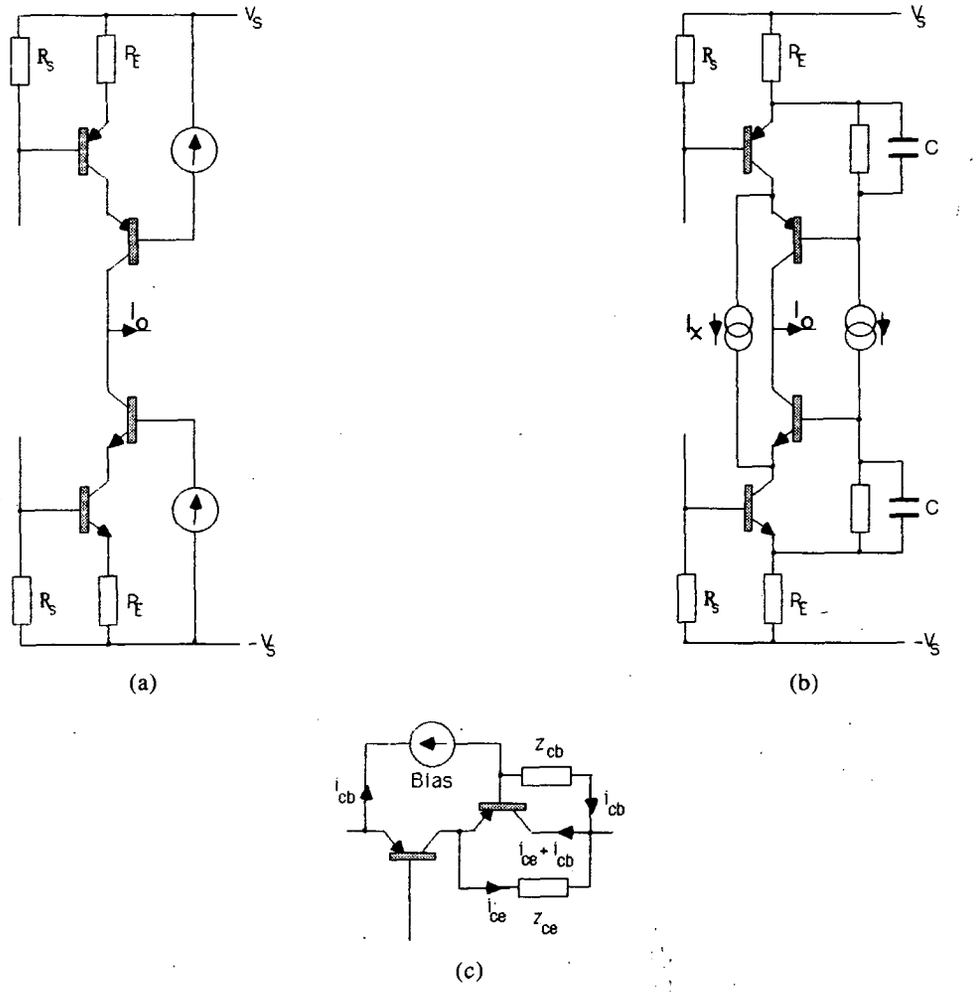


Fig. 6. Slope distortion reduction using feedback topology. (a) Conventional cascode. (b) Enhanced cascode. (c) Illustration of signal current paths i_{ce}, i_{cb} in z_{ce}, z_{cb}.

noise current.

However, an inspection of the noise current paths reveals that in Fig. 8(a) almost all $\overline{i_n^2}$ must flow in the collector, hence effective load, while in Fig. 8(b) virtually all the noise current circulates locally through the common-emitter stage, resulting in only a fraction, $\approx \overline{i_n^2}/[1 + 1/h_{fe} + h_{fe}R_E/(R_s + R_E + z_{be})]^2$, appearing in the collector (assuming similar transistor h_{fe}'s). Consequently with the enhanced topology there is virtually no extra noise generated by the addition of the common-base stage. Hence the output noise current is also $\overline{i_{cn}^2}$.

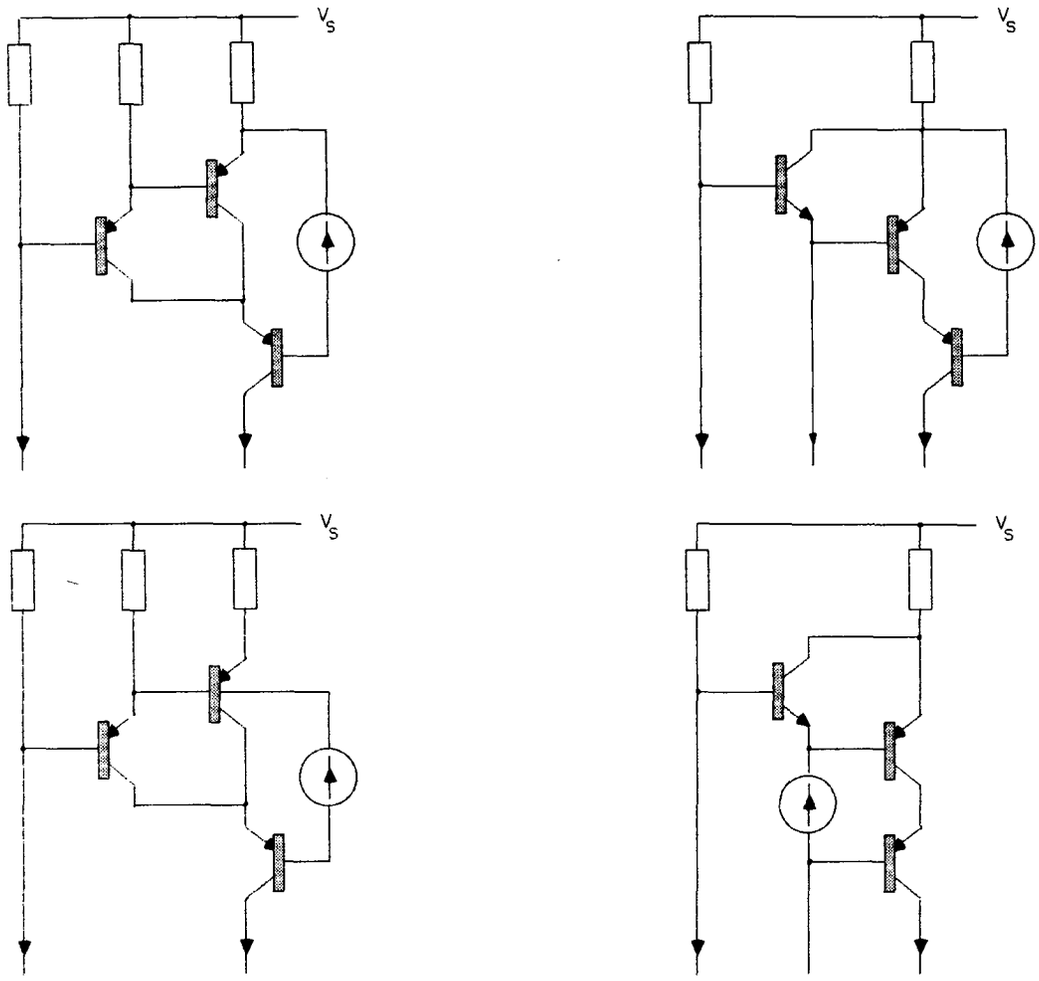


Fig. 7. Circuit examples using two-stage common-emitter amplifier with a common-base output stage.

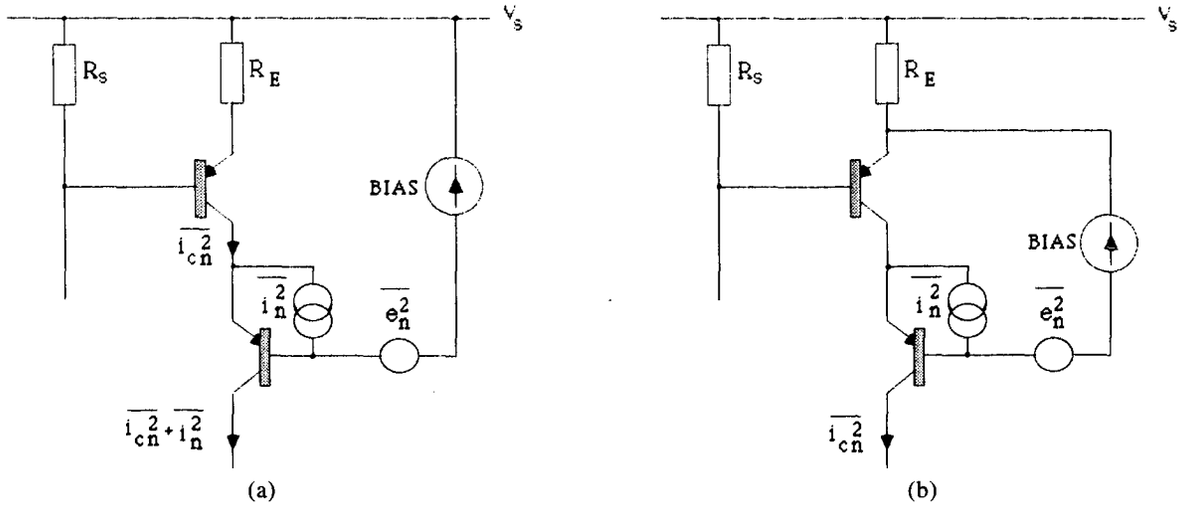


Fig. 8. Noise sources of common-base stage. (a) Conventional cascode. (b) Enhanced cascode.

5 MEASURED PERFORMANCE ADVANTAGE OF ENHANCED TOPOLOGY

To highlight the performance advantage of the modified common-base stage and to demonstrate the significance of slope distortion at large signal levels, a test circuit was constructed to validate the technique and to permit an objective assessment.

Three variants of the circuit were constructed and tested with ascending levels of modification. The enhanced topology is shown in Fig. 9(c), with the comparative output stage variants highlighted in Fig. 9(a) and (b). The circuit is dc coupled and no overall feedback is used. The output voltage is derived using a 10-kΩ gain-defining resistor R_g, and an offset-null potentiometer is provided since no servo amplifier is used. The total harmonic distortion results are given in Table 1. All measurements were performed with a sinusoidal input and an output voltage of 80 V peak to peak.

The results show that the basic circuit exhibits a distortion rising with frequency, reaching an unacceptable 1.9% at 50 kHz. This result is a function of the voltage-dependent nature of the device capacitance and represents a severe dynamic distortion. The conventional cascode exhibits a marked improvement, which reflects the popularity of this topology, where distortions are consistently reduced by 20 dB compared with the no-cascode circuit. However, although distortion products are of a lower order, they are still frequency dependent. This difference in performance arises from the basic common-emitter stage having an output impedance $\approx z_{ce}$, while the common base stage is z_{cb}, where $z_{cb} > z_{ce}$, though they follow the same basic frequency dependence, hence the tracking of the distortion figures.

However, the enhanced cascode, where performance is almost independent of both z_{ce} and z_{cb}, shows a distortion reduction greater than 40 dB at 50 kHz with a very desirable 31.8-dB improvement at 1 kHz over the basic circuit. Of particular significance is the almost frequency-independent nature of the distortion, together with the indication that the two stages of amplification are of inherent low distortion, though clearly they are a limit to linearity for the enhanced circuit. This performance level was masked by slope distortions in the conventional circuit.

These tests are sufficient to validate the technique, especially as the cost overhead is minimal compared with the conventional cascode, and represent a substantial performance enhancement irrespective of whether overall feedback is contemplated in a final design.

6 CONCLUSION

This paper has presented a method of reducing the performance dependence on transistor collector–emitter and collector–base slope impedance parameters, whereby useful distortion reduction can be achieved for large-signal voltage amplifiers.

A theory was presented to demonstrate that for a given input cell transconductance and closed-loop gain, the error signal due to the modulation of output impedance Z_n was not dependent on the level of feedback, provided g_m and target gain γ remained constant. Consequently for the test circuits of Section 5, if overall feedback was applied together with an appropriate increase in the gain-defining resistor R_g, the same level of distortion due to modulation of Z_n should be anticipated. (Note that a unity-gain buffer amplifier would be required.) However, if R_g is raised, the signal current level operating in the transconductance gain stage will fall, resulting in a reduced distortion from modulation in g_m. This later distortion would be particularly evident with the enhanced cascode, where modulation of g_m is now the limiting distortion mechanism.

The enhanced topology has specific application in large-signal voltage amplifiers and, with appropriate circuit additions, to power amplifiers. In particular, MOSFET power amplifiers can benefit by using a more optimum current source to drive the output stage since this reduces dependence on both gate-to-source voltage errors as well as slope impedance modulation errors [8].

A third area of application is RIAA disk preamplifiers that use a transconductance cell and a passive equalization-defining impedance [9], [10]. The more optimum current source will lower distortion and increase EQ accuracy as the current source exhibits a lower output capacitance, together with a higher output resistance, the latter particularly affecting low-frequency performance.

It is interesting to observe that if negative feedback alone were used to reduce error dependence on Z_n by the same factor as the enhanced cascode, at 1 kHz an increase in loop gain of more than 30 dB is required, or at 50 kHz this requirement rises to more than 40 dB. Such factors are often impractical to achieve, thus vindicating the adoption of the enhanced topology. However, more fundamentally, the distortion dependence on transistor slope impedance inevitably rises with both frequency and output voltage level, and moves against the loop gain requirement for stability, thus making negative feedback less effectual in suppressing slope-dependent nonlinearity.

The techniques described in this paper should also find application in circuits that require enhanced supply rail rejection. An appendix outlines how slope impedance distortion reduction can improve the performance of voltage/power amplifiers by enhancing the interface between amplifier stages which alternate their signal reference between ground and supply rail.

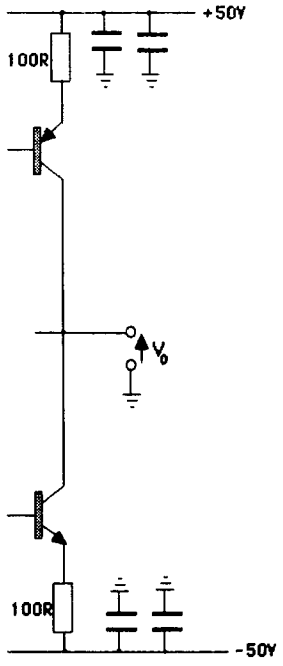
Although the reduction of large-signal-related errors arising from slope distortion has been the central thesis, the reduction of linear distortion at lower signal levels is also welcome. Slope distortion has been shown to involve several factors that depend on both transistors and the associated circuit elements in a particular application. Such device-specific distortion can, in principle, contribute to the subjective performance and

reflects the mutual interrelationship of transistors and circuit construction, which results in small deviations from the target transfer function.

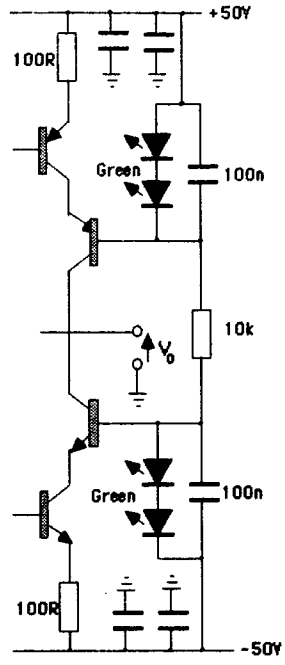
The paper has presented a family of primitive circuit topologies based on the same principle as the enhanced cascode, which are candidates for adoption in transconductance-based amplifiers. There are numerous circuit possibilities for enhancement. However, the two

Table 1. Total harmonic distortion.

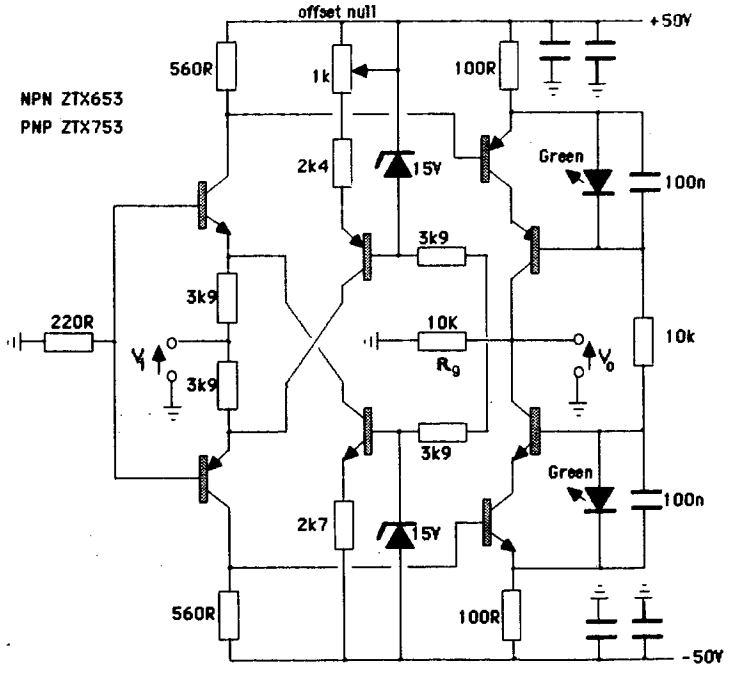
Test frequency, kHz	No cascode, %	Conventional cascode, %	Enhanced cascode, %
1	0.39	0.039	0.010
10	0.47	0.11	0.011
20	0.51	0.14	0.012
50	1.9	0.16	0.016



(a)



(b)



(c)

Fig. 9. Test circuit with three output stage variants. (a) Complementary common-emitter output stage. (b) Complementary cascode output stage. (c) Complete test circuit with enhanced cascode.

basic principles to be observed are

1) Adequately high effective emitter resistance R_E to disassociate z_{ce} from the output impedance at the collector;

2) Addition of base current to collector current, without adding extra circuitry to collector, to disassociate z_{cb} from the output impedance at the collector.

Observation of these two principles then enables a transformation of the signal level from low voltage to large voltage without incurring a significant distortion penalty due to dynamic modulation of the transistor slope parameters, together with a distortion characteristic that is considerably less frequency dependent.

7 ACKNOWLEDGMENT

The author wishes to gratefully acknowledge the assistance of Paul Mills from the Department of Electronic Systems Engineering for his support in constructing and compiling the measured data on the three circuit derivatives.

8 REFERENCES

[1] M. J. Hawksford, "Distortion Correction Circuits for Audio Amplifiers," *J. Audio Eng. Soc.*, vol. 29, pp. 503–510 (1981 July/Aug.).

[2] E. M. Cherry and G. K. Cambrell, "Output Resistance and Intermodulation Distortion of Feedback Amplifiers," *J. Audio Eng. Soc.*, vol. 30, pp. 178–198 (1982 Apr.).

[3] M. J. Hawksford, "Power Amplifier Output-Stage Design Incorporating Error-Feedback Correction with Current-Dumping Enhancement," presented at the 74th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 960 (1983 Dec.), preprint 1995.

[4] M. J. Hawksford, "The Essex Echo: Reflexions," *HFN/RR*, vol. 30, pp. 35–40 (1985 Dec.).

[5] K. Lang, "The Lang 20W Class-A MOSFET Amplifier," *Audio Amateur*, vol. 2, pp. 7–12 (1986).

[6] E. M. Cherry, "Feedback, Sensitivity, and Stability of Audio Power Amplifiers," *J. Audio Eng. Soc.*, vol. 30, pp. 282–294 (1982 May).

[7] P. J. Walker and M. P. Albinson, "Current Dumping Audio Amplifier," presented at the 50th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 23, p. 409 (1975 June).

[8] R. R. Cordell, "A MOSFET Power Amplifier with Error Correction," *J. Audio Eng. Soc.*, vol. 32, pp. 2–17 (1984 Jan./Feb.).

[9] Y. Miloslavskij, "Audio Preamplifier with no TID," *Wireless World*, vol. 85, no. 1524, pp. 58–60 (1979 Aug.).

[10] O. Jones, " . . . about the genesis of the Pip," *HFN/RR* (Letter to the Editor), vol. 30, no. 12, p. 25 (1985 Dec.).

APPENDIX
SUPPLY RAIL REJECTION AS A FUNCTION OF INPUT STAGE AND CURRENT MIRROR SLOPE IMPEDANCES

In this appendix the sensitivity of a two-stage negative-feedback amplifier is determined as a function of the slope impedances Z_{n1} and Z_{n2} of the two stages. The basic circuit is shown in Fig. 10 where g_m is the transconductance of the input stage, m the current gain of the current mirror, R_g a gain-defining resistor, r_2 the input impedance of the current mirror ($r_2 << Z_{n1}$), and k the feedback factor.

Using linear analysis to express V_o as a function of both V_{in} and V_s,

$$V_o = \frac{mg_mR_gV_{in} + R_g[m/Z_{n1} + 1/Z_{n2} + r_2/Z_{n1}Z_{n2}]V_s}{(1 + r_2/Z_{n1})(1 + R_g/Z_{n2}) + kmg_mR_g}. \quad (17)$$

Let δ be the ratio of output to input transfer functions for inputs V_s and V_{in},

$$\delta = \frac{V_o/V_s}{V_o/V_{in}} \quad (18)$$

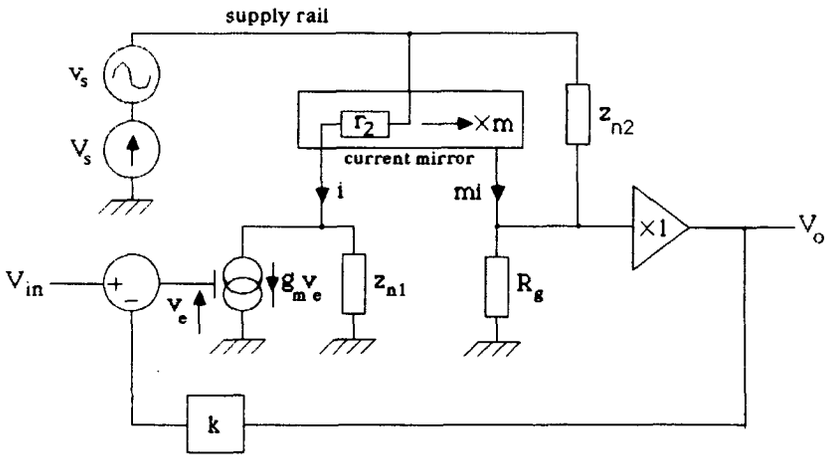


Fig. 10. Two-stage voltage amplifier with v_s representing power supply voltage variation.

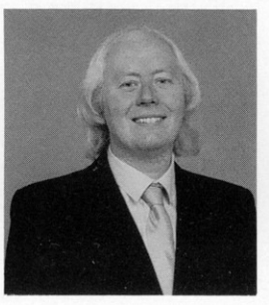
and

$$\delta = \left[\frac{1}{Z_{n1}} + \frac{1}{mZ_{n2}}\left(1 + \frac{r_2}{Z_{n1}}\right)\right]\frac{1}{g_m}. \quad (19)$$

The results show that the slope impedances define the suppression of supply rail rejection together with g_m. This is particularly important in power amplifier applications, where in class AB operation V_s is wide band (>>20 kHz) and a nonlinear function of the input signal due to output stage commutation. The advantages of maximizing both Z_{n1} and Z_{n2} and using separate power supplies for voltage amplifier and output stage in power amplifiers are evident.

Eq. (19) is also shown to be independent of R_g. However, in high loop gain applications where g_m is large, the high-frequency distortion characteristics together with the falling high-frequency gain of g_m may become a limiting factor, particularly if required to suppress wide-band power supply injection. In low-feedback applications, the slope impedance dependent distortion is suppressed more by the presence of R_g than by the presence of g_m. For example, observe how R_g and Z_{n2} form a potential divider to supply injected distortion, but as $R_g \to \infty$, the distortion is processed completely by the feedback loop. Also in low-feedback designs greater local feedback enhances the wide-band distortion characteristics of g_m and helps aid an overall distortion profile which is less frequency dependent.

THE AUTHOR



Malcolm Hawksford is a senior lecturer in the Department of Electronic Systems Engineering at the University of Essex, U.K., where his principal interests are in the fields of electronic circuit design and audio engineering. Dr. Hawksford studied at the University of Aston in Birmingham and gained both a First Class Honors B.Sc. and Ph.D. The Ph.D. program was supported by a BBC Research Scholarship where work on the application of deltamodulation to color television was undertaken.

Since his appointment at Essex, he has established the Audio Research Group, where research on amplifier studies, digital signal processing and loudspeaker systems has been undertaken. Dr. Hawksford has had several AES publications that include topics on error correction in amplifiers and oversampling techniques for ADC and DAC systems. His supplementary activities include designing commercial audio equipment and writing articles for Hi-Fi News—activities that integrate well with visits to Morocco and France. His leisure activities include listening to music, motorcycling and motor mechanics. Dr. Hawksford is a member of the IEE, a Chartered Engineer, Fellow of the AES, and a member of the Review Board of the AES *Journal*.

Distortion Reduction in Moving-Coil Loudspeaker Systems Using Current-Drive Technology*

P. G. L. MILLS** AND M. O. J. HAWKSFORD

University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK

The performance advantages of current-driving moving-coil loudspeakers is considered, thus avoiding thermal errors caused by voice-coil heating, nonlinear electromagnetic damping due to $(Bl)^2$ variations, and high-frequency distortion from coil inductive effects, together with reduced interconnect errors. In exploring methods for maintaining system damping, motional feedback is seen as optimal for low-frequency applications, while other methods are considered. The case for current drive is backed by nonlinear computer simulations, measurements, and theoretical discussion. In addition, novel power amplifier topologies for current drive are discussed, along with methods of drive-unit thermal protection.

0 INTRODUCTION

The moving-coil drive unit is by far the most widely used electroacoustic transducer in both high-performance studio and domestic audio installations, as well as in general-purpose sound reinforcement. Consequently it has attracted numerous studies to investigate its inherent distortion mechanisms (see, for example, [1]–[11]), which as a consequence are well understood. Much work has also been carried out on improving drive-unit linearity by the application of motional feedback techniques, which provide a useful enhancement in performance at low frequencies. Improvements to the basic regime of motional feedback have been made by including an additional current feedback loop [12], [13], which is reported to reduce high-frequency distortion. This method is a specific implementation of what we will term *current drive*, a subject that, it is felt, has not received the attention it deserves.

This paper therefore aims to explore in detail the benefits of current drive in reducing the dependence of drive-unit performance on motor system nonlinearities, in particular the voice-coil resistance which undergoes significant thermal modulation.

In a conventional voltage-driven system (one where the power amplifier output voltage is regarded as the information-representing quantity), the current is initially limited by the series elements of voice-coil resistance and inductance, together with the interconnect and amplifier output impedance. A force related to the current in the system then acts on the drive unit moving elements as a result of the motor principle, and once motion occurs, an electromotive force is induced in the coil to oppose the applied signal voltage, thus constraining the magnitude of current flow. The accuracy to which the drive-unit velocity responds to the applied signal is, therefore, dependent on the series elements in the circuit, and any signal-related changes in their value will result in distortion.

The voice-coil resistance is of specific concern, as it is usually a dominant element. As a result of self-heating in excess of 200°C, a significant increase in coil resistance occurs of typically 0.4%/°C for copper, leading to sensitivity loss, lack of damping, and crossover misalignment. In their paper, Hsu et al. [6] concluded that a satisfactory method of compensating for this effect had yet to be found.

At higher audio frequencies, the coil inductance also becomes significant, resulting in a loss of sensitivity. In addition, the inductance suffers dynamic changes with displacement, providing a distortion mechanism which is further complicated by eddy current coupling to the pole pieces in the magnetic circuit [14, pt. 1]. A further problem is distortion mechanisms at the amplifier–loudspaker interface, such as interconnect errors [14, pt. 4] and interface intermodulation distortion

* Manuscript received 1988 February 17.

** Now at Tannoy Ltd., Rosehall Industrial Estate, Coatbridge, Strathclyde, ML5 4TF, UK.

[15]–[17].

To overcome these limitations, the drive unit should be current rather than voltage controlled and interfaced directly to a power amplifier configured as a current source, thus offering a high output impedance. The performance advantages of this technique are discussed in detail, supported by computer simulation of the nonlinear system together with objective measurement on a prototype two-way active loudspeaker system.

To complement this study, the application of both motional feedback and noninteractive frequency response shaping as a means of aligning the drive unit Q to the required value is discussed. Finally, the topic of current source power amplifier design is considered along with the presentation of some novel types of circuit topology, while the subject of drive-unit protection under current drive is also examined.

The technique of current drive in active loudspeaker systems is seen as being of particular importance in view of the performance advantages demonstrated over conventional systems in terms of both reduced linear and nonlinear distortion. For high-quality system design, current drive is seen as the more logical methodology, with voltage drive appearing as the result of established practice and convenience.

1 LOUDSPEAKER PERFORMANCE UNDER VOLTAGE DRIVE

In order to establish a performance reference, this section considers motor system linearity for a moving-coil drive unit under conventional voltage drive. For the tests, a Celestion SL600 135-mm-diameter bass–midrange driver was used, mounted in its enclosure. It was chosen partly due to the excellent cone and surround behavior, meaning that the distortion contribution of these elements is small.

To enable performance predictions under general signal excitation to be made, the variation of parameters with coil displacement was measured. This is shown in graphic form, in Fig. 1 for Bl product, compliance, and coil inductance. The linear parameters for the model are given in Table 1, which explains the terminology and also the equivalence between the electrical model and the mechanical model used. The approach broadly

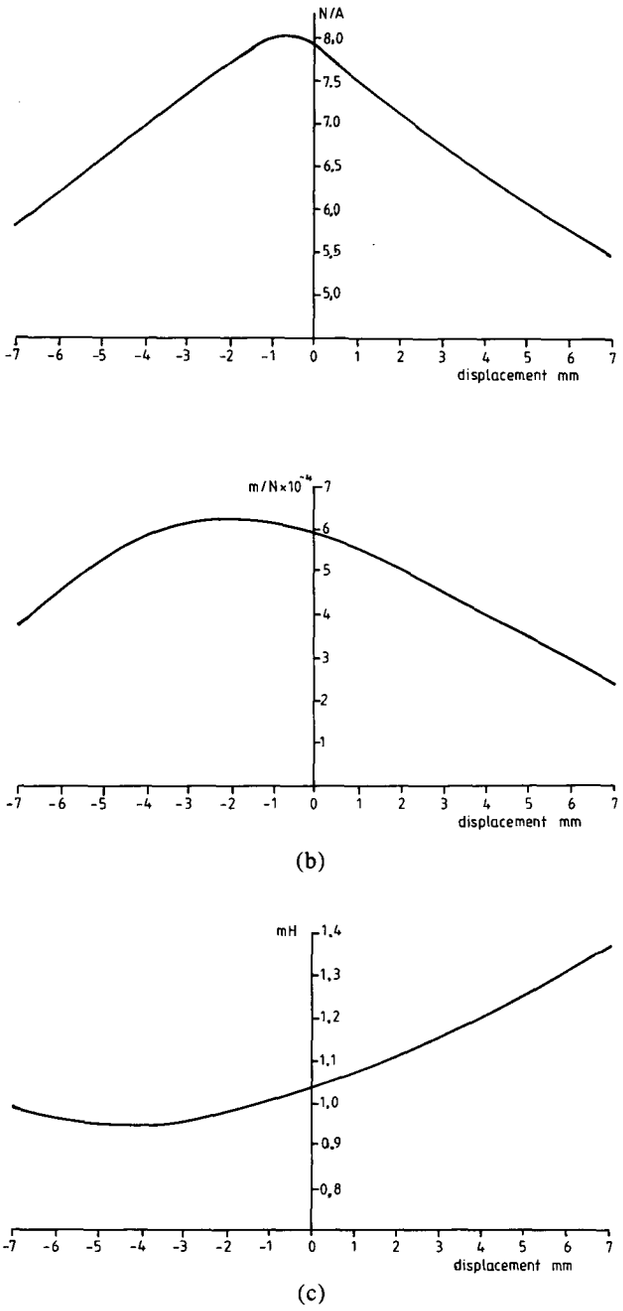


Fig. 1. Variation of model parameters with displacement. Negative displacement indicates motion toward magnet. (a) Bl product. (b) Mechanical compliance. (c) Electric coil inductance.

Table 1. Model parameters for example drive unit.

Parameter	Electrical model	Mechanical model
Voice-coil resistance	$R_e = 7.0\ \Omega$	$R_{me} = (Bl)^2/R_e$ kg/s
Voice-coil inductance	L_e*	$C_{me} = L_e/(Bl)^2$ m/N
Enclosure compliance	$L_{cmb} = C_{mb}(Bl)^2$	$C_{mb} = 750 \times 10^{-6}$ m/N
Suspension compliance	$L_{cms} = C_{ms}(Bl)^2$	C_{ms}* m/N
Moving mass	$C_{mes} = M_{ms}/(Bl)^2$	$M_{ms} = 0.0183$ kg
Mechanical resistive losses	$R_{es} = (Bl)^2/R_{ms}$	$R_{ms} = 2.4336$ kg/s
Source impedance	Z_g (assume zero)	$Z_{mg} = (Bl)^2/Z_g$ kg/s

Bl = force factor (N/A)*

* Indicates nonlinear elements.

follows that of Small [18], except that a mechanical model is used in preference to the acoustic one. Fig. 2(a) shows the equivalent electrical model for the drive unit, connected to an amplifier and interconnect of series source impedance Z_g, showing the mechanical impedance as a lumped quantity Z_m. Analysis of this model gives the transfer function between amplifier output voltage and cone velocity,

$$u = \frac{V_0 Bl}{Z_m [Z_s + (Bl)^2/Z_m]} \quad (1)$$

where

u = cone velocity, meters per second
V_0 = amplifier source voltage, volts
B = flux density for motor system, tesla
l = coil length in field B, meters
Z_m = lumped mechanical impedance, kilograms per second
Z_s = lumped electric impedance (Z_g, R_e, and sL_e), ohms.

Referring the mechanical impedance to the "primary" of the Bl transformer to show its constituents gives the electrical model of Fig. 2(b), while referring the electrical parameters to the "secondary" results in the mechanical model of Fig. 2(c). Both these models are useful in the forthcoming discussion, although emphasis is placed on the mechanical system.

The mechanical model forms the basis of a transient analysis procedure, which can readily incorporate nonlinear parametric variations. The details of this approach

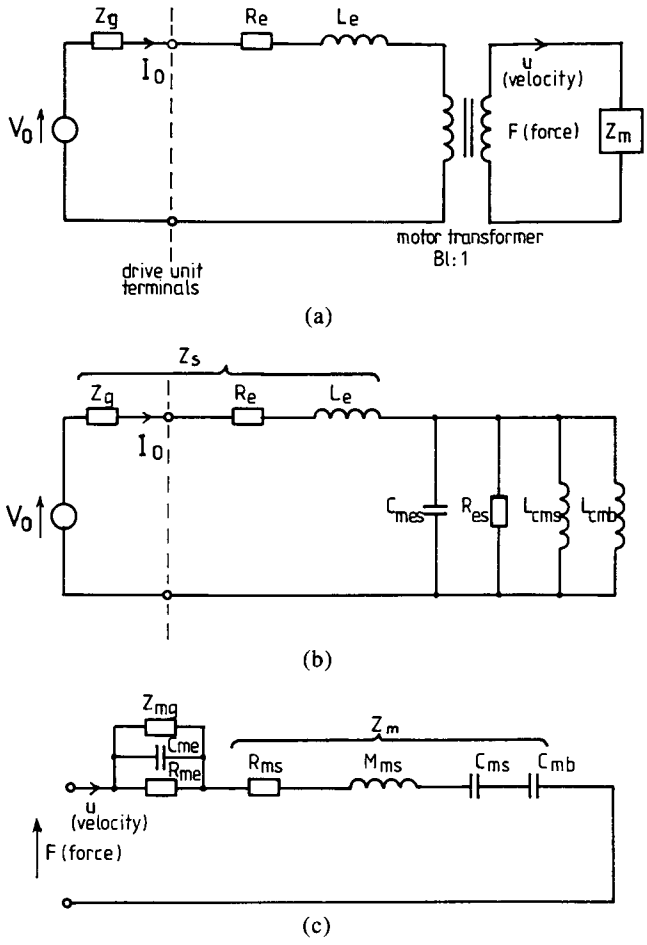


Fig. 2. Modeling of drive unit in sealed enclosure, under voltage drive. (a) Basic electromechanical model. (b) Electrical model. (c) Mechanical model.

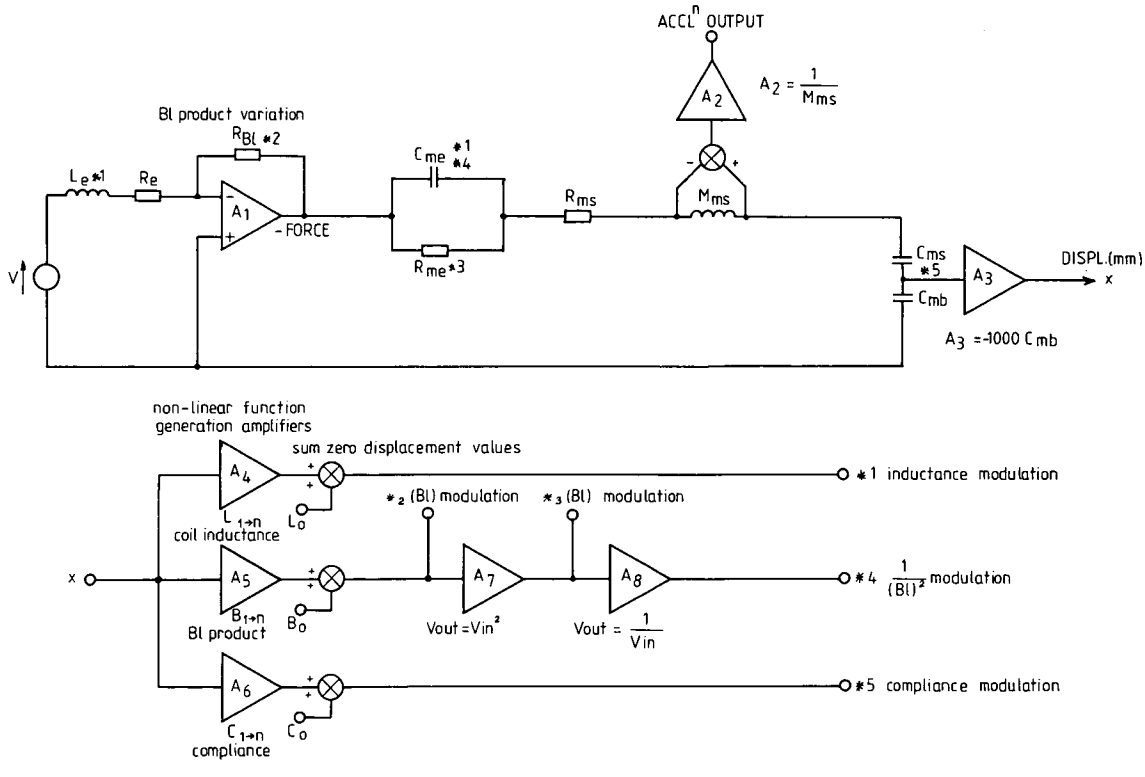


Fig. 3. Simplified nonlinear model for voltage-driven simulation.

were given in an earlier paper [19], where it was seen to avoid the approximations forced by an analytical solution. It may be contrasted to that of Kaizer [3], where drive-unit nonlinearity was modeled by Volterra series expansion. Fig. 3 shows the simplified circuit for the computer model, where it is assumed that the driving amplifier has zero source impedance. The output quantity acceleration is derived from the voltage across the moving mass element M_{ms}, while a signal proportional to displacement is obtained from the voltage across the enclosure compliance C_{mb}. This displacement voltage is used to drive three nonlinear amplifiers A_4, A_5, and A_6, whose transfer characteristics are expressed as polynomials, representing the measured variation in coil inductance, Bl product, and compliance, respectively. A technique was adopted whereby the three nonlinear functions were first represented by a Fourier series from which the corresponding polynomials were generated. A 30th-order approximation to each function was deemed necessary to avoid undue error. The zero displacement values for these parameters were then summed by constant factors L_0, B_0, and C_0. Each of the modulation outputs in the diagram is coded by an asterisk and number to indicate which circuit parameters it modulates.

To produce distortion predictions from this model, a sine wave input is used and the system allowed to reach steady state. A single cycle is then sampled as input data to a fast Fourier transform, which indicates the relative amplitude of the distortion harmonics.

At 100 Hz, with the source voltage chosen to give a current of 1 A peak, a reasonable relation between theoretical and measured distortion spectra can be seen in Fig. 4. The model slightly overestimates most distortion components, probably due to errors in measuring the nonlinear parameters. However, at high frequency the model does not prove usable, due to factors such as the complicated nature of eddy current losses and hysteresis effects in the magnetic circuit. The measured 3-kHz at 1A peak distortion spectra are shown in Fig. 5, while intermodulation products between 50-Hz and 1-kHz sine wave inputs of equal amplitude are shown in Fig. 6.

The effect of voice-coil heating is a major problem under voltage drive, and it is interesting to note the severe difficulty in obtaining these measurements due to the sensitivity loss and frequency response errors which occur as the coil heats up. A further problem caused by heating is that of crossover misalignment in the case of passive systems. To illustrate this effect, Fig. 7 shows an idealized two-way second-order crossover aligned to 3.4 kHz. The drive units are represented by resistive elements, and the overall system transfer function is evaluated by effectively subtracting the high-pass and low-pass outputs. Fig. 8 then compares the system transfer function arising from coil heating to 200°C with the intended response at 20°C. Although oversimplistic, this model does show that large errors can result.

Errors due to interconnect effects are also seen to be of importance. Measurements of the error across a selection of 5-m interconnects have revealed errors up to 15 dB below the main signal. It is worth noting that the error is a function of drive-unit–crossover impedance and, while mainly linear, also contains a nonlinear component due to the nonlinear nature of the load.

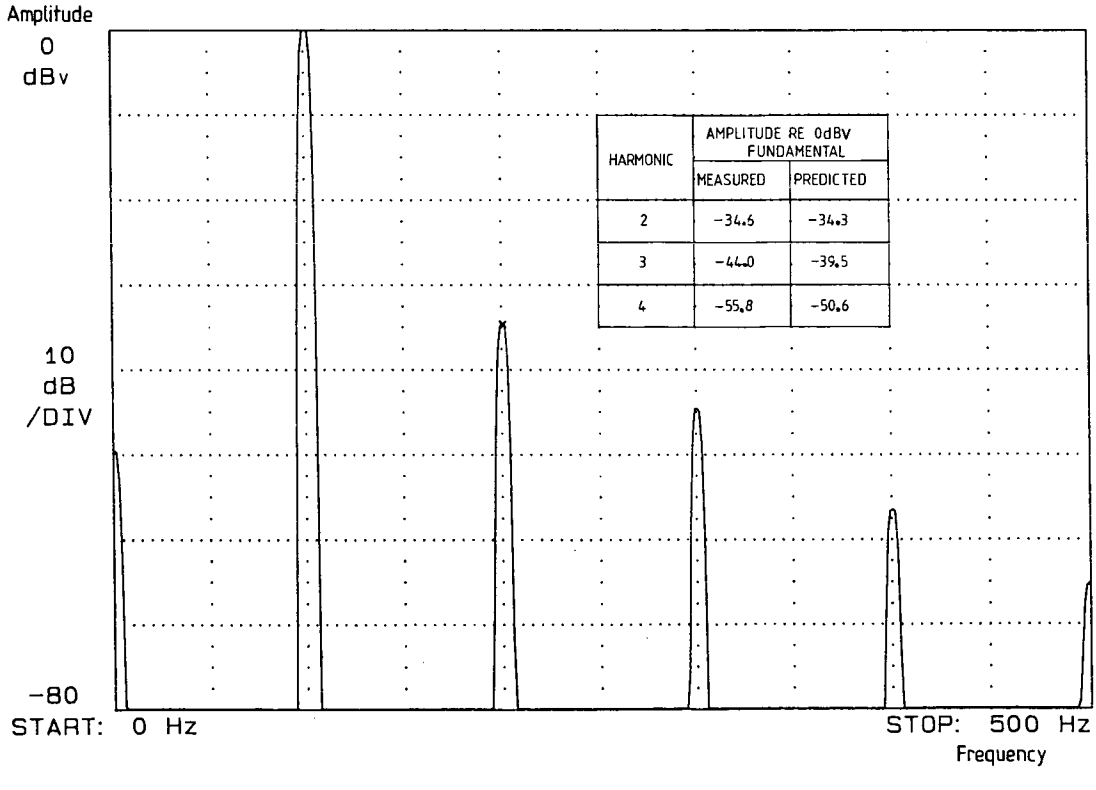


Fig. 4. Measured 100-Hz harmonic distortion, voltage driven. Table compares with predicted result.

2 THE CASE FOR CURRENT DRIVE

To assess the performance advantages of a current-driven moving-coil drive unit, a similar procedure is adopted to that of Sec. 1, though a current source is substituted for the voltage source, with output impedance assumed infinite. An immediate consequence of this strategy is that the series elements of coil resistance, coil inductance (with attendant eddy current losses), and interconnect lumped series elements, together with *Bl* and the lumped mechanical impedance no longer influence the instantaneous driving current. The significance of this observation is best illustrated by examining the current-driven velocity transfer function,

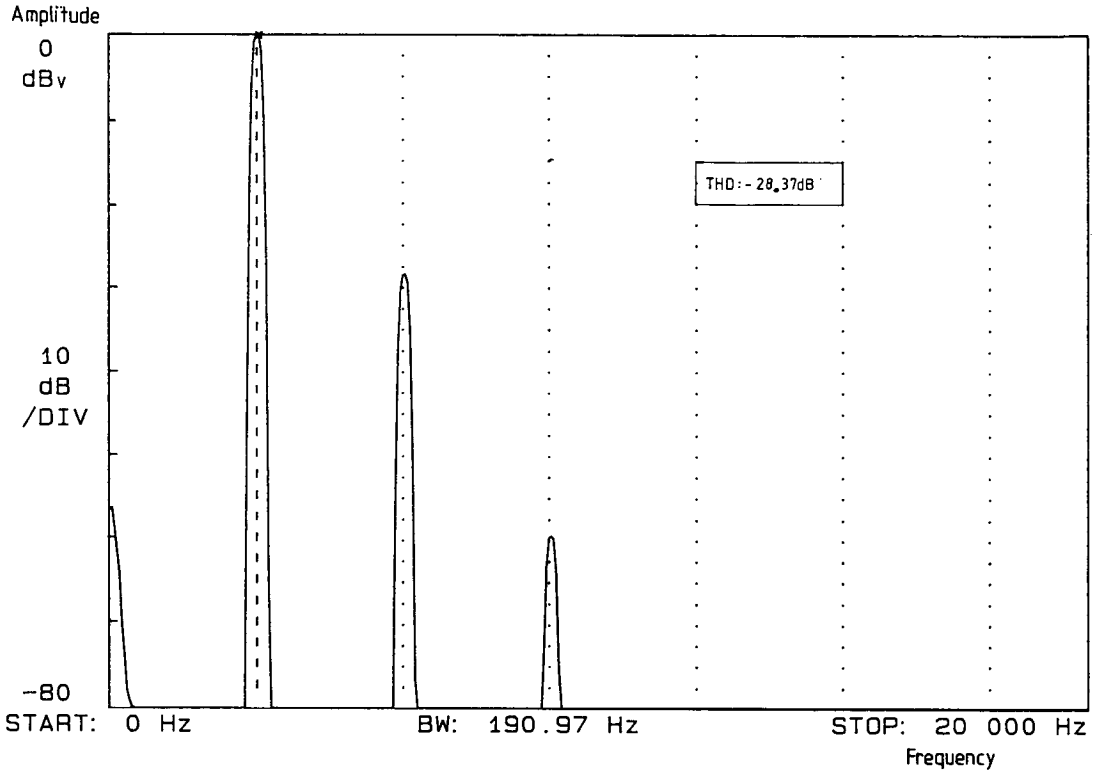


Fig. 5. Measured 3-kHz harmonic distortion, voltage driven.

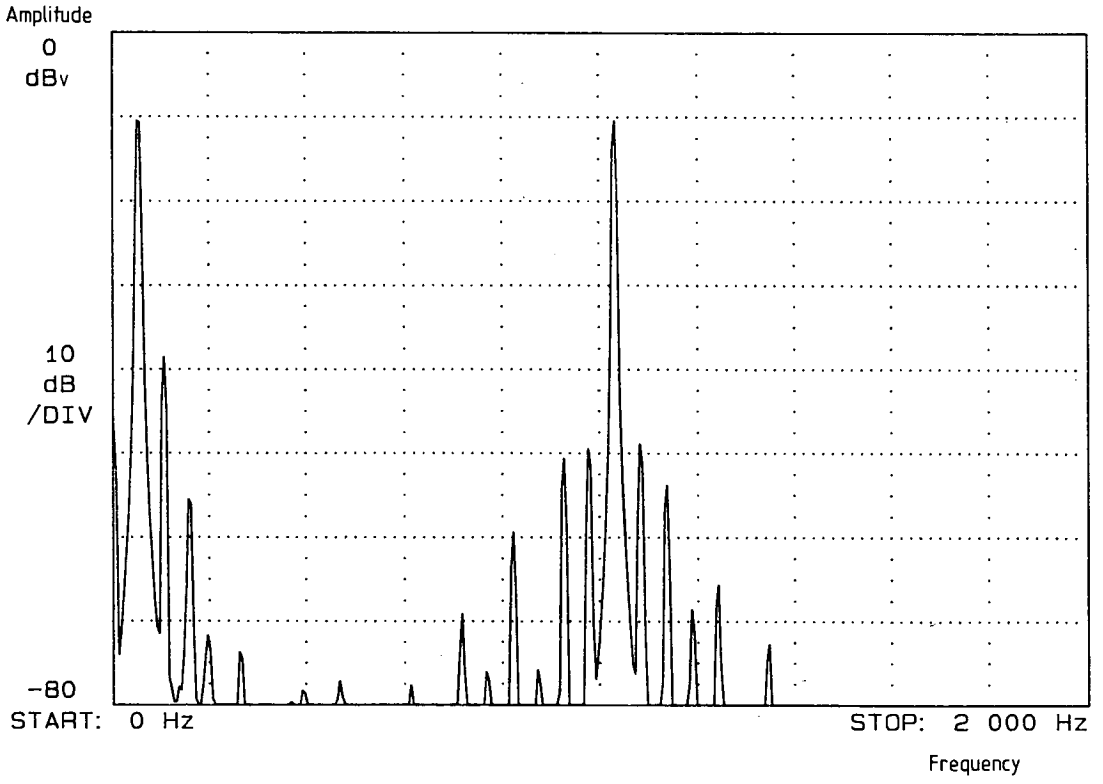


Fig. 6. Measured 50-Hz and 1-kHz intermodulation distortion, voltage driven.

$$u = \frac{I_0 Bl}{Z_m} \quad (2)$$

where I_0 is the amplifier output current in amperes.

Comparison with the voltage-driven case, Eq. (1), shows that for current drive, the transfer function is of a simpler form, independent of the terms Z_s and $(Bl)^2$. It is therefore anticipated that lower distortion will result from elimination of the term $[Z_s + (Bl)^2/Z_m]$. Performance is therefore free of any linear and nonlinear contributions from Z_s, the $(Bl)^2$ term, and shows a reduced dependence on compliance nonlinearity within Z_m, together with any frequency-dependent nonlinear interactions. The mechanical model for the drive unit and enclosure is then reduced to that of Fig. 9.

To demonstrate this claim, a transient analysis at 100 Hz with 1-A peak drive current was performed, using the nonlinear model of Fig. 10. A reasonable match between measured and predicted distortion is again obtained, as shown by Fig. 11.

Comparing this result with the voltage-driven case (Fig. 4) shows a measured and predicted distortion reduction of around 9 dB for the second harmonic, with third- and fourth-order products being reduced by between 3 and 7 dB, depending on whether the measured or the predicted values are taken (the predicted results yielding the better distortion reduction).

Regarding the 3-kHz at 1A peak measurement given in Fig. 12, this shows a substantial reduction of over 26 dB to the voltage-driven result in Fig. 5. Likewise, the 50-Hz–1-kHz intermodulation distortion is improved, as indicated by Fig. 13.

These results show the importance of eliminating the distortion contributions of the $(Bl)^2$ and Z_s terms in a relative comparison between current drive and voltage drive. Thus at the high-frequency end of the drive unit's operating range, the elimination of performance dependence on coil inductance modulation and eddy current losses is seen to be a valuable asset. Further, the current-driven system is completely free from any voice-coil thermal effects. Although the argument is based on a bass–midrange drive unit, with significant cone displacement, tweeters were found to benefit also, with a more modest 3–7-dB measured distortion reduction across the band, along with the elimination of coil-heating effects.

Finally, the performance independence on linear interconnect errors is also welcome when a low shunt capacitance cable is chosen—a high resultant series inductance being of no significant consequence.

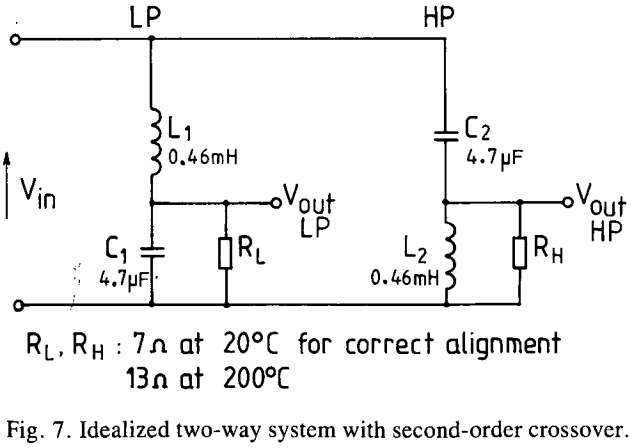


Fig. 7. Idealized two-way system with second-order crossover.

3 DRIVE-UNIT TRANSFER FUNCTION ALIGNMENT UNDER CURRENT DRIVE

Small signal analysis reveals that under current drive, there is a change in frequency response compared with the voltage-driven case, the principal cause being the loss in electromagnetic damping from the low-impedance voice-coil circuit. Consequently, the drive unit Q at fundamental resonance rises to that determined by the mechanical parameters—generally too high for optimal system alignment. To illustrate this, Fig. 14 compares the measured frequency responses of our example drive unit under both current drive and voltage drive. The rise in output around the fundamental resonance under current drive should be noted, along with a reduction in high-frequency rolloff due to the voice-coil inductance no longer appearing in the system transfer function.

In order to realign the acoustic transfer function, three methods have been investigated.

3.1 Electronic Equalization Using Open-Loop Compensation

The addition of a low-level equalizer to redefine the low-frequency alignment of a drive unit under voltage drive is a well-documented technique [20], [21]. The approach is equally applicable to current drive. If the drive-unit transfer function is of the form

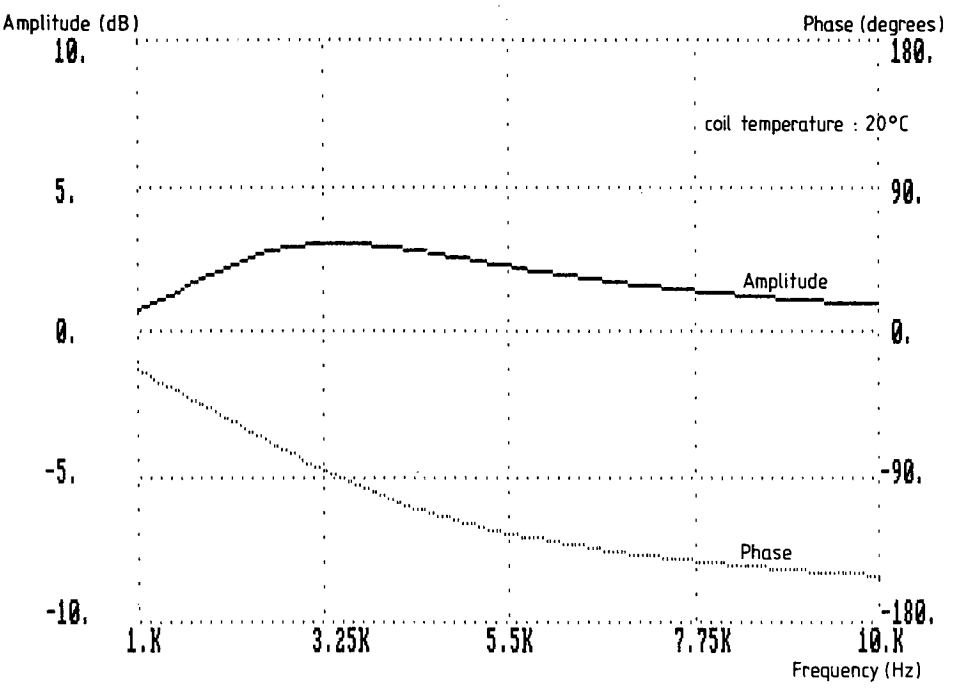
$$G(s) = \frac{s^2 T_s^2}{s^2 T_s^2 + sT_s/Q_m + 1}$$

where T_s is the time constant of fundamental resonance, in seconds, and Q_m is the mechanical drive-unit Q, and the desired low-frequency target alignment is written

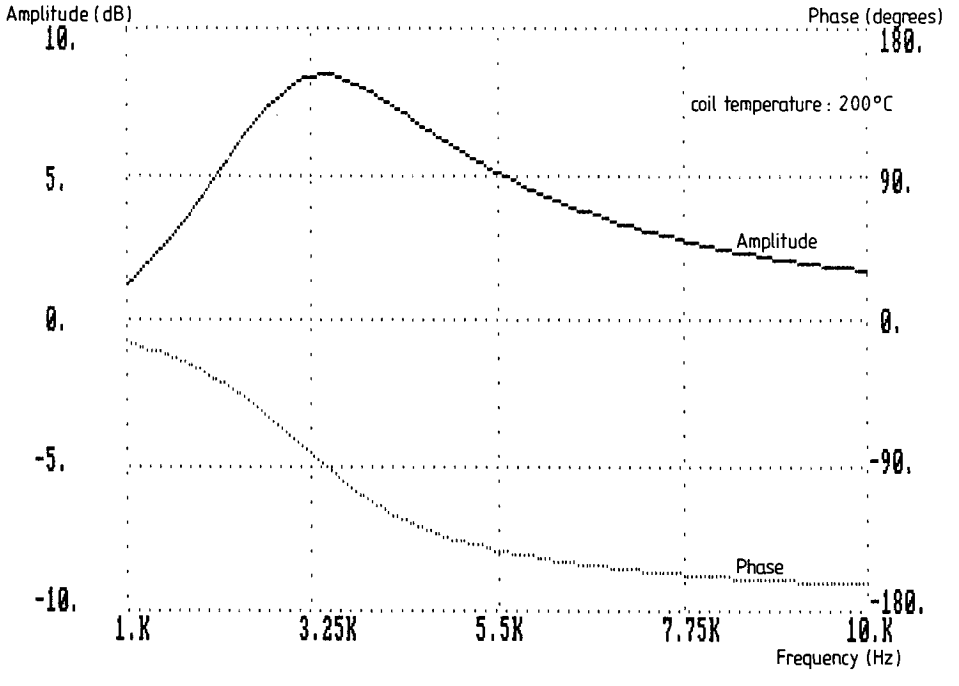
$$G_t = \frac{s^2 T_c^2}{s^2 T_c^2 + sT_c/Q_c + 1}$$

where T_c is the redefined system time constant, in seconds, and Q_c is the compensated Q value, then, assuming a second-order low-frequency alignment is retained, the equalizer transfer function is defined:

$$X(s) = \left(\frac{s^2 T_c^2}{s^2 T_c^2 + sT_c/Q_c + 1}\right) \times \left(\frac{s^2 T_s^2 + sT_s/Q_m + 1}{s^2 T_s^2}\right)$$



(a)



(b)

Fig. 8. Summed high-pass and low-pass outputs for idealized two-way system. (a) at 20°C. (b) at 200°C.

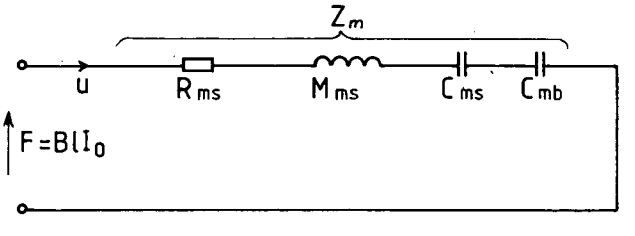


Fig. 9. Mechanical model of drive unit in sealed enclosure under current drive.

that is,

$$X(s) = \frac{s^2T_c^2 + sT_c^2/T_sQ_m + T_c^2/T_s^2}{s^2T_c^2 + sT_c/Q_c + 1} . \quad (3)$$

The equalizer used for experimental purposes is represented by the cascaded integrator structure of Fig. 15. Comparing the transfer function of this system with $X(s)$ in Eq. (3), the time constants, are

$$T_1 = Q_cT_c$$

$$T_2 = \frac{T_c}{Q_c}$$

and the summing constants,

$$C_1 = 1$$

$$C_2 = \frac{Q_cT_c}{Q_mT_s}$$

$$C_3 = \frac{T_c^2}{T_s^2} .$$

This gives the ability to redefine the system Q and, if required, provide low-frequency extension.

Computer simulation of the equalizer and the example drive-unit–enclosure combination shows in Fig. 16(a) the overall system response for a Q realignment to 0.7071 with no resonant frequency shift, while Fig. 16(b) shows the effect of a resonance realignment to 40 Hz, with $Q = 0.7071$.

The disadvantage of this approach is the sensitivity to drive-unit mechanical parameter changes. To investigate this effect, the drive-unit mechanical parameters were subjected to ±20% tolerance and a Monte Carlo analysis based on 25 trials carried out to show the effect of random parametric variations within this range. The results reveal a 2-dB response error standard deviation around the area of fundamental resonance in both cases. However, in practice, mechanical parameter variations are likely to be better controlled with a well-engineered drive unit.

A novel technique of altering low-frequency realignment, which has been described in [22], is the "ace bass" system after Stahl. This method relies on

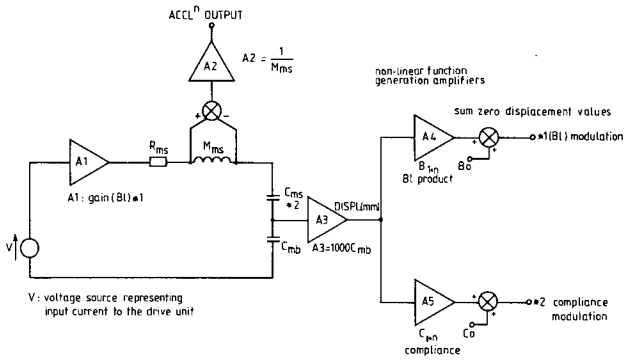


Fig. 10. Simplified nonlinear model for current-driven simulation.

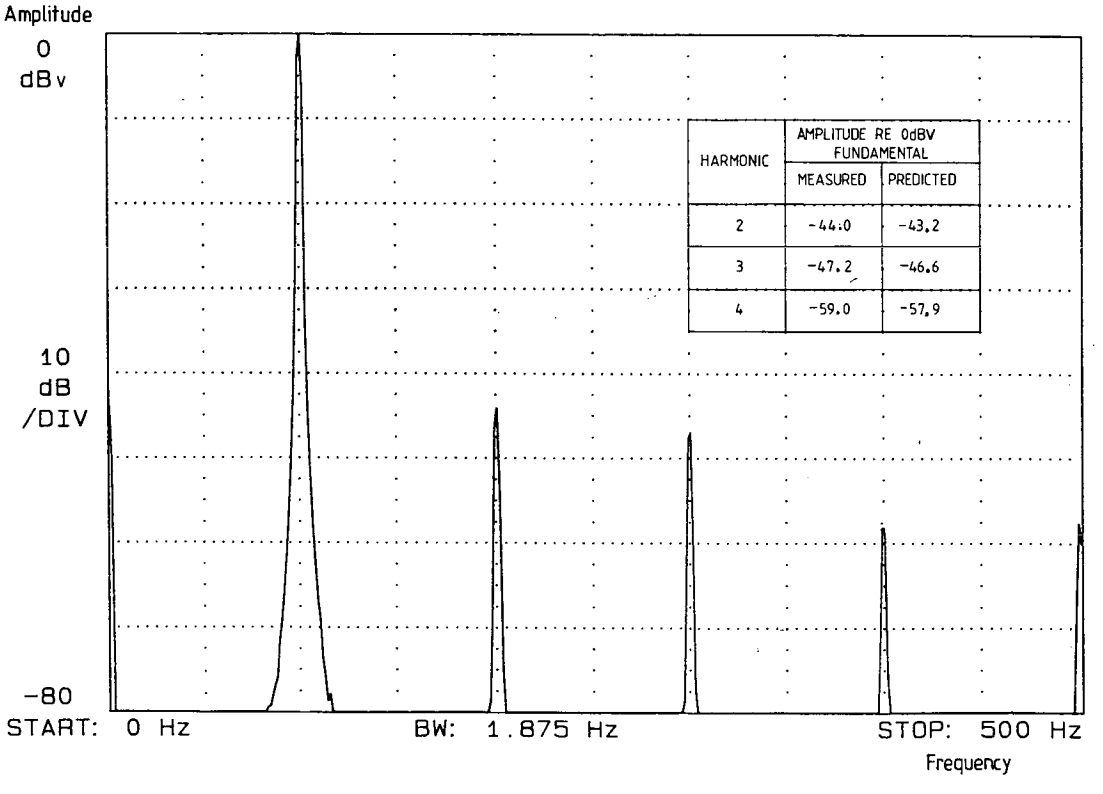


Fig. 11. Measured 100-Hz harmonic distortion, current driven. Table compares with predicted result.

providing the power amplifier with both a negative output resistance to cancel the drive-unit resistance and a synthesized parallel reactance in effect to modify the drive-unit mechanical parameters. While this technique does exhibit insensitivity to mechanical parameter variations (less than 1-dB standard deviation error on the same basis as the open-loop compensator for 40-Hz realignment), it does, as the author admits, incur problems due to voice-coil heating. The error for our example drive unit is shown in Fig. 17, by computer simulation with the voice-coil temperature at 20°C (reference) and increased to 200°C, where the low-

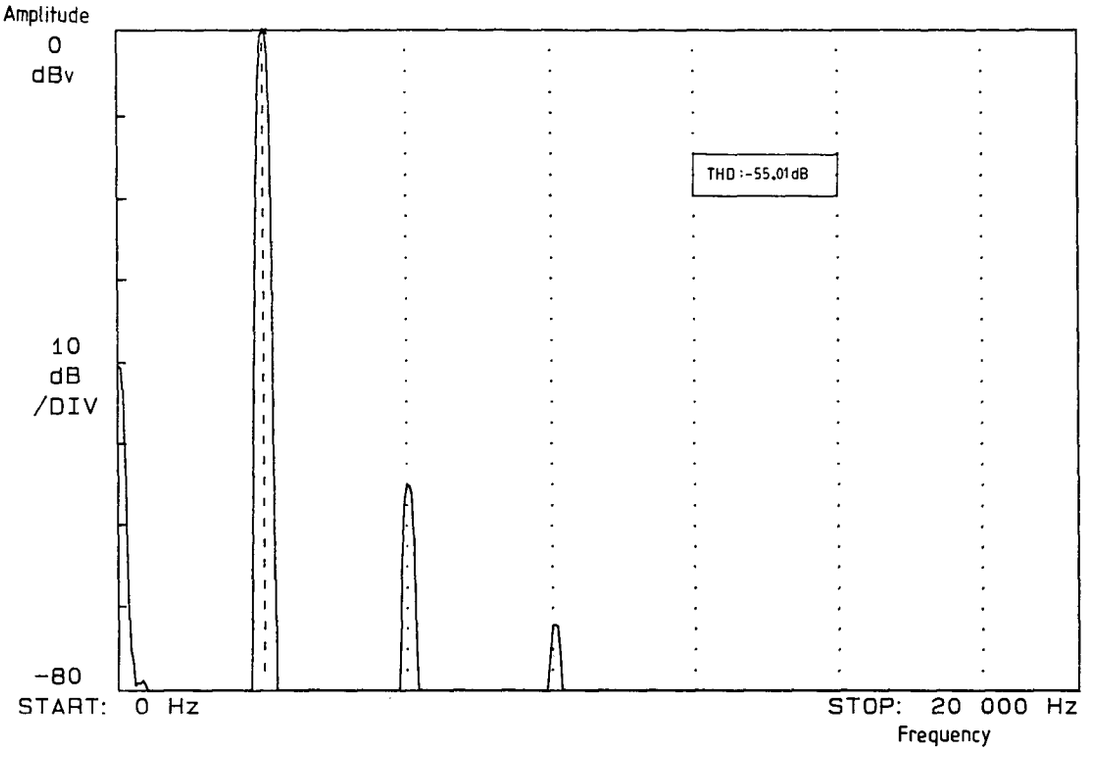


Fig. 12. Measured 3-kHz harmonic distortion, current drive.

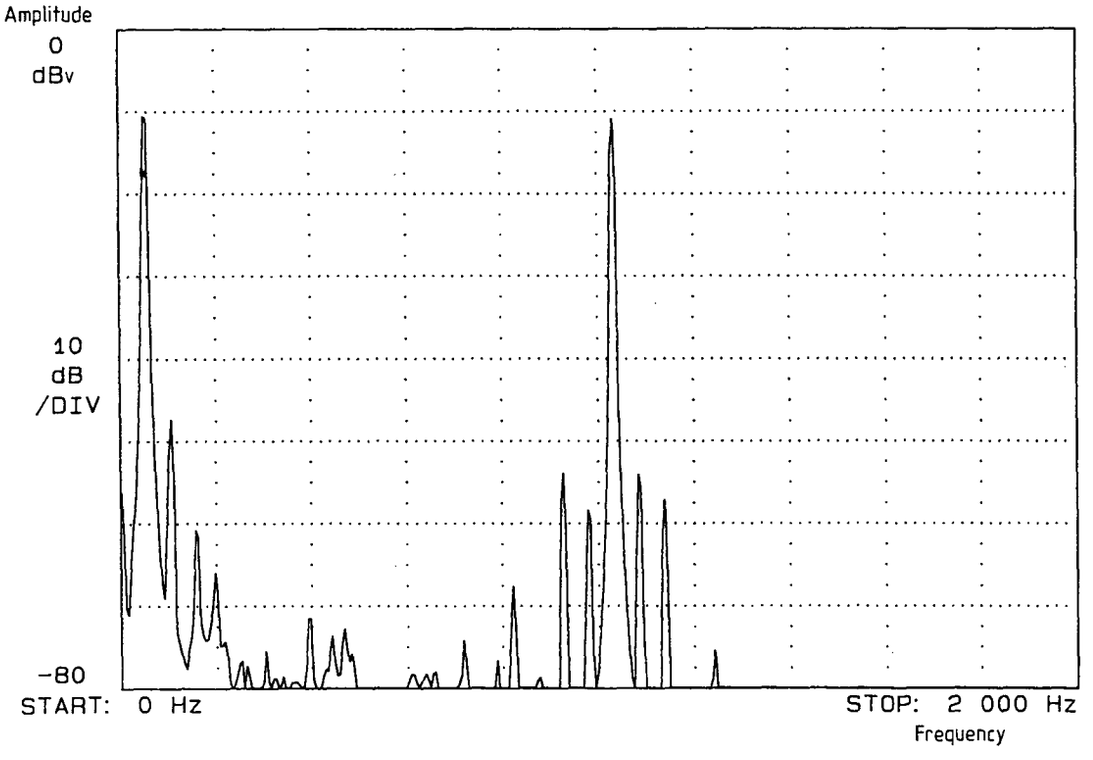
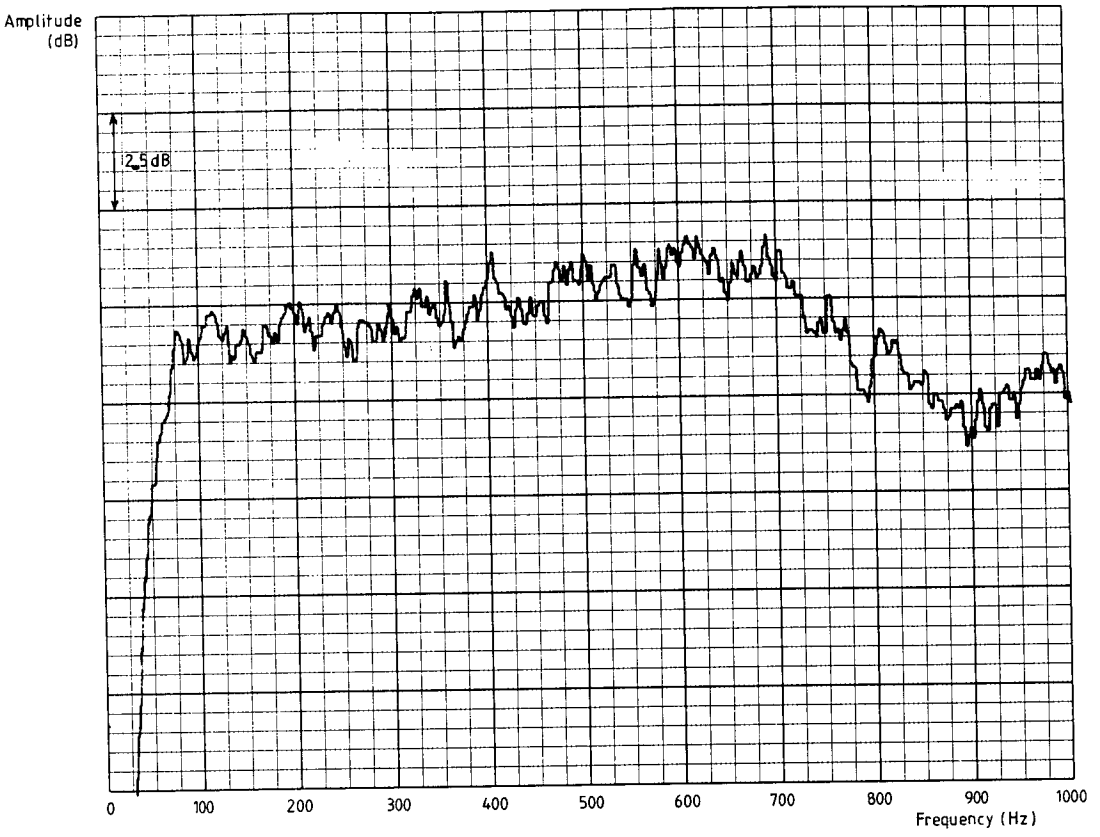
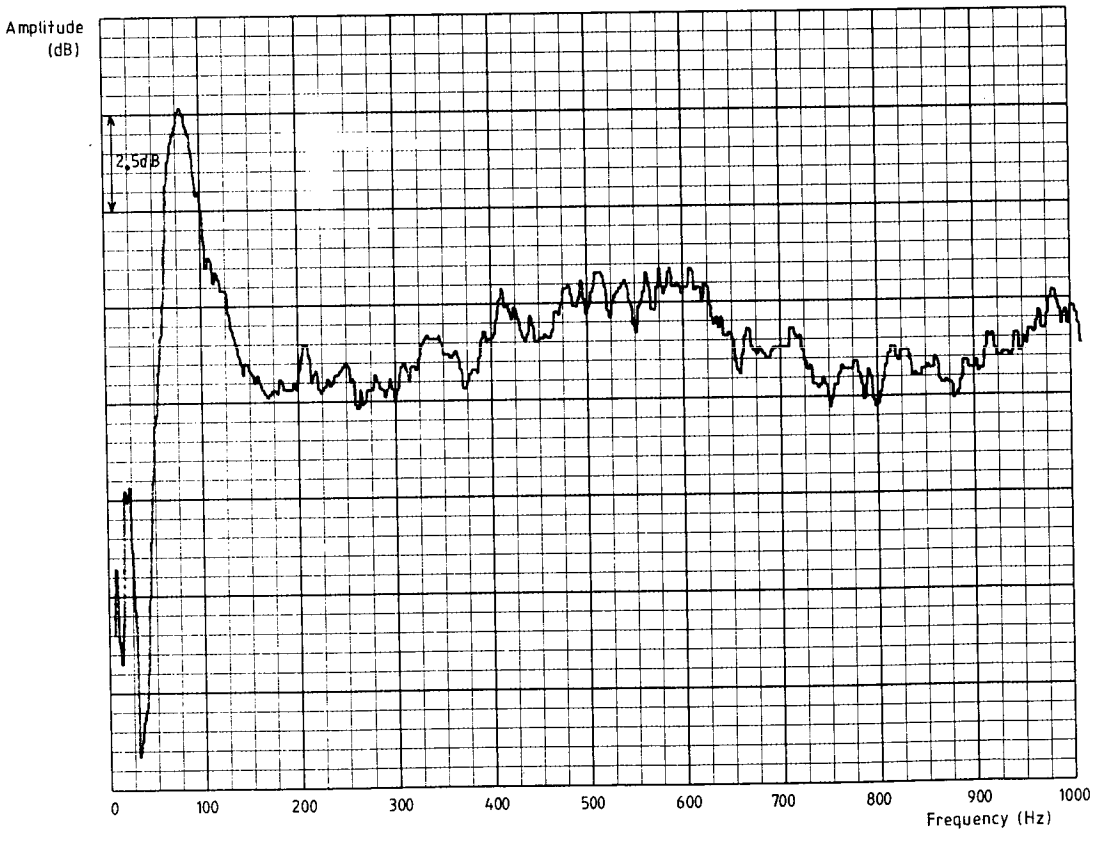


Fig. 13. Measured 50-Hz and 1-kHz intermodulation distortion, current driven.



(a)



(b)

Fig. 14. Measured frequency response of example drive unit. (a) Voltage drive. (b) Current drive.

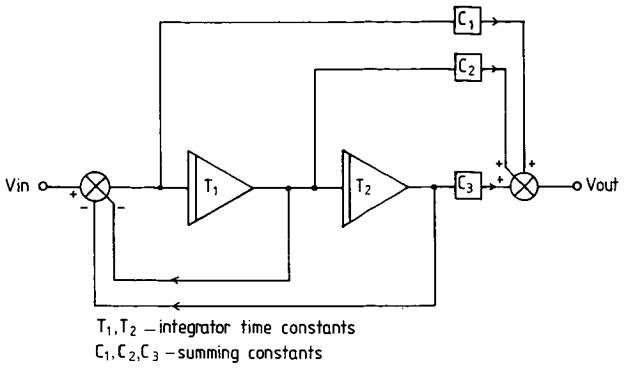


Fig. 15. Representation of open-loop equalizer.

frequency alignment was set to 40-Hz resonance with $Q = 0.7071$. Where low-frequency extension is required, the extra power needed to combat the drive unit's falling response means that coil heating effects are particularly troublesome, hence reinforcing the need for current drive in this type of application.

3.2 Motional Feedback

Motional feedback is considered the optimal method for Q alignment of low-frequency drive units under current control and was consequently incorporated into the prototype development system. Our earliest reference to the technique is due to Voight in 1924 [23],

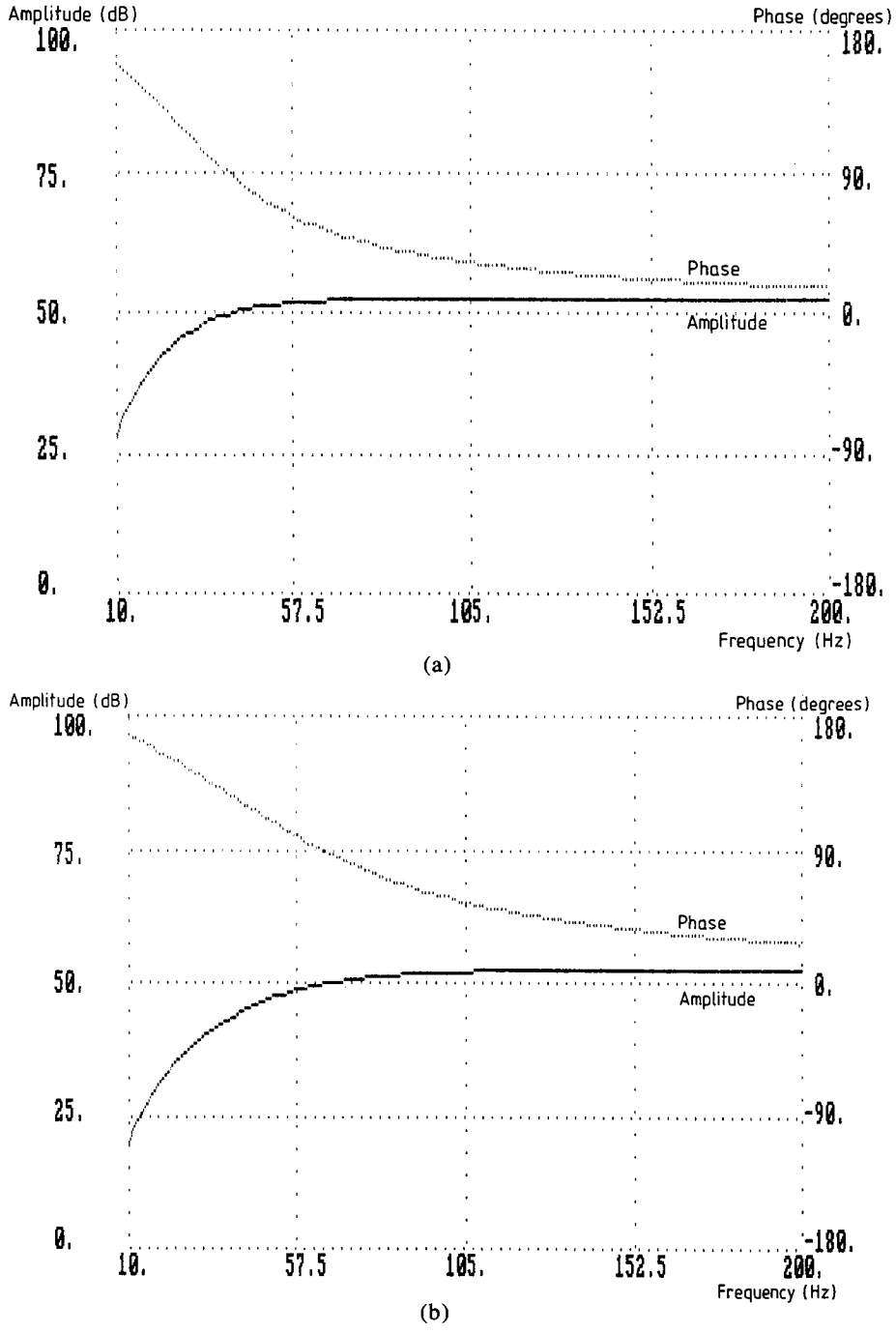
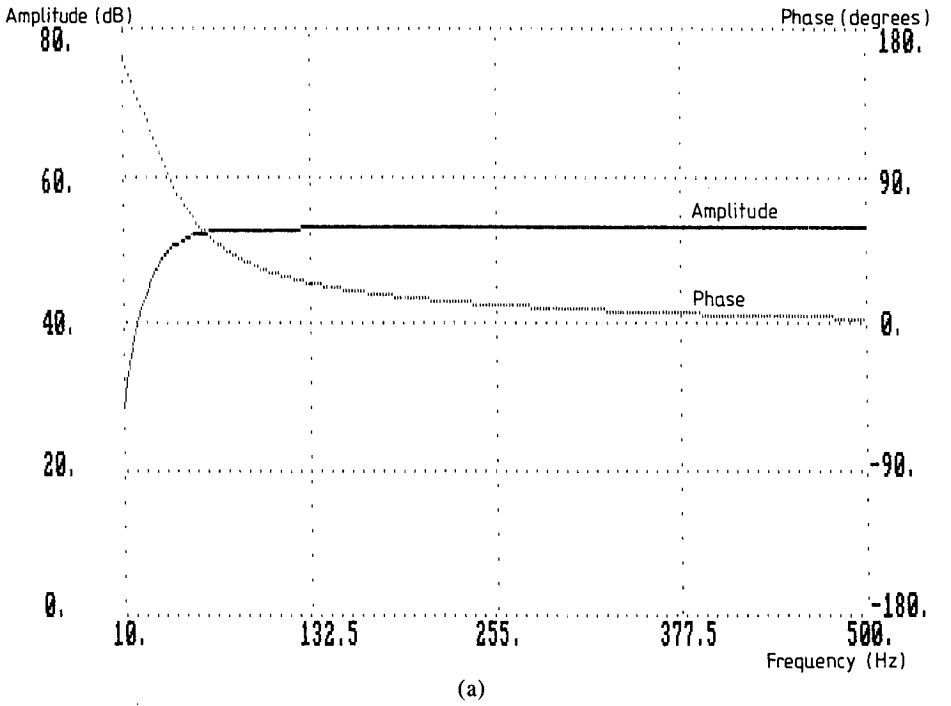


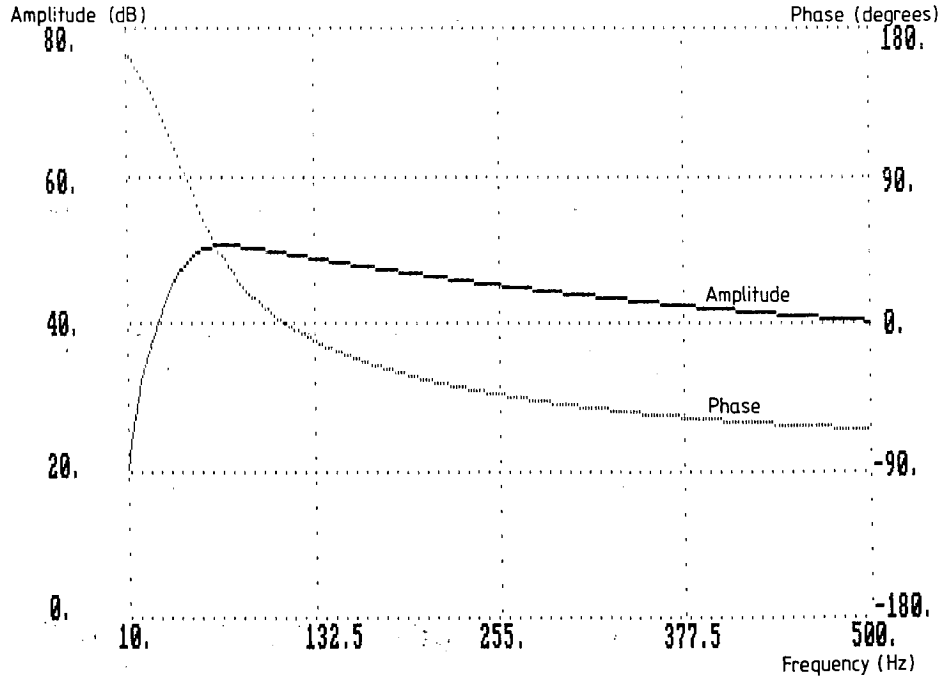
Fig. 16. Simulated current-driven system response with equalizer. (a) No resonance change (65 Hz), $Q = 0.7071$. (b) Resonance lowered to 40 Hz, $Q = 0.7071$.

where the lack in damping from an open-loop tube output stage led to similar problems as faced under pure current drive. The approach was later abandoned for general use when Black [24] formalized negative feedback techniques and amplifier output impedance could be reduced. Since then there has been much interest in motional feedback in high-performance applications, using a variety of sensing methods to obtain velocity, displacement, or acceleration feedback [25]–[28].

The method selected in this study was to wind a sensing coil over the primary drive coil for reasons of cost effectiveness, with little additional complexity over a standard drive unit. The penalty of this mechanical simplicity is that as well as generating a signal proportional to cone velocity, there is also transformer coupling that induces an error from the driving coil into the sensing coil. Methods used to deal with this effect have included the use of additional neutralizing coils [29], [30]. In this case, the rather different method of electronic compensation has been adopted. Fig. 18 shows how this has been achieved, together with a



(a)



(b)

Fig. 17. "Ace bass" system frequency response simulations. Resonance set to 40 Hz, $Q = 0.7071$. (a) Voice-coil temperature 20°C. (b) Voice-coil temperature 200°C.

block diagram of the prototype system. So that the sensing coil does not reintroduce thermal errors, it must be interfaced to a high-input impedance buffer amplifier. The coupling error compensator consists of a filter matched to the transfer function of the coupling error characteristic, which increases by around 15 dB per decade up to 1 kHz at zero coil displacement. Some change in both the magnitude and the slope of this error is apparent with coil displacement, together with an unpredictable response above 1 kHz. To reduce this high-frequency residual error, the second-order low-pass filter at 1 kHz in the velocity feedback loop proves effective and in any event is required to maintain loop stability. Measurements have shown no increase in high-frequency distortion (at 3 kHz) over the open-loop case, indicating the effectiveness of this method.

To analyze the system, consider the simplified representation of Fig. 19, consisting of transconductance power amplifier, drive unit with sensing coil, and feedback path. The output voltage from the sensing coil V_s is written

$$V_s = (Bl)_s u$$

where $(Bl)_s$ is the sensing coil Bl product, in newtons per ampere, and u is the cone velocity, in meters per second. Also,

$$I_0 = [V_{in} - k(Bl)_s u]\, g_m$$

where

I_0 = amplifier output current, amperes
V_{in} = input voltage, volts
k = feedback constant
g_m = amplifier transconductance, siemens.

Using Eq. (2),

$$u = (Bl)\,[V_{in} - k(Bl)_s u]\, Y_m\, g_m$$

where Y_m is the admittance of the mechanical drive-unit model of Fig. 9. Thus

$$u = \frac{V_{in}\,(Bl)\,g_m}{1/Y_m + k(Bl)_s\,(Bl)\,g_m} . \quad (4)$$

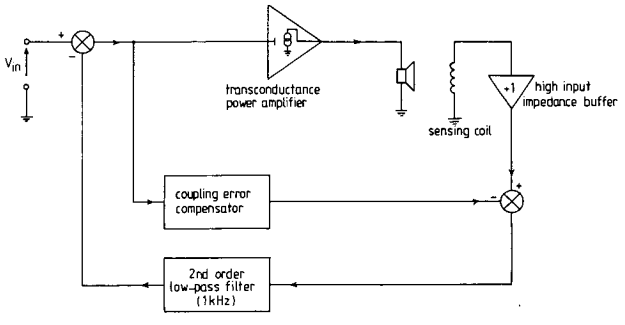


Fig. 18. Block diagram of prototype motional feedback system.

Examining the mechanical drive-unit model reveals

$$Y_m = \frac{sC_{mt}}{s^2 M_{ms} C_{mt} + sC_{mt} R_{ms} + 1} \quad (5)$$

where C_{mt} is the total mechanical compliance of the system, that is,

$$C_{mt} = \frac{C_{ms}\, C_{mb}}{C_{ms} + C_{mb}} .$$

Substituting this result into Eq. (4) gives

$$u = [V_{in}\,(Bl)sC_{mt}\, g_m][M_{ms}C_{mt}\,\{s^2 + s\underbrace{(1/M_{ms})\,[R_{ms} + (Bl)_s\,(Bl)kg_m]}_{\omega_0/Q} + \underbrace{1/M_{ms}\,C_{mt}}_{\omega_0^2}]^{-1} \quad (6)$$

Thus, for a second-order system,

$$Q = \frac{1}{(C_{mt}/M_{ms})^{0.5}\,[R_{ms} + g_m\, k\,(Bl)\,(Bl)_s]} \quad (7)$$

which may be rearranged to give

$$k = \frac{1/Q\,(M_{ms}/C_{mt})^{0.5} - R_{ms}}{(Bl)\,(Bl)_s\, g_m} . \quad (8)$$

Investigation of system performance was carried out as for the open-loop case, with nonlinear transient analysis followed by measurement. It should be noted that in the prototype, the sensing coil followed the same Bl profile as the main coil, although to achieve a further low-frequency distortion reduction over the open-loop case, a more elaborate linear sensing mechanism is required. The transient analysis model will not be detailed, as it follows the earlier methodology—the velocity feedback signal was derived from the voltage across the mechanical resistance R_{ms}, and the feedback path loop stability filter was set to 1 kHz, second order. No transformer coupling effects were included in the model. For a 100-Hz at 1A peak sine-wave excitation, the measured distortion spectra are shown in

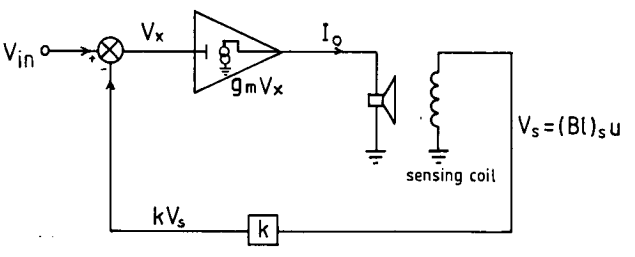
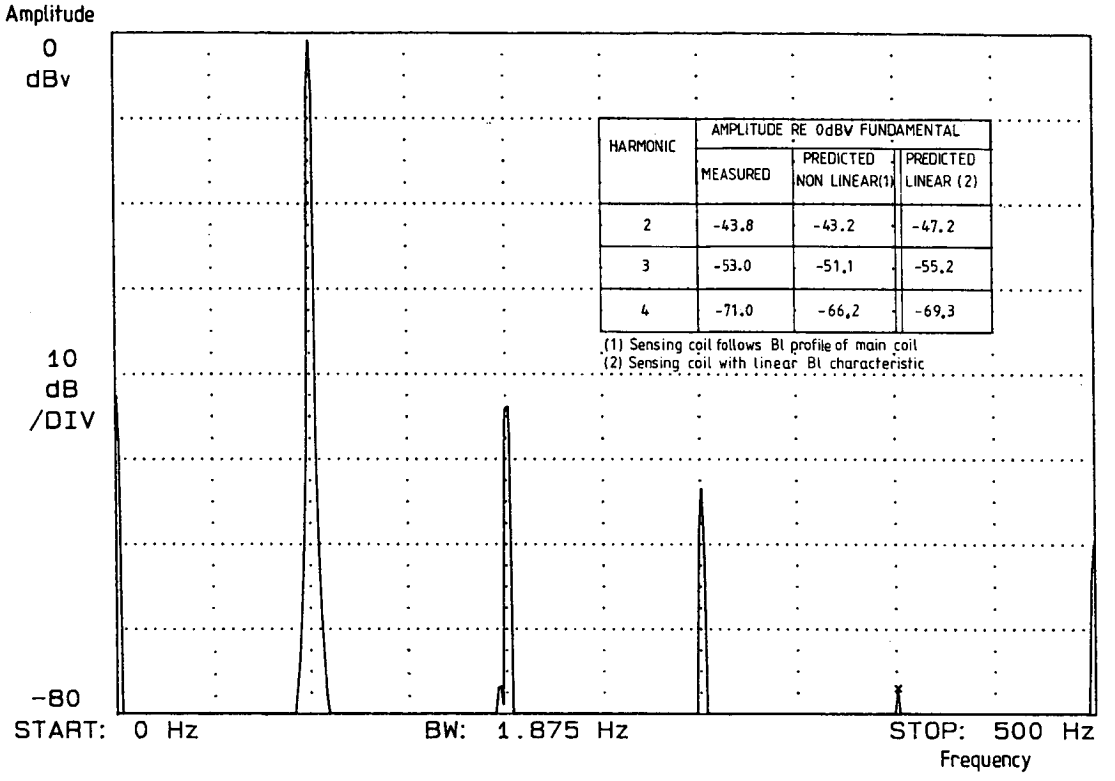


Fig. 19. Simplified motional feedback model for analysis purposes.

Fig. 20, along with comparative data from the simulation for a system Q of 0.7071. The results are seen to be broadly similar to the open-loop case for nonlinear Bl sensing, with a distortion reduction of around 4 dB on the second and third harmonics resulting from predicted linear velocity sensing.

The measured frequency response (Fig. 21) is seen to be flatter than both the voltage-driven and the open-



HARMONIC	AMPLITUDE RE 0dBV FUNDAMENTAL		
	MEASURED	PREDICTED NON LINEAR(1)	PREDICTED LINEAR (2)
2	-43.8	-43.2	-47.2
3	-53.0	-51.1	-55.2
4	-71.0	-66.2	-69.3

(1) Sensing coil follows Bl profile of main coil
(2) Sensing coil with linear Bl characteristic

Fig. 20. Measured 100-Hz harmonic distortion, current driven with velocity feedback. Table compares with predicted results for both linear and nonlinear sensing-coil Bl profiles.

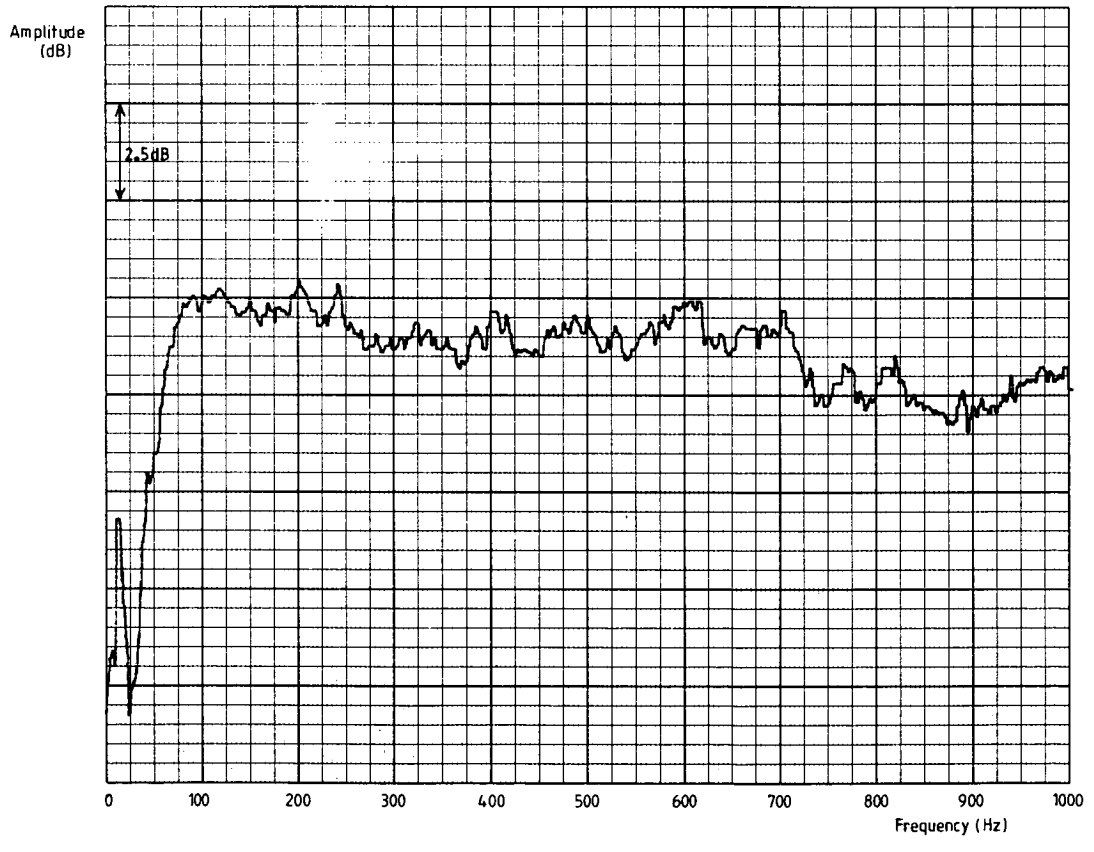


Fig. 21. Measured closed-loop frequency response for velocity feedback current-driven case, Q = 0.7071.

loop current-driven cases (Fig. 14). The effect of altering the feedback factor k is shown by the measured step responses in Fig. 22 for $Q = 3.0$ (open loop), $Q = 1.0$, and $Q = 0.7$.

3.3 High-Frequency Drive Unit with Electrically Conductive Former

At high frequencies, the preferred damping method is to use a drive unit with inherent electromagnetic damping through a conductive coil former. With such a device, experiment has shown a negligible contribution from voice-coil damping under voltage drive. Due to the improved linearity of high-frequency drive units, resulting from low coil displacement, the distortion reduction of current drive is less marked (around 3–7-dB reduction for drive units tested). However, the advantages in terms of freedom from thermally induced response errors and power compression are still valid.

4 POWER AMPLIFIER TOPOLOGIES FOR CURRENT DRIVE

A voltage-driven system requires a power amplifier with adequate bandwidth, low distortion, and a low output impedance which is linear and frequency independent. With current drive, the latter requirement translates to a high output impedance, which again should be linear and frequency independent. Also, the current demand under voltage drive [19], [31]–[34] becomes a problem of voltage demand under current drive. Consequently the maximum current delivery is known, which aids amplifier protection, as the system is inherently self-limiting.

The most basic strategy for generating a high output impedance is by the use of negative current feedback from a sensing resistor in the loudspeaker ground return [12], [13]. A typical configuration is shown in Fig. 23. Analysis of this system reveals that the transconductance g_m is given by

$$g_m = \frac{I_0}{V_{in}} = \frac{A}{(Z_0 + Z_L) + R_f(1 + A)} \qquad (9)$$

where

I_0 = load current, amperes
V_{in} = input voltage, volts
Z_o = open-loop output impedance, ohms
Z_L = drive-unit impedance, ohms
R_f = current-sensing resistor, ohms
A = forward gain of amplifier.

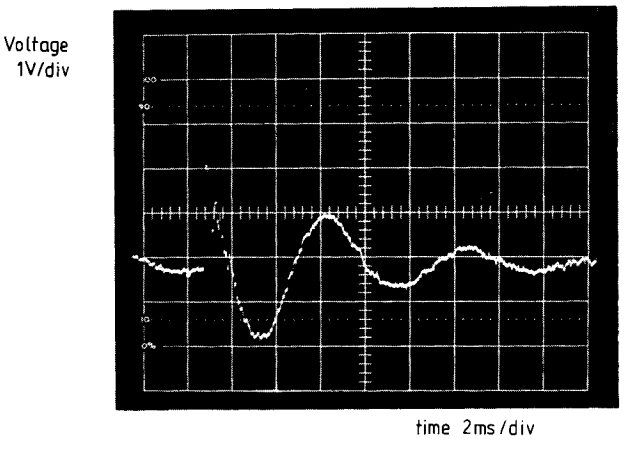
Consequently the output impedance Z_e may be written

$$Z_e = (1 + A)R_f + Z_o \; . \qquad (10)$$

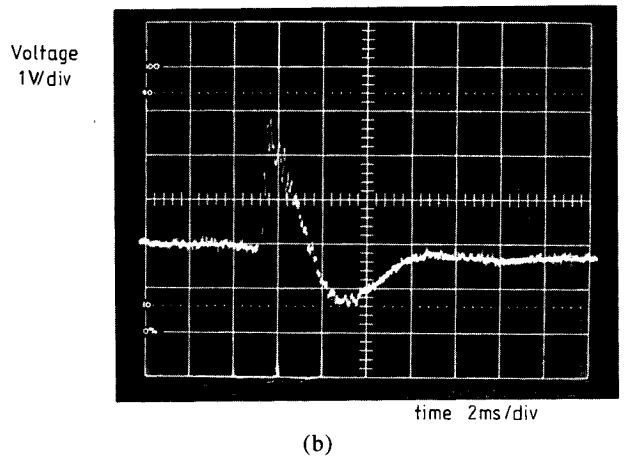
This configuration, although a feasible solution, has two main limitations. First, the forward gain of the amplifier is frequency dependent, falling with increasing frequency as a result of its dominant pole. As a consequence, the output impedance falls with a similar characteristic. Second, the loudspeaker impedance is both frequency dependent and nonlinear, leading to a modulation of the system transconductance [Eq. (9)] and being analogous to interface distortion in a conventional voltage amplifier.

To investigate this effect in more detail, consider an error function E_1 which is defined as

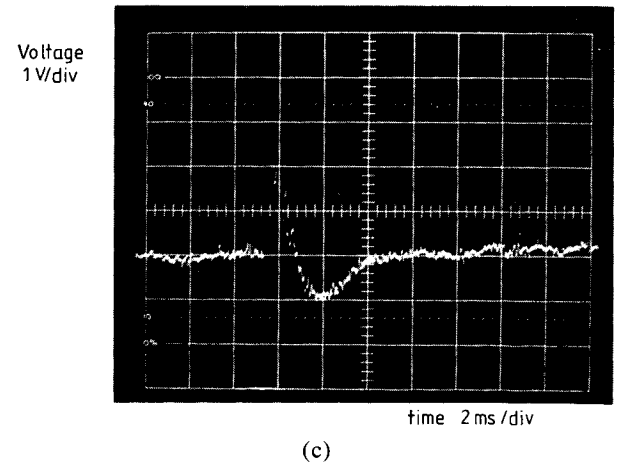
$$E_1 = \frac{g_m}{g_t} - 1$$



(a)



(b)



(c)

Fig. 22. Measured step responses for velocity feedback current-driven system. (a) $Q = 3.0$ (no feedback). (b) $Q = 1.0$. (c) $Q = 0.7$.

where g_t is the target transconductance, that is, $g_t = 1/R_f$. Thus, using Eq. (9),

$$E_1 = \frac{AR_f}{(Z_o + Z_L) + R_f(1 + A)} - 1$$

that is,

$$E_1 = -\frac{Z_o + R_f + Z_L}{(Z_o + Z_L) + R_f(1 + A)} .$$

Assuming $(1 + A)R_f >> (Z_o + Z_L)$,

$$E_1 \approx -\frac{Z_o + R_f + Z_L}{AR_f} \approx -\frac{Z_L}{AR_f} . \quad (11)$$

If we suppose as a numerical example that E_1 should be less than 0.1%, then from Eq. (11),

$$A > \frac{1000Z_L}{R_f} .$$

Hence, if R_f is set to 0.5 Ω and Z_L assumes a maximum value of 20 Ω, then $A_{dB} > 92$ dB. This is seen to be a high open-loop gain to maintain and illustrates well the limitations of the current feedback technique, particularly as Z_L is nonlinear.

A more optimal solution to the problem is to provide a cascaded open-loop grounded-base isolation stage in the amplifier structure to isolate the transconductance amplifier from the load, as shown in Fig. 24. Several advantages result from this enhanced technique.

1) Output impedance is essentially independent of the transconductance amplifier A_t, being a function of the grounded-base isolation stage.

2) Performance of amplifier A_t is isolated from the nonlinear load Z_L, thus eliminating interface distortion through loop gain modulation [see Eq. (11)].

3) Amplifier A_t can, if desired, operate in class A with its own supply $\pm V_{s1}$, which may be of low value to minimize power dissipation.

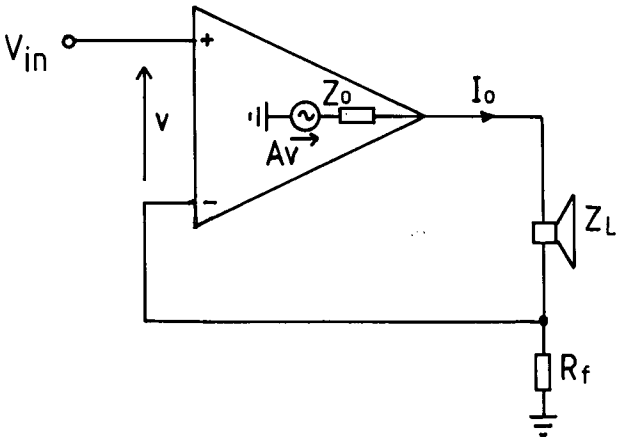


Fig. 23. Basic current feedback derived transconductance amplifier.

4) The grounded-base stage can operate in class AB, with a small standing current giving minimal distortion penalty, as current $I_L = I_0$, except for any base current leakage to ground.

5) Unlike the topology of Fig. 23, the loudspeaker load is referenced to ground, which simplifies installation and reduces the effect of interconnect capacitance at high frequencies.

6) If points P and Q (Fig. 24) are coincident, grounding-related errors are reduced due to signal currents forming well-defined closed paths.

7) The circuit topology is effectively a complementary cascode, therefore offering performance advantages in bandwidth and linearity.

8) As the grounded-base stage operates open loop, it does not degrade the loop gain and bandwidth characteristic of amplifier A_t.

9) The supply voltages $\pm V_{s2}$ can, in principle, be made adaptive to increase efficiency, when used in conjunction with a predictive digital processor.

Fig. 25 shows an alternative power amplifier topology, this time taking the form of a current gain stage. It must therefore be fed from a transconductance preamplifier. It has the advantage of having a ground-referenced power supply $\pm V_{s1}$ for the current amplifier A_i, meaning that in practice, several amplifiers in an active system may share a common supply, reducing complexity and cost.

Several prototype amplifiers have been built using these techniques. The first was based on the Fig. 24 complementary cascode configuration and operated with the transconductance amplifier A_t in class A with error feedback correction [35], [36], while the second was based on the Fig. 25 topology and used class AB operation for the current gain amplifier A_i with more extensive error correction and also moderate overall feedback. Both amplifiers were evaluated in terms of

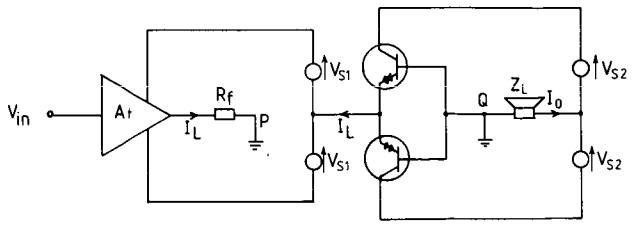


Fig. 24. Transconductance power amplifier using grounded-base output stage in complementary cascode configuration.

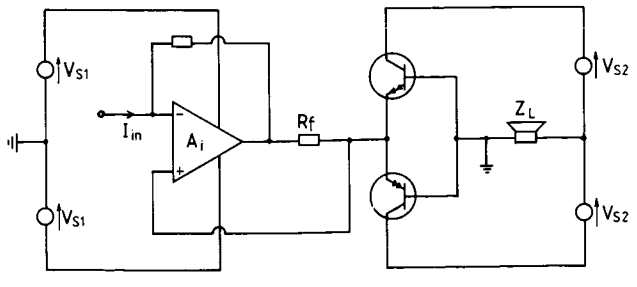


Fig. 25. Alternative power amplifier topology.

conventional measurements (Table 2) and were found comparable to typical high-performance voltage power amplifiers. On a practical note, it is judged important to provide adequate high-frequency current gain in the common-base stage to avoid distortion due to base leakage current to ground.

5 POWER AMPLIFIER AND DRIVE-UNIT PROTECTION UNDER CURRENT DRIVE

As the maximum available current from the transconductance power amplifier is an inherent design parameter, it is therefore self-limiting, so the system is simpler to protect. Indeed, this self-limiting characteristic implies that the designer need not be so concerned about protection circuitry, which has been cited as a source of degradation [37], [38].

Unlike the voltage power amplifier, which requires a series switching element for loudspeaker protection against offsets and other fault conditions, the current power amplifier requires a shunting element across the loudspeaker, thus avoiding the problems of contact degradation with time. Also a series fuse may be added without signal impairment, whereas with a voltage power amplifier, thermal modulation of the fuse wire resistance offers a source of distortion. Whereas a voltage amplifier requires short-circuit protection, a current amplifier is sensitive to open-circuit conditions. However, tests on the experimental amplifiers constructed have not given rise to a failure mode under open circuit.

A major factor concerning system reliability is drive-unit thermal failure. With conventionally powered loudspeakers, the coil current (and hence power dissipation) falls as temperature increases due to the thermal coefficient of the coil, giving a degree of protection. Elaborate protection systems have, however, been described for loudspeakers under voltage drive [39], [40]. Under current drive, no such self-limiting occurs. Indeed, it is an effect we are seeking to avoid. Thus, particularly for high-power and high-reliability installations, a method of sensing voice-coil temperature is required. This is best explained with reference to the block diagram system of Fig. 26. In addition to the main transconductance power amplifier and drive unit, a second low-power transconductance amplifier is provided to drive an impedance scaled model of the drive unit. In this model, R_e' represents the voice-coil resistance at room temperature and Z_m' the drive-unit motional impedance. After taking account of the scaling factors in the current level and impedance of the reference network, the difference in voltage across the drive unit and reference network is obtained by a differential amplifier. The rms values of the input voltage V_{in}, which is proportional to the drive-unit current and of the differential amplifier output, are then fed to a divider network to produce a voltage V_{out} representing any increase in coil resistance due to heating. Presuming the temperature coefficient of the coil material is known, a measure of temperature is then determined.

The output voltage V_{out} may be used to drive a comparator to shut down power to the loudspeaker drive unit at a predetermined temperature. Alternatively, it can be used to progressively attenuate the drive-unit current to a safe level, or to provide curtailment of low-frequency extension to reduce power dissipation. The latter techniques are of particular interest to studio monitors, where high reliability and continuity of operation are paramount.

6 PROTOTYPE TWO-WAY ACTIVE LOUDSPEAKER SYSTEM

The ideas presented in this paper have been incorporated into a working prototype two-way active loud-

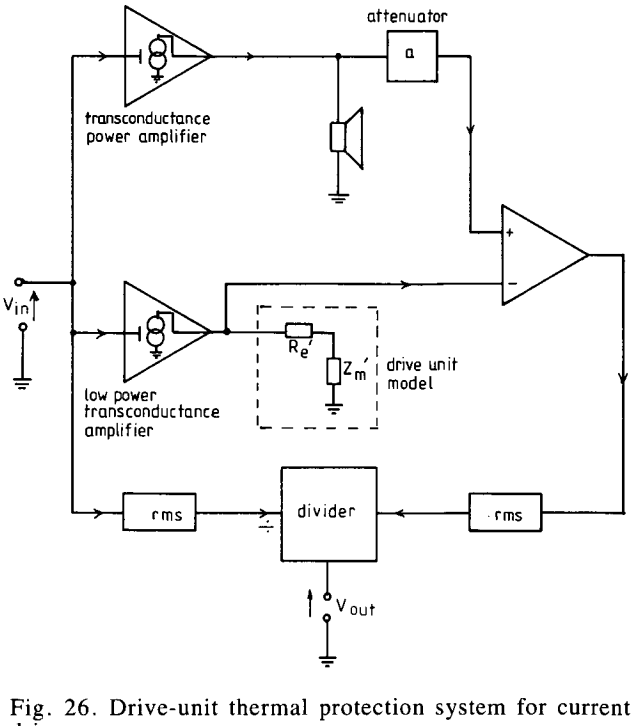


Fig. 26. Drive-unit thermal protection system for current drive.

Table 2. Performance comparison of prototype power amplifiers.

Measurement parameter	Test condition/notes	Class A design	Class AB design
Rated output power	8 Ω resistive load	75 W average	75 W average
Total harmonic distortion re rated power	20 Hz	−88 dB	−79 dB
	1 kHz	−84 dB	−86 dB
	20 kHz	−79 dB	−68 dB
Intermodulation distortion at rated power	19 kHz and 20 kHz at equal levels	<−90 dB	−86 dB
Hum and noise	Unweighted re full power	−91 dB	−90 dB
Small-signal bandwidth	−3 dB	dc–50 kHz	0.1 Hz–50 kHz

speaker system, based on the Celestion SL600 loudspeaker, because its high drive-unit quality and low level of enclosure coloration would theoretically make the benefits of current drive apparent on audition.

The system employs a discrete low-level electronic crossover, feeding individual current power amplifiers. One of the power amplifiers, based on the topology of Fig. 25, running in class AB with error correction is shown in detail in Fig. 27. Both power amplifiers are mounted on the loudspeaker stand, which aids thermal dissipation, with transformers mounted on the base to give mechanical stability. Fig. 28 shows the complete assembly. The crossover, motional feedback control circuits, coupling error compensator, and transconductance line amplifiers are housed in a separate enclosure (Fig. 29), which also incorporates level controls for the input signal. The crossover time constants are each independently adjustable for trimming, to enable comparison with the original voltage-driven loudspeaker to be made on a fair basis.

7 CONCLUSIONS

This paper has presented an alternative approach to the amplifier–loudspeaker interface, where numerous advantages have been cited through technical discussion, nonlinear computer modeling, and measurement. The principal advantages of current drive are seen to be an elimination of performance dependence on voice-coil resistance (which is thermally modulated) and also coil-inductive effects, which give rise to high-frequency distortion, along with nonlinear electromagnetic damping due to Bl variations. The technique is similarly insensitive to the lumped series elements of the amplifier–loudspeaker interconnect. However, it is often necessary to lower the system Q caused by the loss of amplifier-generated damping, either by open-loop compensation, by special drive-unit design, or by motional feedback, where the latter is regarded as the optimal method at low frequencies.

Having attempted a broad coverage of the principles of current drive, it is hoped that a greater interest in and awareness of the technique will result. The authors perceive digital signal processing (DSP) as being integral to further developments in terms of crossovers, motion feedback signal processing, and drive-unit protection against thermal and excursion damage. The subject of digital crossover design has already been researched in depth within the Group. This resulted in a two-way working system, operating on the data stream from a CD player [41]. A further area of DSP to be investigated is compensation for the drive-unit Bl profile with coil displacement sensing, which under voltage drive would not be so amenable to correction, due to the more complicated nature of nonlinearities present in the system transfer function. In addition, DSP techniques have the potential to improve transconductance power amplifier efficiency by using modulated switched-mode power supplies.

While the research has been directed at moving-coil drive units, there is no reason why current drive should not be applied to ribbon transducers, which are often mechanically well damped and would benefit from removal of the matching transformer needed under voltage drive.

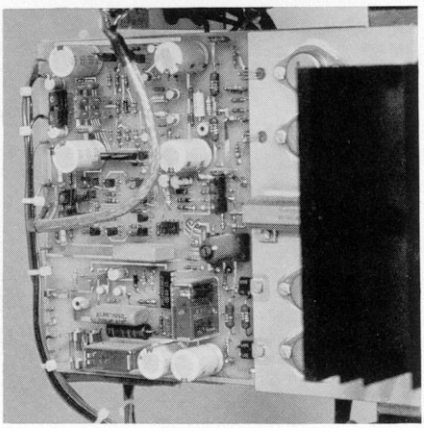


Fig. 27. View of prototype power amplifier based on Fig. 25.

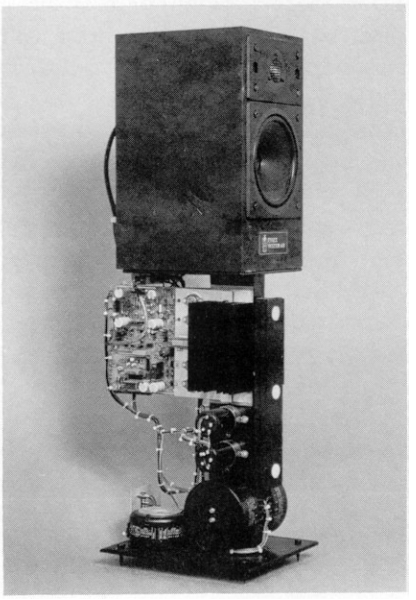


Fig. 28. Prototype active loudspeaker system.

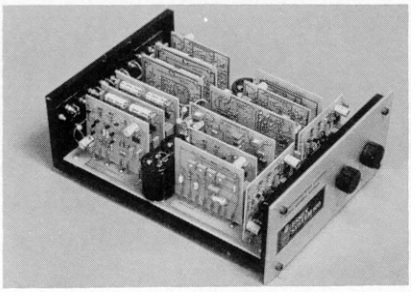


Fig. 29. View of control unit. System includes low-level crossovers, velocity feedback circuitry, and transconductance line amplifiers.

The technique of current drive should find wide applications in both high-performance domestic and studio applications, where it is felt that useful performance gains will be made over conventional systems.

8 ACKNOWLEDGMENT

This research was initially supported by the Science and Engineering Research Council of Great Britain under a research studentship program. Our thanks are due to Ed Form and more recently Graham Bank of Celestion International for providing the loudspeaker system used as the basis for the experimental work, along with the provision of modified drive units.

9 PATENT PROTECTION

The authors would like to point out that aspects of the work documented in this paper are the subject of a U.K. patent application.

10 REFERENCES

[1] H. D. Harwood, "Loudspeaker Distortion Associated with Low-Frequency Signals," *J. Audio Eng. Soc.*, vol. 20, pp. 718–728 (1972 Nov.).

[2] A. Dobrucki and C. Szmal, "Nonlinear Distortions of Woofers in the Fundamental Resonance Region," presented at the 80th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, p. 389 (1986 May), preprint 2344.

[3] A. J. M. Kaizer, "Modeling of the Nonlinear Response of an Electrodynamic Loudspeaker by a Volterra Series Expansion," *J. Audio Eng. Soc.*, vol. 35, pp. 421–433 (1987 June).

[4] R. J. Newman, "Do You Have a Sufficient Quantity of Acoustical Benzin? Aspects Related to the Significance of Diaphragm Excursion," presented at the 80th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, p. 388 (1986 May), preprint 2342.

[5] M. R. Gander, "Dynamic Linearity and Power Compression in Moving-Coil Loudspeakers," presented at the 76th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 32, pp. 1008–1009 (1984 Dec.), preprint 2128.

[6] T. S. Hsu, S. H. Tang, and P. S. Hsu, "Electromagnetic Damping of High-Power Loudspeakers," presented at the 79th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 33, p. 1011 (1985 Dec.), preprint 2297.

[7] C. A. Henricksen, "Heat-Transfer Mechanisms in Loudspeakers: Analysis, Measurement, and Design," *J. Audio Eng. Soc.*, vol. 35, pp. 778–791 (1987 Oct.).

[8] W. J. Cunningham, "Nonlinear Distortion in Dynamic Loudspeakers due to Magnetic Effects," *J. Acoust. Soc. Am.*, vol. 21, pp. 202–207 (1949 May).

[9] J. R. Gilliom, P. L. Boliver, and L. C. Boliver, "Design Problems of High-Level Cone Loudspeakers," *J. Audio Eng. Soc. (Project Notes/Engineering Briefs)*, vol. 25, pp. 294–299 (1977 May).

[10] M. R. Gander, "Moving-Coil Loudspeaker Topology as an Indication of Linear Excursion Capability," *J. Audio Eng. Soc.*, vol. 29, pp. 10–26 (1981 Jan./Feb.).

[11] H. F. Olson, "Analysis of the Effects of Nonlinear Elements upon the Performance of a Back-Enclosed, Direct Radiator Loudspeaker Mechanism," *J Audio Eng. Soc.*, vol. 10, pp. 156–163 (1962 Apr.).

[12] J. A. M. Catrysse, "On the Design of Some Feedback Circuits for Loudspeakers," presented at the 73rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 364 (1983 May), preprint 1964.

[13] R. A. Greiner and T. M. Sims, Jr., "Loudspeaker Distortion Reduction," *J. Audio Eng. Soc.*, vol. 32, pp. 956–963 (1984 Dec.).

[14] M. J. Hawksford, "The Essex Echo—Unification," *Hi-Fi News Rec. Rev.*, pt. 1 (1986 May), pt. 2 (1986 Aug.), pt. 3 (1986 Oct.), pt. 4 (1987 Feb.).

[15] M. N. T. Otala and J. Lammasniemi, "Intermodulation Distortion at the Amplifier–Loudspeaker Interface," presented at the 59th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 26, p. 382 (1978 May), preprint 1336.

[16] J. Lammasniemi and M. Otala, "Power Amplifier Design Parameters and Intermodulation Distortion at the Amplifier–Loudspeaker Interface," presented at the 65th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 28, p. 380 (1980 May), preprint 1608.

[17] R. R. Cordell, "Open-Loop Output Impedance and Interface Intermodulation Distortion in Audio Power Amplifiers," presented at the 64th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 27, p. 1022 (1979 Dec.), preprint 1537.

[18] R. H. Small, "Direct-Radiator Loudspeaker System Analysis," *J. Audio Eng. Soc.*, vol. 20, pp. 383–395 (1972 June).

[19] P. G. L. Mills and M. J. Hawksford, "Transient Analysis: A Design Tool in Loudspeaker Systems Engineering," presented at the 80th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 35, p. 386 (1986 May), preprint 2338.

[20] S. H. Linkwitz, "Loudspeaker System Design," *Wireless World*, pp. 79–83 (1978 Dec.).

[21] W. M. Leach, Jr., "Active Equalization of Closed-Box Loudspeaker Systems," *J. Audio Eng. Soc.*, vol. 29, pp. 405–407 (1981 June).

[22] K. E. Stahl, "Synthesis of Loudspeaker Mechanical Parameters by Electrical Means: A New Method for Controlling Low-Frequency Loudspeaker Behavior," *J. Audio Eng. Soc.*, pp. 587–596 (1981 Sept.).

[23] P. G. A. H. Voight, "Improvements in or Relating to Thermionic Amplifying Circuits for Telephony," UK Patent 231972 (1924 Jan.).

[24] H. Black, "Inventing the Negative Feedback Amplifier," *IEEE Spectrum*, pp. 55–60 (1977 Dec.).

[25] J. A. Klaassen and S. H. de Koning, "Motional Feedback with Loudspeakers," *Philips Tech. Rev.*, vol. 29, no. 5, pp. 148–157 (1968).

[26] D. de Greef and J. Vandewege, "Acceleration Feedback Loudspeaker," *Wireless World*, pp. 32–36 (1981 Sept.).

[27] G. J. Adams, "Adaptive Control of Loudspeaker Frequency Response at Low Frequencies," presented at the 73rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 361 (1983 May), preprint 1983.

[28] E. de Boer, "Theory of Motional Feedback," *IRE Trans. Audio*, pp. 15–21 (1961 Jan./Feb.).

[29] A. F. Sykes, "Damping Electrically Operated Vibration Devices," UK Patent 272622 (1926 Mar.).

[30] R. L. Tanner, "Improving Loudspeaker Response with Motional Feedback," *Electronics*, pp. 142 ff. (1951 Mar.).

[31] I. Martikainen, A. Varla, and M. Otala, "Input Current Requirements of High-Quality Loudspeaker Systems," presented at the 73rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 364 (1983 May), preprint 1987.

[32] D. Preis and J. Schroeter, "Peak Transient Current and Power into a Complex Impedance," presented at the 80th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, p. 386 (1986 May), preprint 2337.

[33] M. Otala and P. Huttunen, "Peak Current Requirement of Commercial Loudspeaker Systems," *J. Audio Eng. Soc.*, vol. 35, pp. 455–462 (1987 June).

[34] J. Vanderkooy and S. P. Lipshitz, "Computing Peak Currents into Loudspeakers," presented at the 81st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, pp. 1036–1037 (1986 Dec.), preprint 2411.

[35] M. J. Hawksford, "Distortion Correction in Audio Power Amplifiers," *J. Audio Eng. Soc.*, vol. 29 (*Engineering Reports*), pp. 27–30 (1981 Jan./Feb.).

[36] M. J. Hawksford, "Power Amplifier Output-Stage Design Incorporating Error-Feedback Correction with Current-Dumping Enhancement," presented at the 74th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 960 (1983 Dec.), preprint 1993.

[37] T. Holman, "New Factors in Power Amplifier Design," *J. Audio Eng. Soc. (Engineering Reports)*, pp. 517–522 (1981 July/Aug.).

[38] M. Huber, "Important Aspects of Power Amplifiers," *Studio Sound*, pp. 66–74 (1985 Nov.).

[39] H. D. Harwood, "Improvements Relating to Loudspeakers," UK Patent 1520156 (1976 Mar.).

[40] D. R. von Recklinghausen, "Dynamic Equalizer System for Loudspeakers," UK Patent 2050754A (1980 Jan.).

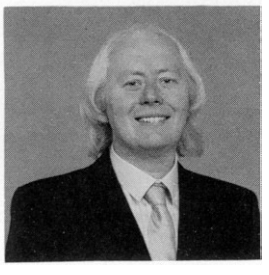
[41] R. M. Bews, "Digital Crossover Networks for Active Loudspeaker Systems," Ph.D. dissertation, University of Essex, UK (1987 Sept.).

THE AUTHORS

P. Mills

Paul Mills graduated with a B.Eng. degree in engineering science and industrial management from Liverpool University in 1980. After working for GEC in power and control systems engineering, he began a 3-year period of postgraduate study at the University of Essex in 1983, during which time he formulated the ideas presented in this paper. He then taught at Essex for two years in the Electronics Systems Engineering Department, while writing his Ph.D. thesis on active loudspeaker systems with transconductance amplification. In 1988 July he was appointed senior design engineer with Tannoy Limited. He is a member of the Audio Engineering Society.

•



M. O. Hawksford

Malcolm Omar Hawksford is a senior lecturer in the Department of Electronic Systems Engineering at the University of Essex, where his principal interests are in the fields of electronic circuit design and audio engineering. Dr. Hawksford studied at the University of Aston in Birmingham and gained both a First Class Honors B.Sc. and Ph.D. The Ph.D. program was supported by a BBC Research Scholarship, where the field of study was the application of delta modulation to color television and the development of a time compression/time multiplex system for combining luminance and chrominance signals.

Since his employment at Essex, he has established the Audio Research Group, where research on amplifier studies, digital signal processing, and loudspeaker systems has been undertaken. Since 1982 research into digital crossover systems has begun within the group and, more recently, oversampling and noise shaping investigated as a means of analog-to-digital/digital-to-analog conversion. Dr. Hawksford has had several AES publications that include topics on error correction in amplifiers and oversampling techniques. His supplementary activities include writing articles for *Hi-Fi News* and designing commercial audio equipment. He is a member of the IEE, a chartered engineer, a fellow of the AES, and a member of the review board of the *Journal*. He is also a technical adviser for *HFN* and *RR*.

Transconductance Power Amplifier Systems for Current-Driven Loudspeakers*

P. G. L. MILLS

Tannoy Limited, Coatbridge, Strathclyde ML5 4TF, UK

AND

M. O. J. HAWKSFORD

University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK

Moving-coil loudspeakers generally provide a substantial improvement in linearity when current driven, together with the elimination of voice-coil heating effects. Consequently there is a need to investigate low-distortion power amplifier topologies suitable for this purpose. After considering established current feedback approaches, a novel method using a common-base isolation stage is outlined and extended to show a prototype amplifier circuit in detail. In addition, the elements of a two-way active current-driven system are described, with low-frequency velocity feedback control derived from a sensing coil. The coupling error between this coil and the main driving coil is nulled by electronic compensation.

0 INTRODUCTION

The moving-coil drive unit can readily be shown to benefit in terms of linearity when controlled by a current source rather than the more conventional voltage source. Throughout this paper we will term this mode of operation *current drive*, whereby the amplifier source impedance can, to all intents and purposes, be considered infinite compared to the drive unit impedance.

Of the drive unit error mechanisms that can be countered by current drive, the voice-coil resistance is of particular interest. As a result of self-heating in excess of 200°C, the increase in coil resistance leads to sensitivity loss (often referred to as power compression [1], [2]), loss in electrical damping of the fundamental resonance, and crossover filter misalignment. In their paper Hsu et al. [3] concluded that a satisfactory method of compensating for the effect had yet to be found.

At higher frequencies, nonlinearity occurs as the coil inductance is modulated by movement in the magnetic circuit and by other effects such as magnetic hysteresis [4]. Measurements under current drive have shown, in comparison with voltage drive, a high-frequency distortion reduction of typically 20–30 dB for a bass–midrange drive unit.

These performance advantages arise from the coil resistance and inductance being totally eliminated from the system transfer function. The force on the cone is proportional to the voice-coil current, not the applied voltage. Analysis also shows a reduced dependence on nonlinearity within the force factor and mechanical impedance of the drive unit.

Thus as a result of the performance gains that can be demonstrated using current drive, there arises the need to investigate suitable power amplifier topologies to make the best of the technique. This paper therefore aims to review some of the earlier published work on transconductance amplifier design, while presenting new topologies and detailed circuitry of a two-way active prototype system. In addition, due to the loss of voice-coil damping under current drive, control circuitry for restoring damping by means of motional

* Manuscript received 1988 July 6. This paper expands on some areas covered by the authors in "Distortion Reduction in Moving-Coil Loudspeaker Systems Using Current-Drive Technology," volume 37, number 3 (1989 March).

feedback applied to the bass–midrange drive unit is described, along with the low-level crossover circuitry of the prototype system.

1 POWER AMPLIFIER TOPOLOGIES FOR CURRENT DRIVE

1.1 Review of Transconductance Amplifier Techniques

A transconductance power amplifier requires a high output impedance that is linear and frequency independent. It must also possess the attributes of a conventional voltage power amplifier such as high linearity, wide bandwidth, freedom from slewing-induced errors, and insensitivity to load variations (be they linear or nonlinear).

The most commonly used technique to obtain a high output impedance is to apply current feedback around a conventional power amplifier by means of a sensing resistor in the loudspeaker earth return [5], [6], as illustrated in Fig. 1. The transconductance g_m is defined

$$g_m = \frac{1}{R_f} . \quad (1)$$

The method has also been used in high-current industrial applications. There are two main disadvantages with such a system. First, the open-loop gain of the amplifier is frequency dependent as a result of the amplifier's dominant pole, and this is reflected in the output impedance. Second, the loudspeaker impedance, which is both frequency dependent and nonlinear, tends to modulate the transconductance of the amplifier. The fact that the load is not ground referenced may be considered inconvenient in some applications.

A refinement of the basic technique was described in Lewis [7]. The circuit was symmetrical in nature, using two current-sensing resistors, with a ground referenced load fed from MOSFET output devices. Good linearity was indicated at 10-W average power into a 5-Ω load. However due to class A operation, the design would be inefficient at the power levels necessary for a moving-coil drive unit (between 50 and 100 W typically). Care must be taken to minimize output offset current with this scheme.

A further ground-referenced current feedback scheme was described in Nedungadi [8], but this required the complexity of a differential voltage-to-current converter in conjunction with a floating sensing resistor in order to maintain current feedback.

Another technique used in implementing transconductance amplifiers involves the combination of supply current sensing around a follower, together with current mirrors feeding the load [9]. The arrangement is shown in Fig. 2, where R_f is a dummy load and the transconductance is again defined by Eq. (1). Operation of the circuit is typically in class AB.

While suitable for low output currents (<50 mA peak), the approach is difficult to extend to the levels required for driving a loudspeaker (typically 5 A peak or more) due to the linearity of the mirrors and also power loss in R_f. Although the mirrors could be arranged to provide current gain and could be partially linearized by error-correction techniques [10], the technique is not felt to offer a particularly practical solution.

1.2 Methods Using a Common-Base Isolation Stage

The approach devised to overcome the limitations cited as being inherent to existing topologies is illustrated in basic form by Fig. 3. The notable aspect of this strategy is the open-loop grounded base stage, which isolates the load Z_L from the main amplifier A_t while providing a naturally high output impedance without the use of overall current feedback. In addition a cascode configuration is formed in conjunction with the output devices in the main amplifier A_t. Resistor R_f defines the transconductance, driven from amplifier A_t, a voltage source, which may operate with low values of supply voltage $\pm V_{SI}$ to reduce power dissipation. The loudspeaker is referenced to ground and isolated from any feedback loop used to linearize the amplifier A_t. Although a successful prototype based on this scheme has been constructed, with amplifier A_t running in class A and with a class AB output stage, it is to some extent an uneconomical solution due to the need for two pairs of floating power supplies.

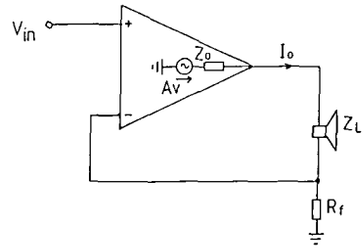


Fig. 1. Basic current-feedback-derived transconductance amplifier.

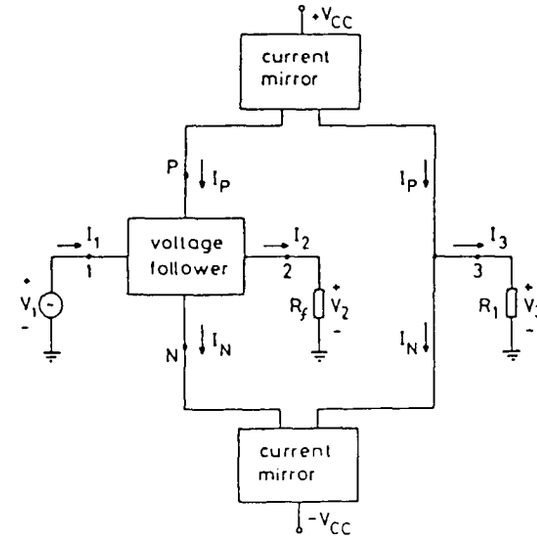


Fig. 2. Voltage-current converter. After Rao and Haslett [9].

A more viable alternative is the revised topology illustrated by Fig. 4. This circuit takes the form of a current amplifier of current gain

$$\alpha = -\frac{R_x}{R_f} \tag{2}$$

The first-stage power supply $\pm V_{S1}$ is ground referenced, unlike the previous case, meaning that several power amplifiers within an active system may share a common supply, thus reducing complexity and cost. Like the previous scheme, the current flowing in the transconductance defining resistor R_f is that which flows in the load Z_L, except for any base current lost to ground in the common-base stage. The fact that the amplifier A_i is referenced to the input of the common-base stage and not to ground tends to decouple it from any distortion appearing at the emitters of the common-base stage.

This topology forms the basis of the prototype system, the detailed circuitry of which is described in Sec. 2. On a practical note, it is important to provide adequate current gain in the common-base stage in order to prevent nonlinear current loss to ground, which introduces distortion.

1.3 Alternative Approaches

All of the circuits described so far rely on a current-sensing resistor to define the overall system transconductance. Even when this resistor is of a low value (about 1 Ω), it still tends to dissipate an appreciable amount of power. This element would at first seem to be fundamental to the design of a transconductance amplifier, but it is interesting to note the possibilities of transformer-derived feedback in perhaps reducing such losses.

Nordholt, in his classification of feedback configurations [11], described how transformer-derived feedback could be used to generate transconductance and current gain functions, as shown in Fig. 5. In Fig. 5(a) resistor R_f is still necessasry in order to define the stage transconductance.

Although no research has been directed in this area and the approach is only conceptual in nature, it may be worth further investigation, given a wide-bandwidth transformer design.

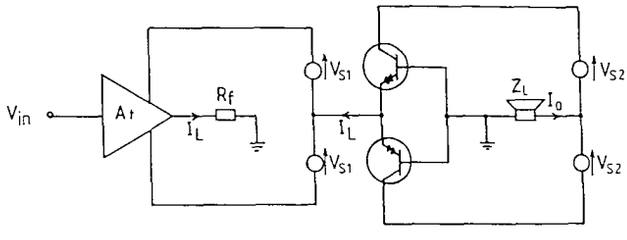


Fig. 3. Basic transconductance power amplifier using grounded-base output stage.

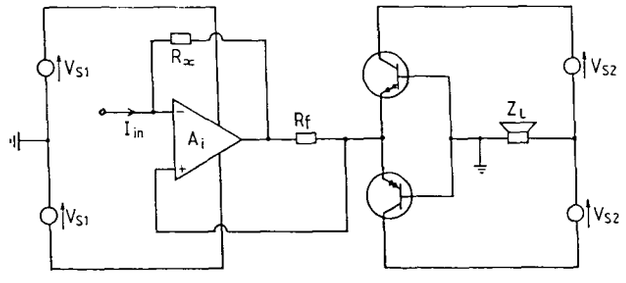


Fig. 4. Alternative configuration for current gain.

2 PROTOTYPE AMPLIFIER SYSTEM

2.1 General Overview

The two-way active loudspeaker system constructed to validate the basic approach proposed for high output impedance power amplifier design was based on the Celestion SL600 loudspeaker. In this section the current gain power amplifier is considered in detail along with the necessary transconductance preamplifier, while Sec. 3 considers the associated motional feedback control circuitry, which is required for the bass–midrange drive unit.

Throughout the design, the underlying philosophy has been to use symmetrical direct-coupled circuitry to give good transfer function linearity without recourse to high levels of overall negative feedback [12], [13]. DC stability is taken care of by servo amplifiers (feedback integrators).

2.2 Transconductance Preamplifier

Fig. 6 shows a two-stage design, the basic topology of which has often been used with overall feedback as

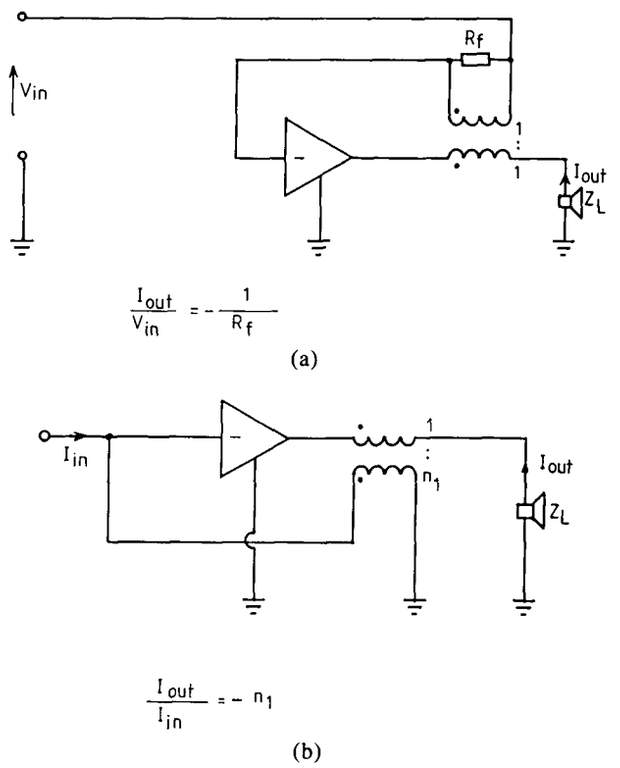


Fig. 5. Transformer-derived feedback systems. After Nordholt [11]. (a) Transconductance stage. (b) Current gain stage.

a voltage gain stage [14]. It is operated here open loop to provide a high output impedance and consequently must be capable of good linearity.

Transistor pairs Q_3/Q_4 and Q_5/Q_6 form cascodes to increase high-frequency linearity and give a high output impedance. Bias arrangements for the cascode are somewhat unusual in that resistors R_{14} and R_{15} are not returned to the supply rails, but are connected to the emitters of the common-emitter part of the cascode, thus avoiding nonlinearity from base current loss in the common-base devices [15]. This reduces high-frequency distortion by typically a factor of 10 at 20 kHz over the conventional bias method.

Operational amplifier IC_1 with associated passive components forms a current-sensing differential servo amplifier to null any output offset current due to imbalances in the main circuit and has no effect on performance within the audio band. This configuration of servo amplifier, to the authors' knowledge, has not been seen before in the literature.

At frequencies within the passband of the amplifier, the transconductance g_m may be approximated by the expression

$$g_m = \frac{I_{out}}{V_{in}} \approx \frac{R_6}{R_{11}(R_8/2 + R_{10})} . \quad (3)$$

With the component values shown, $g_m \approx 4$ mS.

In addition to the main input and output, an auxiliary velocity feedback input is provided along with an error-nulling output. These are only required for low-frequency use, and their function is described in Sec. 3 when considering the velocity feedback control circuitry.

2.3 Current Gain Power Amplifier

The current gain power amplifier, which accepts the output of the transconductance preamplifier, is based on the structure shown in Fig. 4. For the purpose of description, it is split into three sections: input amplifier, power follower, and common-base output stage. Both input amplifier and follower are represented by the gain block A_i in this simplified representation.

We consider first the input amplifier, Fig. 7. This is essentially the same topology as the transconductance preamplifier, but with a few refinements. Input stage biasing is performed with current sources based around transistors Q_1 and Q_2, instead of resistive biasing. This is a result of the need to provide immunity to the greater level of supply rail contamination caused by class AB operation of the power follower stage. The output from the transconductance preamplifier is fed to the emitters of the input devices Q_3 and Q_4, which thus operate in common-base mode. The first and second stages of the amplifier are coupled together by current mirror pairs Q_5/Q_8 and Q_9/Q_{12} to reduce loading effects and interaction between the two stages. These mirrors are themselves linearized by local error feedback correction consisting of transistor pairs Q_6/Q_7 and Q_{10}/Q_{11}. This approach has been previously documented [10], al-

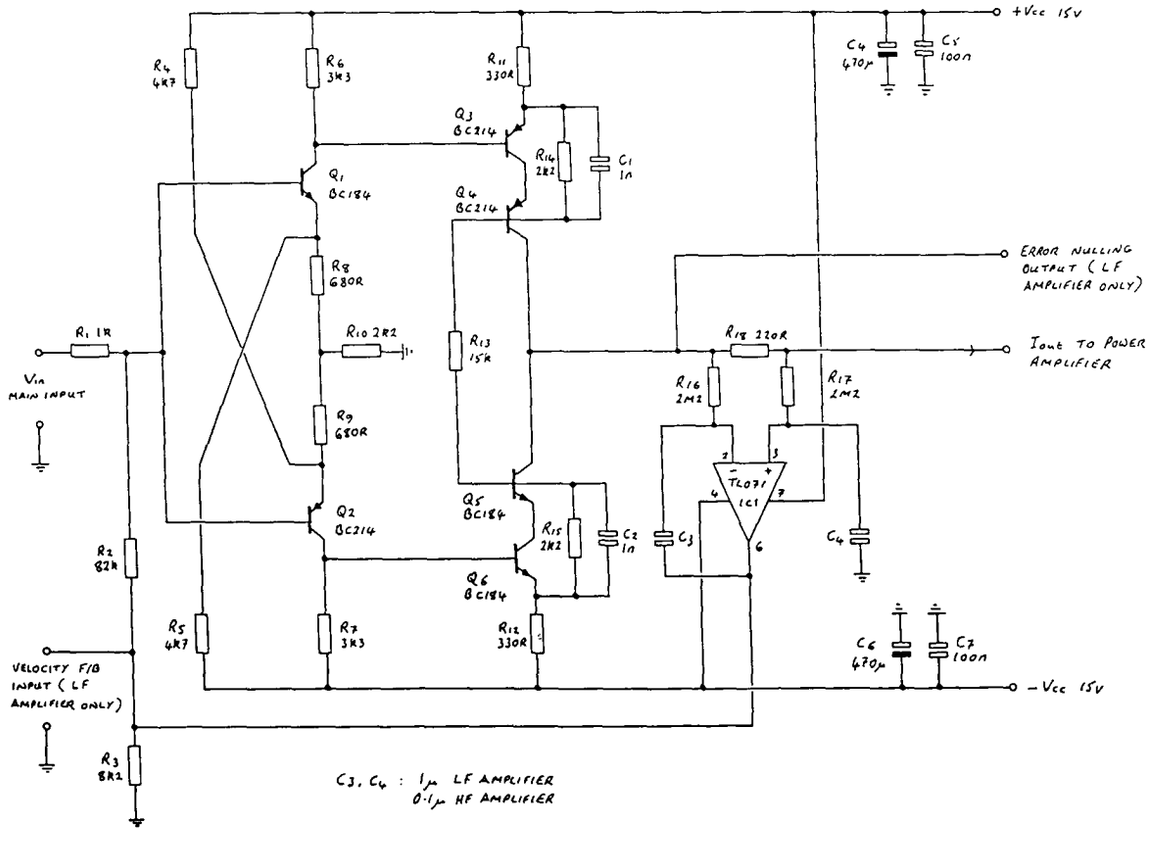


Fig. 6. Transconductance preamplifier.

though in this case some current gain has been introduced into the mirrors to enable correct quiescent operating conditions to be established in the first and second gain stages.

The outputs from the cascode pairs Q_8/Q_{13} and Q_{12}/Q_{14} are displaced ± 4 V about ground by green LEDs D_1–D_4, in order to bias the next stage. Resistor R_{28} and capacitor C_3 are included to define the open-loop gain characteristics of the amplifier, to ensure that stability is maintained under closed-loop conditions.

Fig. 8 shows the next section, which is a follower with extensive error-correction circuitry and is essentially similar to a previously published topology [16], but with improvements to biasing arrangements. It is worth briefly reviewing the principle of operation.

Transistor pairs Q_{16}/Q_{18} and Q_{17}/Q_{19} form a Darlington follower, preventing loading of the previous stage and driving the Darlington output devices Q_{30} and Q_{31}. Transistors Q_{28} and Q_{29} form V_{be} multipliers to bias the output Darlingtons, but are also configured as error amplifiers, which together with Q_{22} and Q_{23} form the main error feedback loop, delivering a correction current through resistors R_{38} and R_{78} in response to any nonlinearity in the output devices Q_{30} and Q_{31}. R_{48} is included as an adjustment to achieve the best distortion null.

In order to linearize Q_{18} and Q_{19}, which have to drive the output Darlingtons, additional error correction in the form of feedforward is applied with the aid of Q_{20} and Q_{21}, in combination with the input transistors Q_{16} and Q_{17}. Further linearization is achieved by current mirror transistors Q_{25} and Q_{26}, which form a negative feedback loop, thus reducing the source impedance seen by output Darlingtons Q_{30} and Q_{31}.

Moving now to Fig. 9, which shows the output common-base stage, the preceding follower drives current through resistor Q_{67}, which in conjunction with R_2 (Fig. 7) sets the midband current gain of the complete amplifier to around 800. Inductor L_2 serves to reduce the high-frequency current gain of the amplifier to ensure stability. The current in R_{67} flows into the common-base output stage, consisting of Darlingtons Q_{32} and Q_{33}, along with driver devices Q_{36} and Q_{37}, the bases of which are referenced to ground. Except for any current loss to ground, such as through the bases of these devices and through the biasing current sources (Q_{34}, Q_{35}), the current in R_{67} flows through the load via floating power supplies $\pm V_{CC2}$.

In order to establish a low-output offset current for the amplifier (typically less than ± 2 mA), a servo based around IC_1 and referenced to the input of the common-base stage, is used to feed a dc compensation current back to the input of the amplifier.

To prevent switch-on and switch-off transients from reaching the load, relay RL_1 is included, controlled by a time-delay circuit on startup and almost instantaneously dropping out on power down. The control circuitry to perform this function is not shown.

The power amplifier together with the transconductance preamplifier was evaluated in terms of standard

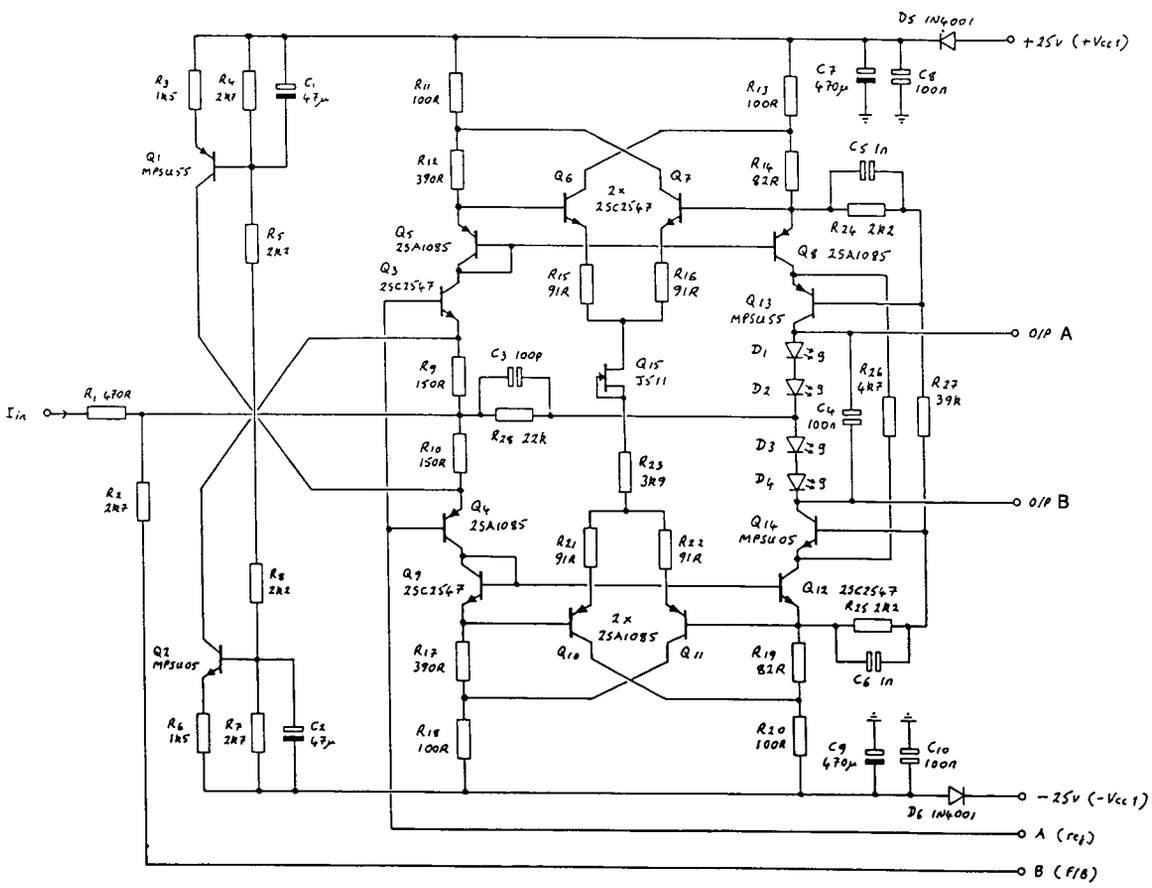


Fig. 7. Prototype amplifier, input stage.

measurements and found to be comparable with a typical high-performance conventional amplifier. The results are as follows:

Rated power output (8-Ω resistive load)	75 W average
Total harmonic distortion at rated power	
20 Hz	−79 dB
1 kHz	−86 dB
20 kHz	−68 dB
Intermodulation distortion (19 and 20 kHz at equal levels at rated power)	−86 dB
Hum and noise (re maximum output)	−90 dB
Small-signal bandwidth, −3 dB	0.1 Hz to 50 kHz
Output impedance*	
20 Hz	4.1 MΩ
1 kHz	106 kΩ
20 kHz	11.4 kΩ

* From computer simulation, due to the difficulty in performing these measurements.

It is interesting to note that the distortion measurements may only easily be made indirectly by converting the output current to a voltage, by means of a resistive load bank. The measurements as shown will thus reflect any nonlinearity in the load.

The protective features, consisting of output fuses and relay contact, should not introduce any degradation in performance, as they are in series with a high source impedance, which is not the case with a conventional power amplifier.

3 VELOCITY FEEDBACK CONTROL SYSTEM

3.1 Outline Approach

In order to compensate for the loss in electric damping of the bass–midrange unit caused by the high amplifier output impedance, velocity feedback was used to restore damping [5], [6]. While many forms of sensing arrangement have been described ([17]–[23], for example), the method adopted here is attractive for reasons of mechanical simplicity and cost effectiveness. The technique used is to wind a sensing coil over the main voice coil of the drive unit. The output voltage of the sensing coil will ideally be defined by

$$V_s = (Bl)_s u \tag{4}$$

where $(Bl)_s$ is the sensing coil Bl product, N/A, and u is the cone velcity, m/s.

Unfortunately an error is induced in the sensing coil by transformer action from the main driving coil. In the previously documented work induced errors were overcome by neutralizing coils or by an altogether more elaborate mechanical arrangement to physically isolate the driving and sensing coils. With the approach considered here, a procedure of electronic compensation has been chosen in order to avoid expensive tooling costs for a specialized drive unit.

The physical arrangement of the assembly is shown in Fig. 10. It should be noted that in this case, the sensing coil follows roughly the same Bl profile as the main driving coil, so the action of velocity feedback

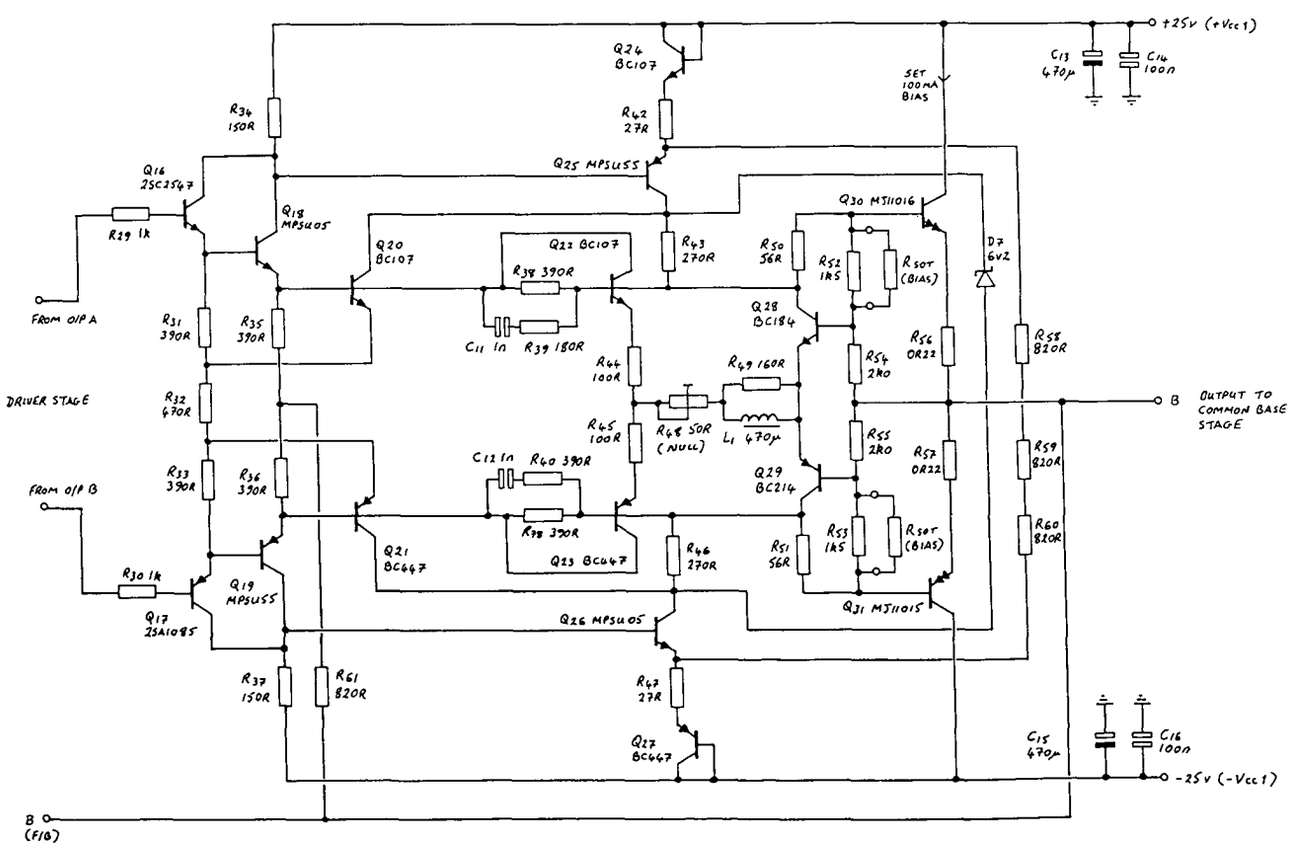


Fig. 8. Prototype amplifier, follower stage.

does not improve linearity above that already afforded by the current drive.

If a longer sensing coil could be accommodated (or indeed a very short coil that remained well within the magnet gap), a further reduction in distortion would be possible.

3.2 Coupling Error Compensation

In order to investigate the nature of the transformer coupling error, Fig. 11 shows the error magnitude with respect to frequency for the coil assembly at equilibrium and also at both extremes of travel. For this measurement the driving coil was powered from the prototype transconductance amplifier system. The level of error is seen to be frequency dependent, rising initially at a rate of approximately 4.6 dB/octave. This unusual characteristic is considered to be a function of pole-piece coupling with the magnetic circuit, but a full analysis of the mechanisms at work has not been undertaken. In addition, some positional dependence of the error magnitude is also apparent. At 100 Hz the coupled error is around 15 dB below the voltage appearing on the driving coil, thus illustrating the need for an effective compensation system.

To implement the compensator, it is necessary to derive a signal proportional to the current in the driving coil and to subject this signal to the same frequency dependence as the error mechanism itself in order to null the error from the sensing coil output. The variation in error level with displacement (typically ± 3 dB) has not been accounted for.

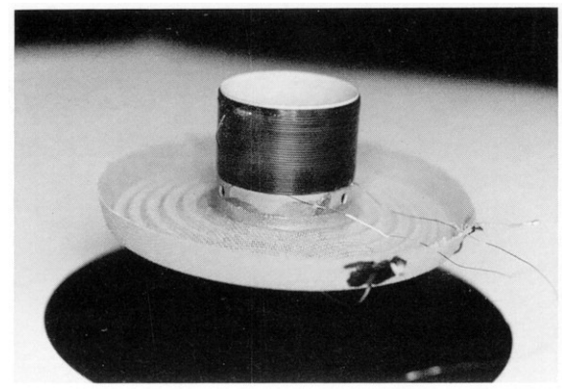


Fig. 10. Sensing-coil assembly.

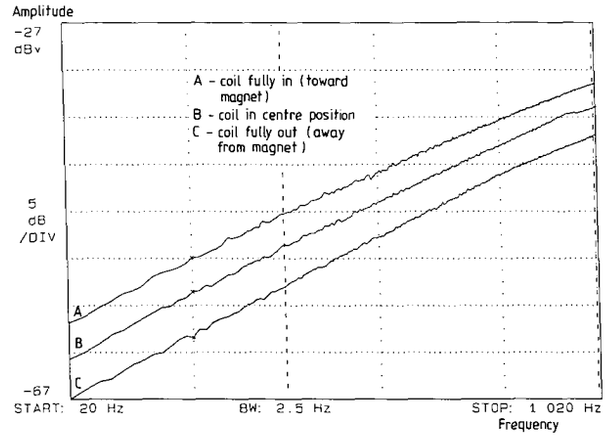


Fig. 11. Measured transformer coupling error.

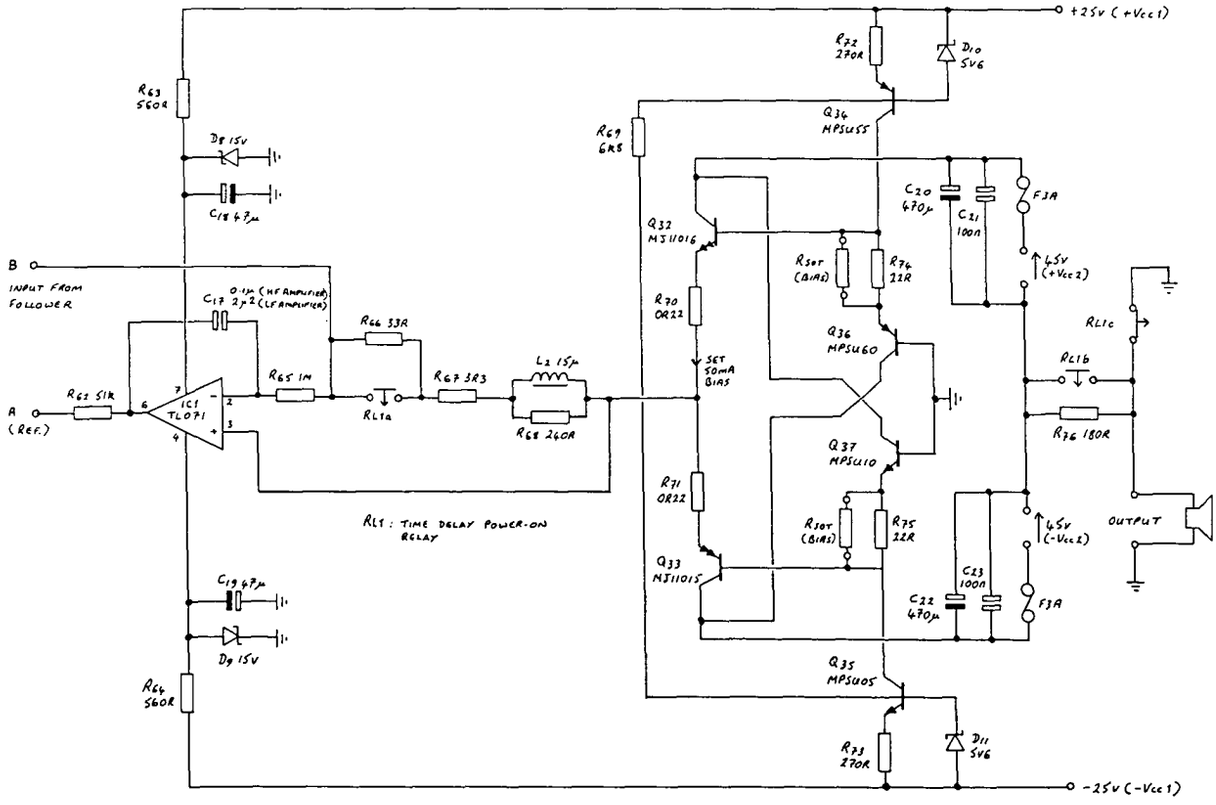


Fig. 9. Prototype amplifier, common-base stage. Protection relay contacts RL_1 are shown with amplifier shut down.

Fig. 12 illustrates the general approach to synthesis of the frequency-dependent element of the compensator. A number of first-order sections are combined, with pole–zero locations set to produce a slope approximating that desired. The general circuit configuration to give this response is shown in Fig. 13 for an nth-order compensator. The transfer function of this circuit is written

$$\frac{V_{out}}{V_{in}} = -\sum_{r=0}^{n}\left(\frac{j\omega RC_r}{1 + j\omega R_r C_r}\right) . \quad (5)$$

A software optimization routine was used to select component values in order to match the 4.6-dB/octave slope required. For a 6th-order compensator, the computer-predicted frequency response is shown in Fig. 14, which also lists the nearest preferred value component values chosen. The result is deemed more than adequate for our purposes, bearing in mind that some positional dependence of the coupling error is present, together with a gradual deviation from the idealized 4.6-dB/octave response with increasing frequency.

3.3 Complete Control System

We continue by considering the complete velocity feedback control system shown in Fig. 15. The sensing coil (source impedance 28 Ω) is connected to a high input impedance buffer stage IC_{2a} via an attenuator network to avoid overload. IC_{2b} forms a summing amplifier in order to subtract the signal derived from the coupling error compensator.

The compensator input is differential, accepting the voltage across the servo current-sensing resistor R_{18} of the transconductance preamplifier (Fig. 6). Thus the input to the compensator is proportional to the drive unit current. This differential signal is converted to single-ended format before the 4.6-dB/octave weighting is applied by the circuitry based around IC_{1d}. R_{22} provides an adjustment to enable the best error null to be obtained with a static motor coil assembly connected to the velocity feedback input.

In order to maintain stability of the closed-loop system, a second-order low-pass filter at 500 Hz is included in the feedback control loop. This also has the benefit of reducing any residual transformer coupling error at high frequencies, where the compensator is no longer as effective due to the changing slope of the error. Finally the output of the controller is summed with the main signal at the velocity feedback input of the transconductance preamplifier, with R_{27} (Fig. 15) providing an adjustment of the low-frequency Q alignment.

To illustrate the performance of the velocity feedback control system, a number of frequency and time domain measurements were obtained. First, Fig. 16(a) shows

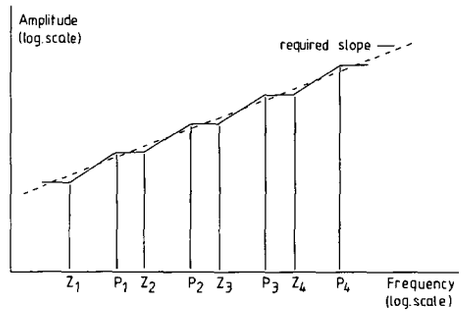


Fig. 12. Basis for synthesis of coupling error compensator.

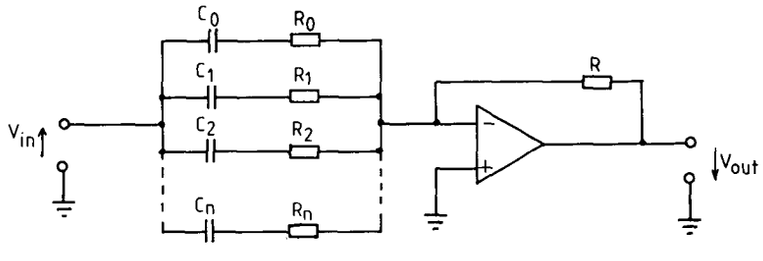


Fig. 13. General configuration for nth-order compensator.

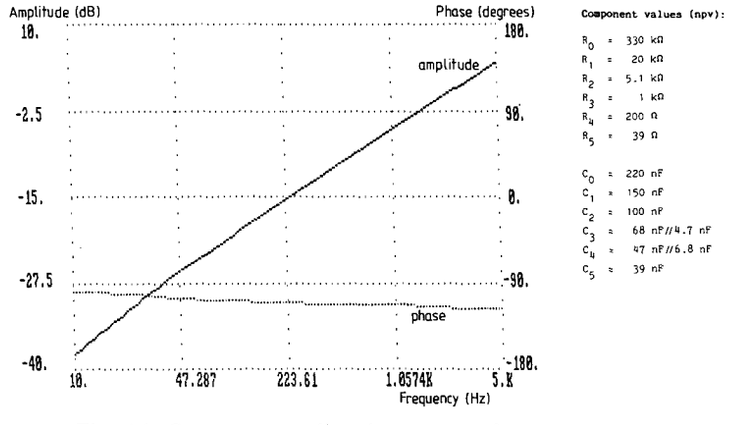


Fig. 14. Computer-predicted response of compensator.

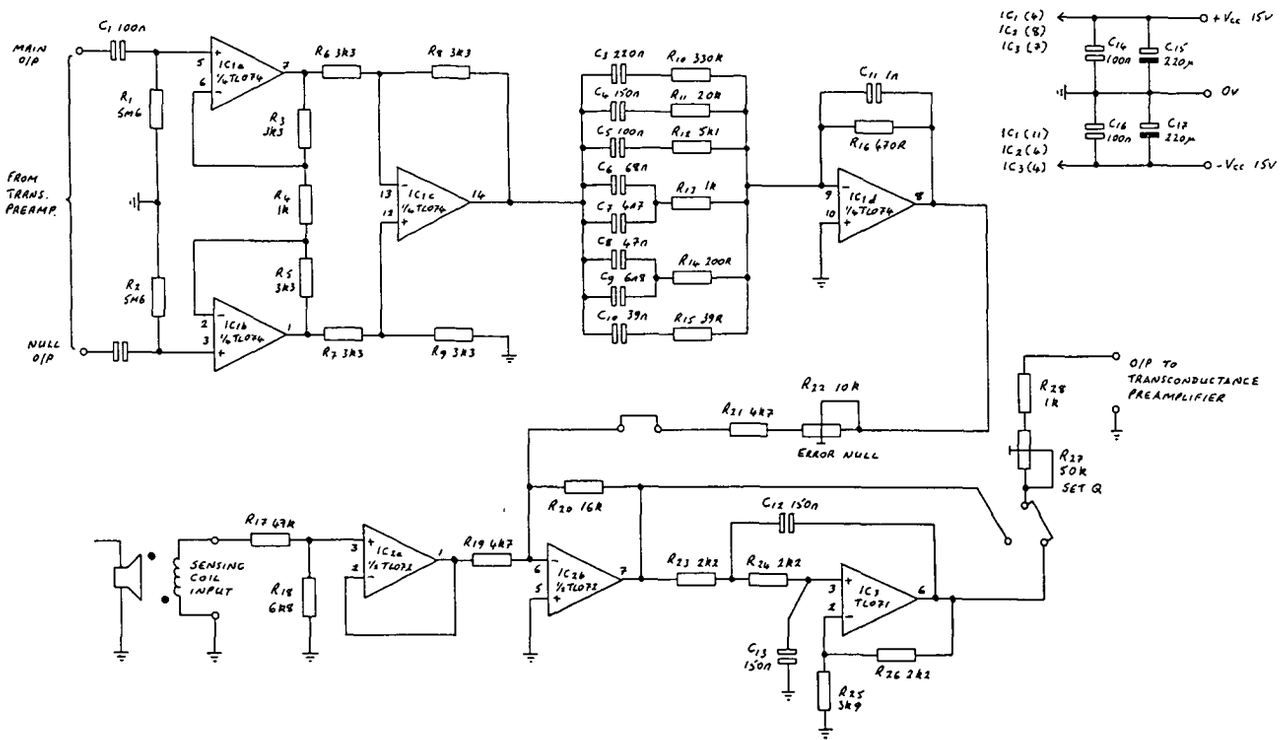


Fig. 15. Velocity feedback control system.

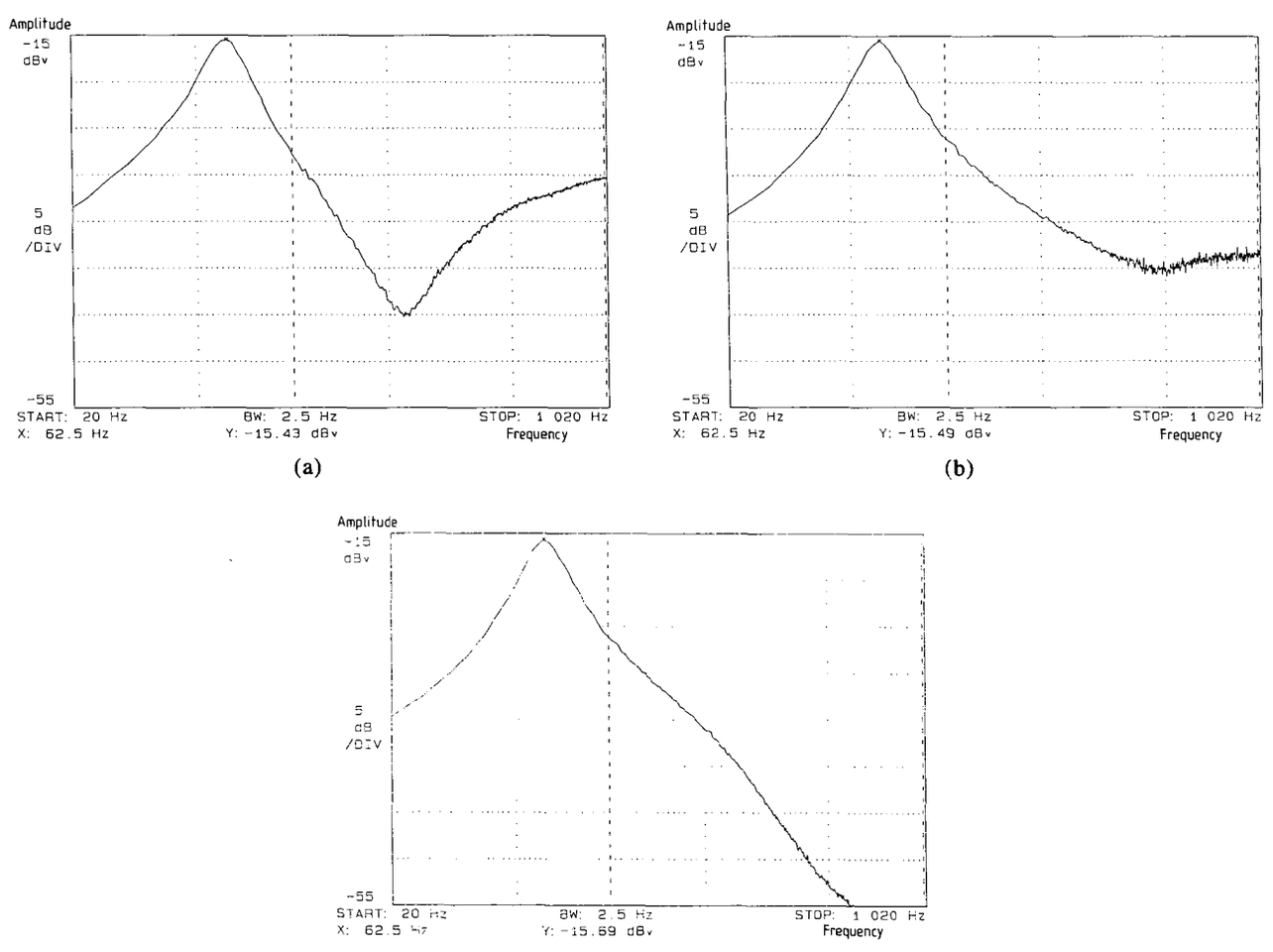


Fig. 16. Velocity feedback control signal. (a) Sensing-coil output voltage. (b) Addition of compensator. (c) Addition of compensator and low-pass filter.

the output of the sensing coil, corresponding to velocity, with frequency. The peak at 62.5 Hz corresponds to the drive unit–enclosure fundamental resonance, while the rising high-frequency output is due to the coupling error between drive and sensing coils. Fig. 16(b) shows the addition of the coupling error compensator, giving a much reduced spurious high-frequency output. The further addition of the second-order low-pass filter at 500 Hz gives the response of Fig. 16(c), which is close to an idealized velocity function.

Steady-state sine-wave measurements of the acoustic output suggest a worthwhile improvement in linearity of the bass–midrange drive unit compared to voltage drive. The following acoustic distortion measurements at a drive current of 1 A peak are illustrative:

	Voltage Drive (dB)	Current Drive* (dB)
Total harmonic distortion at 100 Hz re fundamental	−34.1	−43.3
Total harmonic distortion at 3 kHz re fundamental	−28.4	−55.0

* Under closed-loop conditions.

The effectiveness of the coupling error compensator and filter is confirmed by the fact that no increase in harmonic distortion is measurable up to 3 kHz and beyond (that is, over the full operating range of the drive unit) when the feedback loop is closed. The actual increase in distortion level present on the unfiltered and uncompensated velocity signal, compared to the drive unit acoustic output, ranges from 11 to 23 dB as frequency is increased from 500 Hz to 3 kHz.

The ability to vary the system Q with the velocity feedback control circuit is shown by means of near-field acoustic step response measurements. Fig. 17(a) is without the velocity feedback operational, showing a Q of around 2.5, which is the natural mechanical Q of the drive unit. Fig. 17(b)–(e) shows compensated Q alignments of 1.5, 1.0, 0.7, and 0.5, respectively. A value of $Q = 0.7$ preserves the low-frequency characteristics of the unmodified loudspeaker under voltage drive.

4 LOW-LEVEL CROSSOVER

To complete the two-way prototype system, a second-order low-level high- and low-pass crossover was included to integrate the drive units together, with a nominal crossover point of 3 kHz.

The crossover is implemented by passive RC elements, the time constants of which are individually adjustable to give the flattest frequency response, with buffer amplifiers between stages, to avoid loading effects. The complete system is shown in modular form by Fig. 18. After the input level control, amplifier A_1 provides a low-impedance drive to the first low- and high-pass filters, which are buffered by amplifiers A_2 and A_3 before the second set of filter sections. A_4 and A_5 are the transconductance preamplifiers previously

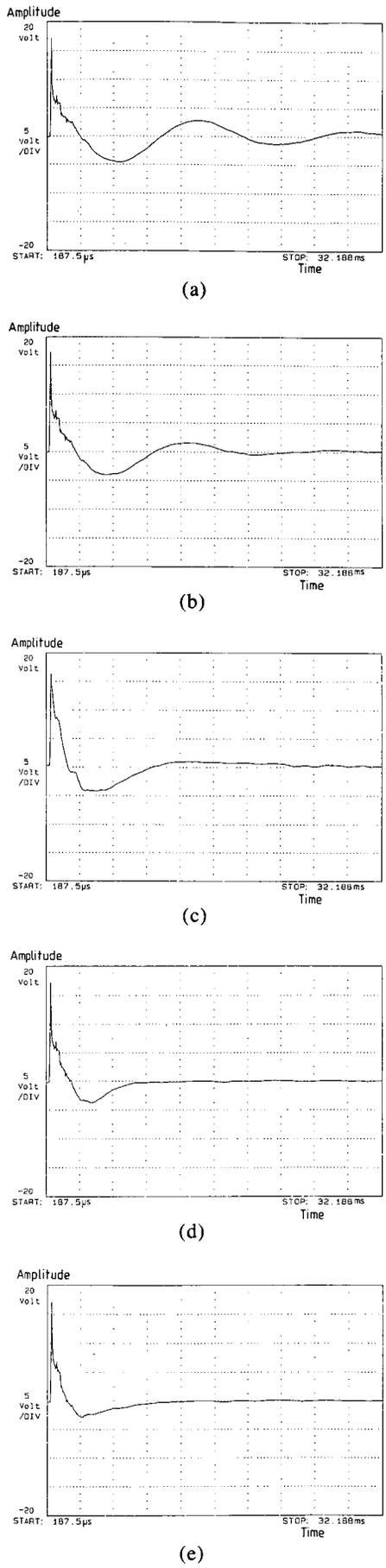


Fig. 17. Measured step responses of bass–midrange drive unit. (a) $Q = 2.5$ (no feedback). (b) $Q = 1.5$. (c) $Q = 1.0$. (d) $Q = 0.7$. (e) $Q = 0.5$.

described, which drive the high- and low-frequency power amplifiers, respectively.

The circuit topology of buffer amplifiers A_1, A_2, and A_3 (Fig. 19) is similar to the transconductance preamplifier, but with the addition of an output follower and overall negative feedback to provide a low output impedance. Certain gain and frequency response defining components are specific to individual amplifier stages as indicated.

The performance of the system is shown by the in-room measured frequency response curve of Fig. 20, with the measurement microphone 1 m on axis. The uneven high-frequency response is a function of the tweeter characteristics, with the resonant peak near 19 kHz being due to the first bending mode resonance of the copper dome. There is no discernible frequency response deviation in moving from voltage drive to current drive with this device, due to its high level of intrinsic damping.

While the low-frequency drive unit benefits sub-

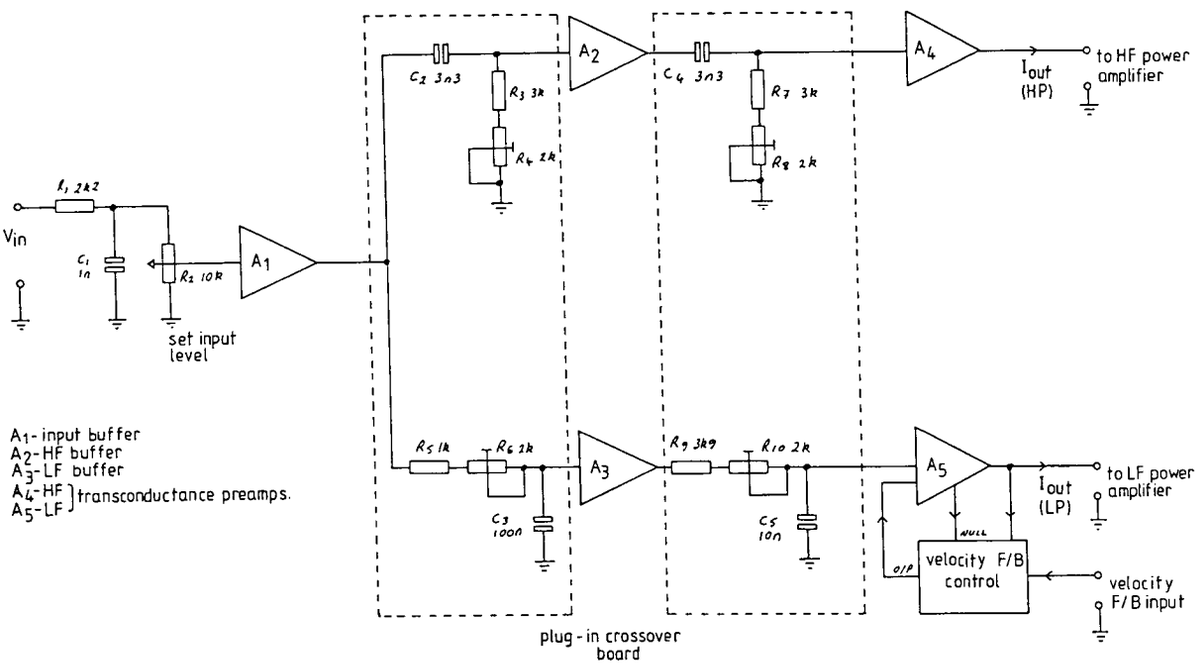


Fig. 18. Block diagram of low-level crossover.

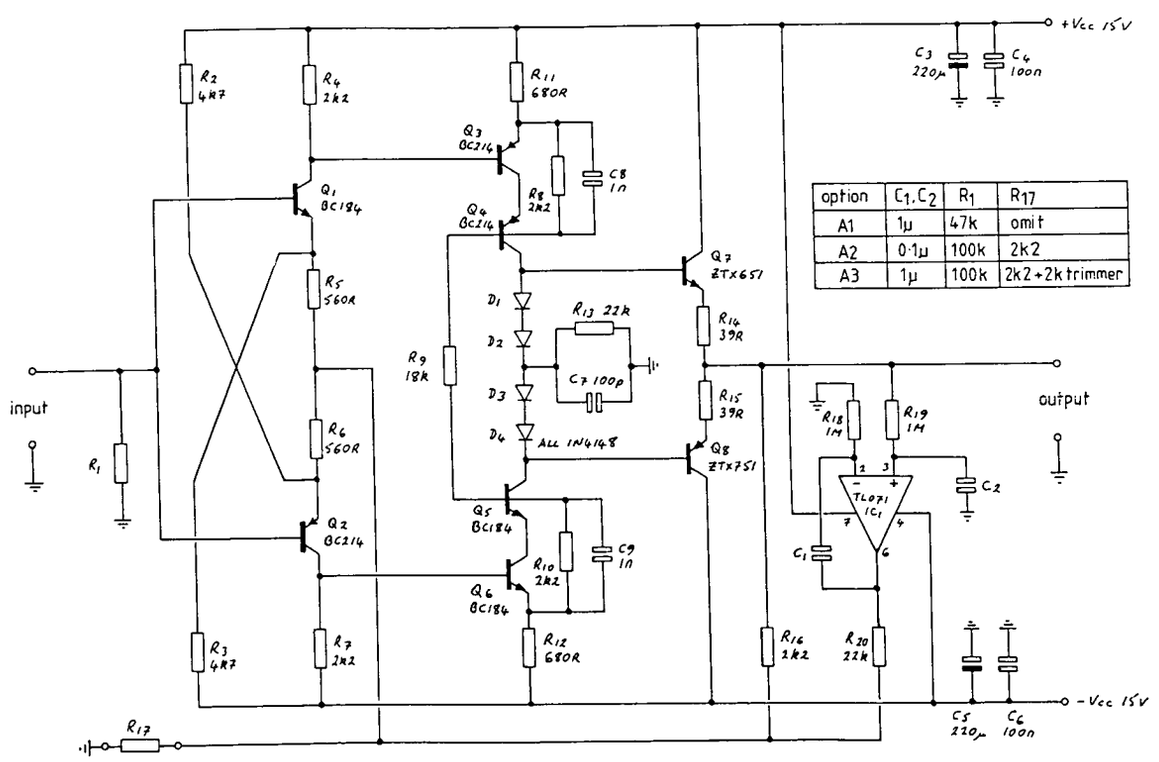


Fig. 19. Buffer amplifiers for crossover network.

stantially from current drive, the improvements in linearity to the tweeter are more modest, largely as a result of a more linear magnetic circuit and lower cone displacements. A distortion improvement of typically 3–7 dB is afforded. However, benefits in terms of the elimination of thermally induced errors are still apparent.

5 DYNAMIC CURRENT AND VOLTAGE DEMAND

Under certain signal conditions, drive units and loudspeaker systems have, under voltage drive, been shown to exhibit an instantaneous impedance modulus lower than might initially be suggested from the steady-state impedance characteristics, thus stressing the power amplifier in terms of current delivery [24]–[28]. In this section we consider the implications of this work in relation to current drive.

As an example, the bass–midrange unit and enclosure combination is considered. The equivalent electrical model is shown in Fig. 21. Under voltage drive, a pulse is applied to the drive unit, the duration of which is set to excite the large negative-going current excursion shown in Fig. 22. The voltage signal has been second-order low-pass filtered at 3 kHz to represent realistic operating conditions. At the point of maximum negative-going current, the instantaneous impedance modulus is 4.15 Ω, lower than the steady-state minimum of 7 Ω.

Under current drive it is only realistic to consider the drive unit when equalized by velocity feedback to give the same Q at fundamental resonance as in the voltage-driven case ($Q \approx 0.7$). Under these conditions, also with a 3-kHz crossover point, the current waveform is seen to be similar to the voltage-driven case, while the voltage waveform shows peaks due to the voice-coil inductance (Fig. 23). The instantaneous impedance modulus is similar to that under voltage drive, but slightly lower at 4.05 Ω.

The main significance of these results is that while the power amplifier is similarly stressed under both voltage and current drive, allowance must be made for sufficient headroom in the power amplifier for the voltage peak resulting from the coil inductance. The problem is worsened by voice-coil heating. The effect of a temperature rise of 200°C (meaning that the coil resistance increases to 13 Ω using copper) is shown by the waveforms of Fig. 24. While the current waveform is identical to that at normal temperature, as expected, the voltage waveform is increased in magnitude in order to keep the current constant. The negative-going excursion is seen to be 1.7 times greater. Although the performance of the drive unit is unaffected by the in-

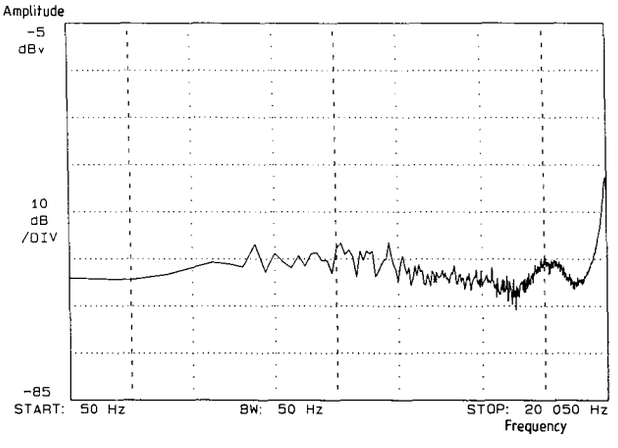


Fig. 20. Measured frequency response of complete system.

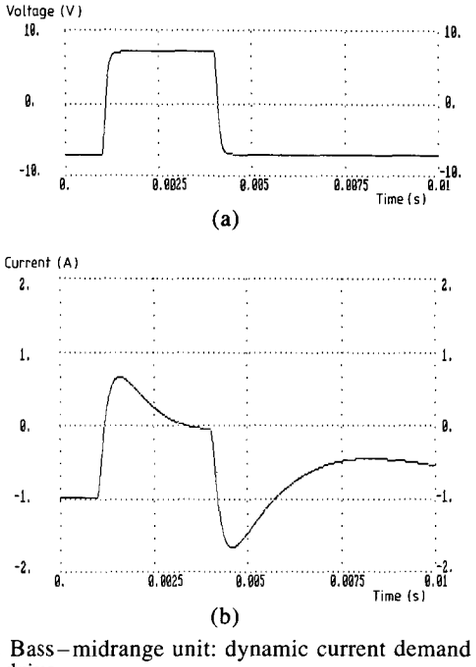


Fig. 22. Bass–midrange unit: dynamic current demand under voltage drive.

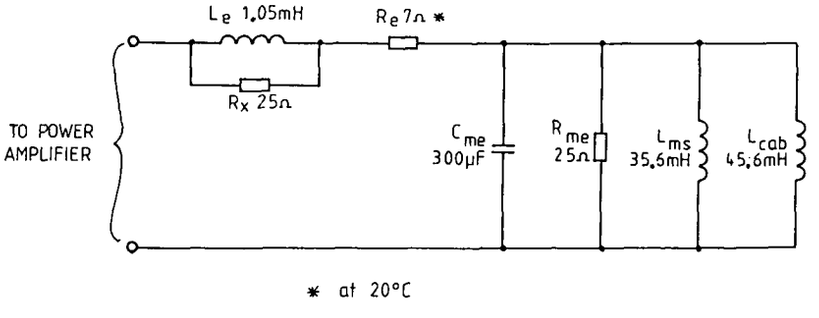


Fig. 21. Electrical model of bass–midrange drive unit and enclosure combination. L_e—voice-coil inductance; R_x—losses due to pole piece coupling; R_e—voice-coil resistance; C_{me}—capacitance due to moving mass; R_{me}—resistance due to mechanical losses; L_{ms}—inductance due to suspension compliance; L_{cab}—inductance due to volume of air in cabinet.

creased temperature, care must be taken in selecting the supply voltages for the power amplifier to avoid voltage clipping, with implications regarding the safe operating area of power devices.

6 CONCLUSIONS

In this paper we have considered a prototype two-way active current-driven loudspeaker system. The main benefits of current drive are seen to be a freedom from thermally induced distortion effects and also high-frequency nonlinearity caused by the voice-coil inductance.

By reviewing earlier work on transconductance power amplifiers, the limitations of these approaches were noted and a novel topology described, employing a common-base isolating stage to drive the loudspeaker. This decouples the nonlinear load from the feedback loop of the amplifier and provides a naturally high output impedance. For low-frequency operation, motional feedback is acknowledged as providing a suitable method of damping the drive unit fundamental resonance. A velocity-sensing coil was employed for this purpose, wound over the main driving coil, with electronic compensation used to null the transformer coupling error between the two. At high frequencies, where motional feedback is no longer feasible, a drive unit with good self-damping properties is recommended, although open-loop compensation could also be considered.

While the power amplifier circuits presented may be regarded as being somewhat complicated, they demonstrate some of the earlier work carried out within the group on distortion correction techniques. The availability of new hybrid gain stage devices (such as the Deltec DH-OA32) with high open-loop bandwidth and good output voltage and current-driving capabilities considerably simplifies the task of power amplifier design.

The provision of velocity feedback control circuitry introduces additional complexity over voltage drive, but this must be considered the price to be paid for what is regarded as a most worthwhile improvement in loudspeaker performance.

7 REFERENCES

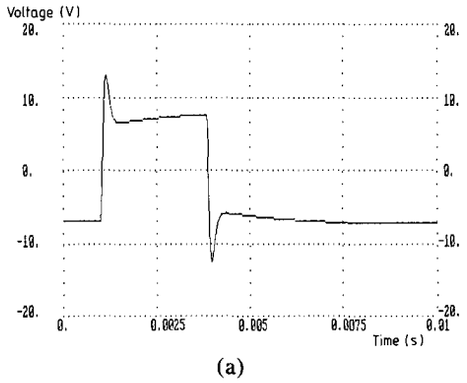
[1] M. R. Gander, "Dynamic Linearity and Power Compression in Moving-Coil Loudspeakers," presented at the 76th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 32, pp. 1008–1009 (1984 Dec.), preprint 2128.

[2] J. R. Gillion, P. L. Boliver, and L. C. Boliver, "Design Problems of High-Level Cone Loudspeakers," *J. Audio Eng. Soc. (Project Notes/Engineering Briefs)*, vol. 25, pp. 294–299 (1977 May).

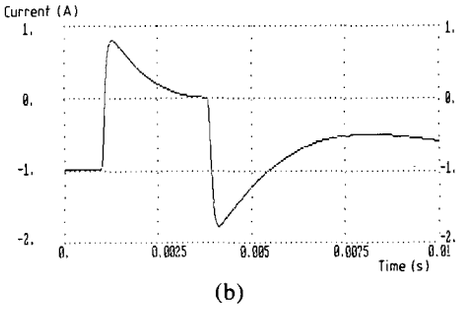
[3] T. S. Hsu, S. H. Tang, and P. S. Hsu, "Electromagnetic Damping of High-Power Loudspeakers," presented at the 79th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 33, p. 1011 (1985 Dec.), preprint 2297.

[4] W. J. Cunningham, "Nonlinear Distortion in Dynamic Loudspeakers Due to Magnetic Effects," *J. Acoust. Soc. Am.*, vol. 21, pp. 202–207 (1949 May).

[5] J. A. M. Catrysse, "On the Design of Some

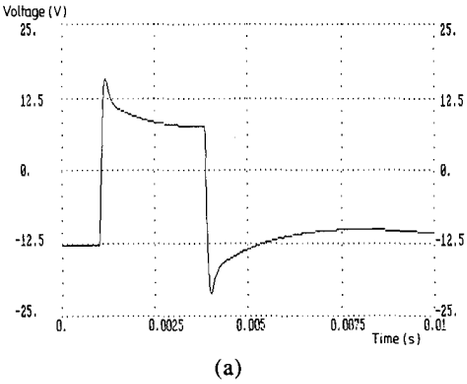


(a)

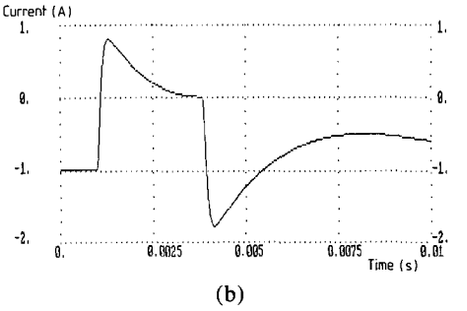


(b)

Fig. 23. Dynamic excitation under current drive.



(a)



(b)

Fig. 24. Dynamic excitation as for Fig. 23, but voice-coil temperature raised to 200°C (R_e = 13 Ω).

Feedback Circuits for Loudspeakers," presented at the 73rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 364 (1983 May), preprint 1964.

[6] R. A. Greiner and T. M. Sims, Jr., "Loudspeaker Distortion Reduction," *J. Audio Eng. Soc.*, vol. 32, pp. 956–963 (1984 Dec.).

[7] K. Lewis, "Transconductance Amplifiers," *Electron. Wireless World*, pp. 580–582 (1987 June).

[8] A. Nedungadi, "High Current Class AB Converter Technique," *Electron. Lett.*, vol. 16, pp. 418–419 (1980 May).

[9] M. K. N. Rao and J. W. Haslett, "Class AB Voltage-Current Converter," *Electron. Lett.*, vol. 14, pp. 762–764 (1978 Nov.).

[10] M. J. Hawksford, "Low-Distortion Programmable Gain Cell Using Current-Steering Cascode Topology," *J. Audio Eng. Soc.*, vol. 30, pp. 795–799 (1982 Nov.).

[11] E. H. Nordholt, *Design of High-Performance Negative-Feedback Amplifiers* (Elsevier, Amsterdam, 1983).

[12] J. J. Davidson, "A Low-Noise Transistorized Tape Playback Amplifier," *J. Audio Eng. Soc.*, vol. 13, pp. 2–16 (1965 Jan.).

[13] J. L. Linsley Hood, "Symmetry in Audio Amplifier Circuitry," *Electron. Wireless World*, pp. 31–34 (1985 Jan.).

[14] R. N. Marsh, "A Passively Equalised Phono Pre-amplifier," *Audio Amateur*, no. 3, pp. 18 (1980).

[15] M. Hawksford, "Reduction of Transistor Slope Impedance Dependent Distortion in Large-Signal Amplifiers," *J. Audio Eng. Soc.*, vol. 36, pp. 213–222 (1988 Apr.).

[16] M. J. Hawksford, "Power Amplifier Output-Stage Design Incorporating Error-Feedback Correction with Current-Dumping Enhancement," presented at the 74th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 960 (1983 Dec.), preprint 1993.

[17] P. G. A. H. Voight, "Improvements in or Relating to Thermionic Amplifying Circuits for Telephony," *UK patent 231972* (1924 Jan.).

[18] J. A. Klaassen and S. H. de Koning, "Motional Feedback with Loudspeakers," *Philips Tech. Rev.*, vol. 29, pp. 148–157 (1968).

[19] D. de Greff and J. Vandewege, "Acceleration Feedback Loudspeaker," *Wireless World*, pp. 32–36 (1981 Sept.).

[20] G. J. Adams, "Adaptive Control of Loudspeaker Frequency Response at Low Frequencies," presented at the 73rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 364 (1983 May), preprint 1983.

[21] E. de Boer, "Theory of Motional Feedback," *IRE Trans. Audio*, pp. 15–21 (1961 Jan./Feb.).

[22] A. F. Sykes, "Damping Electrically Operated Vibration Devices," *UK patent 272622* (1926 Mar.).

[23] R. L. Tanner, "Improving Loudspeaker Response with Motional Feedback," *Electronics*, pp. 142 ff. (1951 Mar.).

[24] P. G. L. Mills and M. J. Hawksford, "Transient Analysis: A Design Tool in Loudspeaker Systems Engineering," presented at the 80th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, p. 386 (1986 May), preprint 2338.

[25] I. Martikainen, A. Varla, and M. Otala, "Input Current Requirements of High-Quality Loudspeaker Systems," presented at the 73rd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 31, p. 364 (1983 May), preprint 1987.

[26] M. Otala and P. Huttunen, "Peak Current Requirement of Commercial Loudspeaker Systems," *J. Audio Eng. Soc.*, vol. 35, pp. 455–462 (1987 June).

[27] D. Preis, "Peak Transient Current and Power into a Complex Impedance," presented at the 80th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, p. 386 (1986 May), preprint 2337.

[28] J. Vanderkooy and S. P. Lipshitz, "Computing Peak Currents into Loudspeakers," presented at the 81st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, pp. 1036–1037 (1986 Dec.), preprint 2411.

Biographies for Drs. Mills and Hawksford were published in the March issue.

Differential-current derived feedback in error-correcting audio amplifier applications

M.O.J. Hawksford

Indexing terms: Feedback, Audio

Abstract: A dominant distortion mechanism in audio power amplifiers results from nonlinearity in the voltage and current transfer functions within the power output stage. A technique is presented that uses a magnetic differential sensor to sense this distortion and to apply a supplementary error-correcting feedback path within the amplifier, in addition to overall negative feedback. Differential-current derived feedback is explored and the technique is shown to be applicable to both voltage and current transfer errors. Analysis reveals the system alignment for obtaining a distortion null and these results are confirmed using nonlinear transient analysis. Basic system topologies are presented using voltage transfer and transconductance gain cells as an aid to complete amplifier synthesis. Comparisons are made with feedforward-feedback error correction where the current-dumping amplifier is reconfigured using differential-current feedback. Generalisations are made to a multi-loop nest of amplifiers using individually aligned error correcting paths to achieve corresponding improvements in distortion reduction.

1 Introduction

A cardinal objective of analogue amplifier design is to identify circuit techniques that can minimise the distortion induced by transistor nonlinearity, especially in the output stage of power amplifier circuits. It has already been demonstrated that error-correction topologies exist that can substantially linearise the earlier stages of an amplifier [1], and the use of an enhanced cascode [2] can further reduce distortion under large voltage swing conditions. It is well understood that feedback alone [3] cannot completely eliminate distortion because of the need to reduce the loop gain at high frequency so as to maintain closed-loop stability. In theory, to eliminate broadband distortion, a feedforward path is required that can either extend beyond the main nonlinear feedback loop or correct locally for a nonlinearity.

This paper proposes differential-current derived feedback (DCDF) as a means of error correction that can accommodate both voltage and current transfer nonlinearities in the amplifier output stage. System validation is by transient analysis, and comparisons are made with the Albinson–Walker current-dumping amplifier [4–14], where it is shown that a dominant differential-error feedback loop is supplemented by feedforward, rather than comprising just feedforward.

Finally, DCDF is extended to an mth-order nest of DCDF loops, which yields greater distortion reduction both by increasing loop gain and by introducing m zeros in the output-stage-related error function.

Paper 1013G (E10), first received 31st August and in revised form 6th December 1993

The author is with the Department of Electronic Systems Engineering, University of Essex, United Kingdom

2 Differential-current sensing

DCDF uses a current transformer to sense the rate of change of a current that is related to an error signal, and it will be shown that this arrangement can compensate for the frequency-dependent and distortion-reduction characteristics of the closed-loop gain. The current transformer shown in Fig. 1 consists of a magnetic circuit and two loosely coupled windings. For the purpose of definition, the single-turn winding carrying the current to be sensed is the primary one, and the multi-turn winding used to derive a voltage proportional to the rate of change of the primary current is the secondary winding.

The open-circuit secondary voltage e_s expressed as a function of primary current I_p follows from standard electromagnetic theory as

$$e_s = L\frac{\partial I_p}{\partial t} = \frac{\mu_0 \mu_r NA}{2\pi R}\frac{\partial I_p}{\partial t} \tag{1a}$$

where the mutual inductance L for the transformer is defined as

$$L = \frac{\mu_0 \mu_r NA}{2\pi R} \tag{1b}$$

where R = mean radius of toroid, μ_0 = permeability of free space (H/m), A = mean cross-sectional area of toroid (m^2), and μ_r = relative permeability of ferrite.

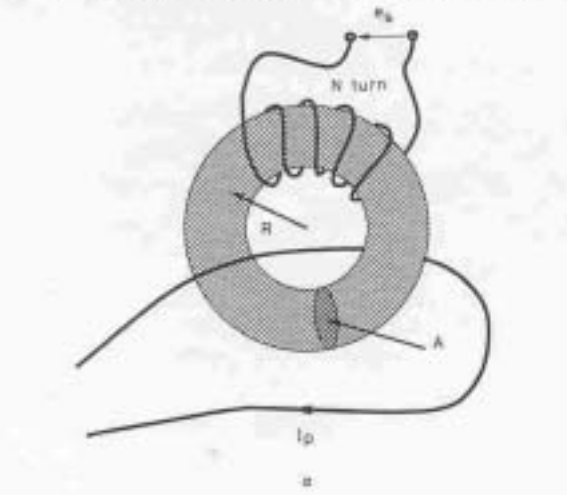
To illustrate the performance of an experimental differential-current sensor using a ferrite toroid, a current transformer was designed using the following (approximate) parameters: internal diameter = 10 mm, secondary turns = 5, external diameter = 40 mm, and primary turns = 1.

When a primary current of triangular waveform was injected, close adherence to a differential operator was demonstrated, as shown in Fig. 2, which was limited only by the coil's resonant frequency, ~10 MHz.

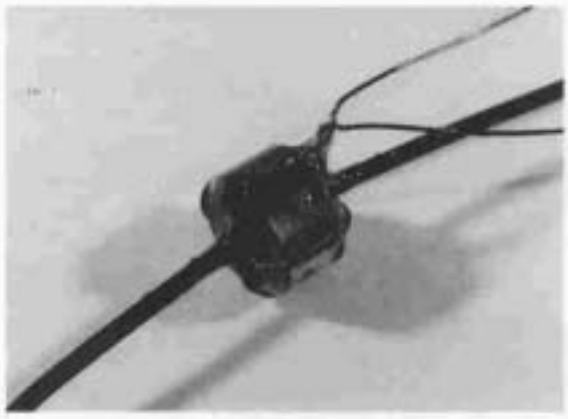
3 DCDF amplifier without error correction

An amplifier is shown in Fig. 3 with a forward gain A, an open-loop output impedance Z_0 and both unity-gain voltage feedback and DCDF. DCDF is modelled with a floating-voltage generator related to the rate of change of

output current I_0 where, using the Laplace operator s, this is described by $sL_0 I_0$ (the notation used in this paper gives the same subscript to the mutual inductance L_x and



a



b

Fig. 1 *Toroidal current sensing element*
a Actual
b Experimental

corresponding current to be sensed I_x). The closed-loop gain is

$$\frac{V_0}{V_i} = \frac{A}{1 + A + \left(\dfrac{Z_L - sL_0 A}{Z_L}\right)} \tag{2a}$$

However, by aligning inductance L_0 such that

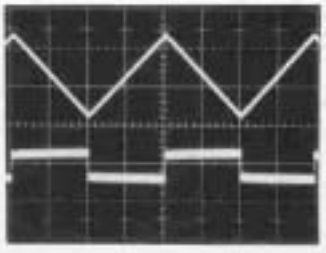
$$sL_0 A = Z_L \tag{2b}$$

the closed-loop gain becomes independent of the load impedance Z_L and reduces to

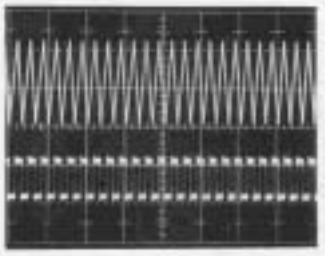
$$\frac{V_0}{V_i} = \frac{A}{1 + A} \tag{2c}$$

Eqn. 2b describes a balance condition that reduces the amplifier's closed-loop output impedance to zero, providing Z_L is resistive and A takes the form

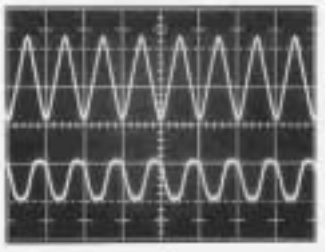
$$A = \frac{1}{sT_0} \tag{2d}$$



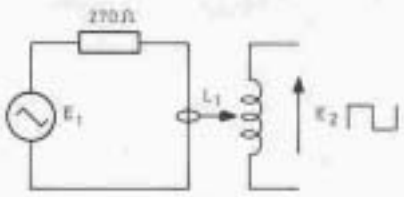
a



b



c



d

Fig. 2 *Example measurements on toroidal transformer for triangular-wave excitation*
a 50 kHz triangular waveform
Top trace: 10 V/div
Bottom trace: 100 mV/div
Timebase: 5 μs/div
b 500 kHz triangular waveform
Top trace: 10 V/div
Bottom trace: 1 mV/div
Timebase: 5 μs/div
c 5 MHz triangular waveform
Top trace: 10 V/div
Bottom trace: 10 V/div
Timebase: 5 μs/div
d Measurement system

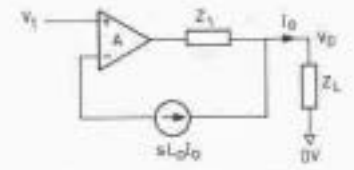


Fig. 3 *Basic amplifier using both voltage and DCDF*

Hence, from eqns. 2*b* and *d*, the optimum mutual inductance L_0 is calculated as

$$L_0 = Z_1 T_0 \tag{2e}$$

A computer simulation of the topology in Fig. 4 was performed where a family of closed-loop gains was as shown in Fig. 5. A gain parameter γ in the DCDF loop is varied from 0 to 1 (1 is optimum) in five steps to demonstrate the effect of varying degrees of current feedback. Without DCDF, a second-order response is visible at a frequency of ~154 kHz, whereas for optimum DCDF the resonance is almost suppressed.

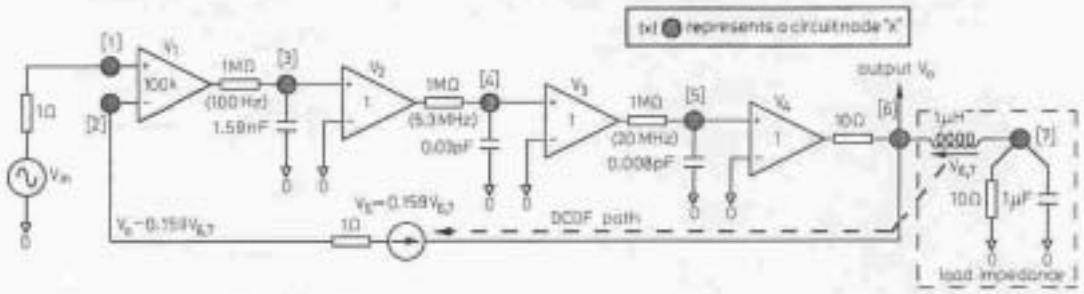


Fig. 4 *Simulation model of 3-pole DCDF amplifier*

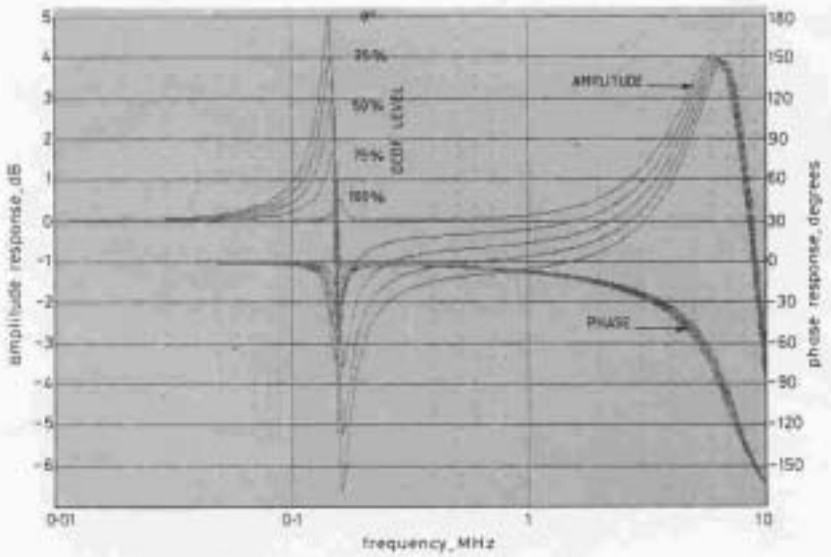


Fig. 5 *Section 2 example amplifier with 5 levels of DCDF ranging from 0 to 100%*

Although a DCDF enhanced amplifier can achieve zero output impedance with a finite forward-path gain, it is only realised if the output impedance Z_1 has a constant resistance and if the current transformer and amplifier transfer functions are optimally aligned. For high-output current amplifiers, the impedance Z_1 is nonlinear, and these conditions are not met. In Section 4, DCDF is extended to accommodate amplifiers with nonlinear output impedance.

4 Error-correction amplifiers using DCDF

4.1 Correction of voltage-transfer errors

Consider a linear amplifier of transfer function A that has a constant-value output resistance R_1, where an appropriate level of DCDF makes the closed-loop output impedance zero. If a nonlinear resistance is now connected to the amplifier output, then, because of the zero output impedance under optimal balance, the output voltage remains undistorted. This observation can be extended to a nonlinear stage N (where $N \approx 1$ and, at this stage of discussion has a very high input impedance) connected across the resistor R_1, as shown in Fig. 6. The stage N contributes to the total output current, but, because the system has a theoretical output impedance of zero, the amplifier closed-loop gain remains independent of the nonlinear gain N.

The amplifier in Fig. 6 uses three current-sensing paths (sensors L_0, L_1 and L_2) that feedback the differentials of I_1, I_2 and I_0, respectively. Initially, the forward-path amplifier is assumed to have both a zero output resistance and a voltage transfer function A, defined by eqn. 2*d*, where the closed-loop gain can be expressed as

$$\frac{V_0}{V_i} = \frac{A}{1+A} - \frac{1}{V_i} \times \left\{ \frac{sA(L_0 + L_2)I_0 + [R_1 - sA(L_1 - L_2)]I_1}{1+A} \right\} \tag{3a}$$

To make eqn. 3*a* independent of current I_1 and thus independent of amplifier nonlinearity,

$$R_1 = sA(L_1 - L_2) \tag{3b}$$

whereby eqn. 3*a* reduces to

$$V_0 = \frac{A}{1+A}[V_1 - s(L_0 + L_2)I_0] \tag{3c}$$

Hence, substituting for A from eqn. 2*d*, the closed-loop output impedance Z_{out} is

$$Z_{out} = \frac{s(L_0 + L_2)}{1 + sT_0} \tag{3d}$$

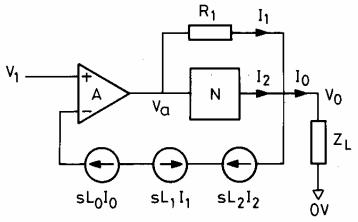


Fig. 6 *Error correcting amplifier using three current sensing loops*

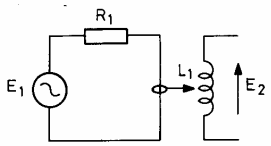


Fig. 7 *Measurement scheme for determining coil mutual inductance*

Eqns. 3*b* and *d* identify three correction scenarios:

(*a*) L_0, L_1, L_2 all finite: in general Z_{out} is finite, as defined by eqn. 3*d*; however, for the special case of $L_0 = -L_2$, then $Z_{out} = 0$. This condition can be implemented using a single sensing toroid that is simultaneously linked by I_0 and I_2, flowing in opposite directions.

(*b*) L_0 or L_2 set zero: balance remains possible, but the output impedance is finite.

(*c*) L_0 and L_2 zero, only I_1 sensed: under optimal balance $Z_{out} = 0$, and toroid magnetic flux is reduced as $I_1 \ll I_2$ for $N \approx 1$; only a single sensor is required.

An expression for closed-loop gain follows from eqn. 3*a* when the output stage is considered to have an inverse operator N^{-1}, where, if $I_1 = V_0(N^{-1} - 1)/R_1$ and $L_0 = L_2 = 0$,

$$\frac{V_0}{V_1}\left[1 + \frac{(R_1 - sAL_1)}{1+A}(N^{-1} - 1)\right] = \frac{A}{1+A} \tag{3e}$$

Defining an error function $E(s)$ where the actual gain $G(s) = V_0/V_1$ and the target gain $G_t(s) = G(s)$ for $N^{-1} = 1$, and assuming $(N^{-1} - 1)(1 - sAL_1/R_1)/(1 + A) \ll 1$, then

$$E(s) = \frac{G(s)}{G_t(s)} - 1 = (N^{-1} - 1)\left[\frac{1 - sAL_1/R_1}{1+A}\right] \tag{3f}$$

The circuit in Fig. 7 enables L_1 to be determined as $E_1 = E_2$ at the frequency where $|A| = 1$.

4.2 Error correction with finite output impedance of amplifier A

The system diagram in Fig. 8 shows a DCDF scheme where amplifier A has a finite output impedance Z_0. Both the input current I_n to the output stage and the current I_1 in impedance Z_1 can now inject nonlinear signal components into the overall feedback loop, which in turn distort the output signal. The proposed correction procedure is to use individual differential-current sensors for *both* currents I_n and I_1, where, in general,

$$I_1Z_1 + I_1Z_0 + I_nZ_0 - sA(I_nL_n + I_1L_1) = A(V_1 - V_0) - V_0$$

Error correction is achieved when the left-hand side is equated to zero, when, to desensitise system dependence on *both* I_1 and I_n, two conditions must be met where, for

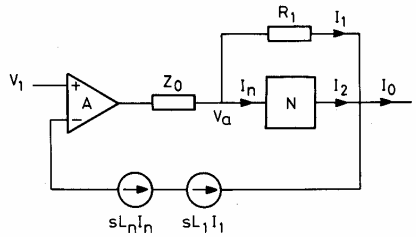


Fig. 8 *Error correction accommodating both voltage transfer errors in N and finite output impedance of A in association with nonlinear loading of output stage N*

the special case of A given by eqn. 2*d* and for resistive components $Z_0 = R_0$ and $Z_1 = R_1$,

$$L_1 = \frac{Z_0 + Z_1}{sA} = (R_0 + R_1)T_0 \tag{3g}$$

$$L_n = \frac{Z_0}{sA} = R_0 T_0 \tag{3h}$$

that is, $L_1/L_n = 1 + R_1/R_0$.

4.3 DCDF error-correction scheme using cascaded transconductance stages

Two cascaded transconductance amplifiers g_1, g_2 are shown in Fig. 9, where g_2 has local capacitive feedback via Z_0 and a local load impedance Z_2. The second stage

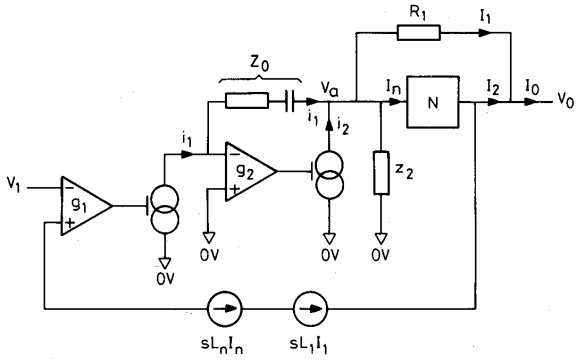


Fig. 9 *DCDF error-correction amplifier using transconductance stages*

also drives the nonlinear output stage and supplies current to the feedforward bypass impedance Z_1. As in Section 4.2, a dual DCDF scheme is used to correct both voltage-transfer errors and the effect of nonlinear loading on the frequency-dependent output impedance of the second stage.

Defining α as

$$\alpha = \frac{g_2 Z_2}{1 + g_2 Z_2} \tag{4a}$$

the closed-loop transfer function is expressed as

$$\frac{V_0}{V_1} = \frac{\alpha g_1(Z_0 - 1/g_2) - \dfrac{I_1 Z_1 - s\alpha g_1(Z_0 - 1/g_2)(L_1 I_1 + L_n I_n) + \alpha(I_1 + I_n)/g_2}{V_1}}{1 + \alpha g_1(Z_0 - 1/g_2)}$$

Equating the terms in I_1 and I_n to zero, setting $Z_0 = r_0 + 1/(sC_0)$ and $Z_1 = R_1$, $Z_2 = R_2$ and aligning $r_0 = 1/g_2$, the two balance conditions follow as

(*a*) eliminating I_1

$$L_1 = \frac{1 + g_2 Z_1 + \dfrac{Z_1}{Z_2}}{sg_1(g_2 Z_0 - 1)} = \left[\frac{1}{g_1 g_2} + \frac{R_1}{g_1} + \frac{R_1}{R_2 g_1 g_2}\right] C_0 \quad (4c)$$

(*b*) eliminating I_n

$$L_n = \frac{1}{sg_1(g_2 Z_0 - 1)} = \frac{C_0}{g_1 g_2} \quad (4d)$$

that is, $L_1/L_n = 1 + g_2 R_1 + R_1/R_2$. R_2 represents the output impedance of the second transconductance stage, which can be made large by using either a grounded base stage or an enhanced cascode topology [2], whereby, if $R_2(1/R_1 + g_2) \gg 1$, then eqn. 4*c* reduces to $L_1 = C_0(R_1 + 1/g_2)/g_1$. Hence, providing the core cells g_1, g_2 exhibit a wide bandwidth with constant transconductance and L_1, L_n are optimally aligned, then, by setting $Z_0 = 1/g_2 + 1/(sC_0)$ with $g_2 Z_2 \gg 1$, a first-order closed-loop gain follows from eqns. 4*a* and *b*

$$\frac{V_0}{V_1} = \frac{g_1 g_2 Z_2(Z_0 - 1/g_2)}{1 + g_2 Z_2[1 + g_1(Z_0 - 1/g_2)]} = \frac{g_1}{g_1 + sC_0} \quad (4e)$$

4.4 High-frequency performance limitation of current transformer

Fig. 10 shows a DCDF amplifier with feedback factor k that includes a second-order model of the HF resonance

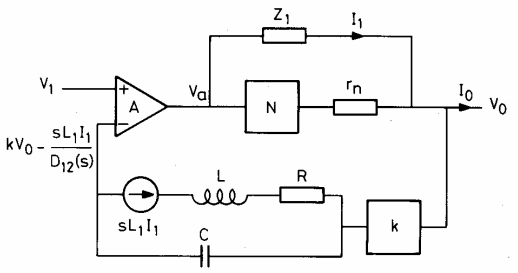


Fig. 10 *DCDF with second-order bandlimited current transformer*

of the current transformer. Using the transformation

$$sL_1 \cdots \frac{sL_1}{D_{12}(s)}$$

where

$$D_{12}(s) = 1 + \frac{1}{Q_1}\left(\frac{s}{\omega_1}\right) + \left(\frac{s}{\omega_1}\right)^2$$

enables the closed-loop gain to be expressed as

$$G_{n0}(s) = \left\{k + \frac{1}{A}\left[\frac{1}{D_{12}(s)} + \frac{1}{N_T}\left(1 - \frac{1}{D_{12}(s)}\right)\right]\right\}^{-1} \quad (4f)$$

To determine the sensitivity of $G_{n0}(s)$ to $D_{12}(s)$, an error function $E(s)$ is defined

$$E(s) = \frac{G_{n0}(s)}{G_{n0}(s)|_{D_{12}(s)=1}} - 1$$

where the advantage of making $\omega_1 >$ {amplifier gain–bandwidth product} is shown by

$$E(s) = -\left[\frac{D_{12}(s) - 1}{D_{12}(s)}\right]\left[\frac{N_T - 1}{N_T}\right]\frac{1}{1 + kA}$$

$$= -\frac{s}{\omega_1}\left[\frac{1}{Q_1} + \frac{s}{\omega_1}\right]\left[\frac{N_T - 1}{N_T}\right]\frac{1}{1 + kA} \quad (4g)$$

5 Current-dumping amplifier

The *current-dumping* amplifier illustrated in elemental form in Fig. 11*a* was introduced in 1975 by Albinson and Walker [4] and is described as a feedback amplifier with feedforward error correction [14]. The basic operation suggests that the pre-output-stage amplifier (A in the present notation) both drives the output stage N and supplies an error-correction current via a resistor R_1 ($Z_1 = R_1$) to compensate for the nonlinearity in N. An alternative thesis is proposed here that the principal distortion-correction process in the Albinson–Walker amplifier is differential-error feedback, although the supporting role of a partial bypass around N is recognised. We also show how DCDF can compensate for the finite output impedance of the class A stage, which has to supply current to both the output transistors and the feedforward resistor.

5.2 Comparison of current dumping and DCDF amplifiers

A progressive transformation from current dumping to DCDF is shown in Fig. 11*a*, *b*, *c* and *d*, where the output stage is modelled as a nonlinear transfer characteristic N with nonlinear output impedance r_n. In Fig. 11*c* and *d* this is combined with Z_2, where $Z_{2n} = r_n + Z_2$. Thus, defining $N_T^{-1} = 1 + NZ_1/Z_{2n}$, the transfer function for the unbalanced amplifier is

$$\frac{V_0}{V_1} = \frac{1}{1 + s\tau}\left[1 - \frac{I_0 Z_1}{V_1}\left(s\tau + \frac{1 - sA\tau}{AN_T^{-1}}\right)\right] \quad (5a)$$

Under optimum balance $sA\tau = 1$, where $R_1 = Z_1$ and $\tau = L_2/R_1$ are substituted,

$$\frac{V_0}{V_1} = \frac{A}{1 + A} - \frac{I_0\{(sL_2)//(R_1)\}}{V_1} \quad (5b)$$

The schematic diagram of Fig. 11*d* distinguishes between error feedforward and error feedback. Stage N has partial feedforward via Z_1 but appears within the feedback path, whereas DCDF senses the output-stage error. That is, feedforward is not fundamental to this process. However, the feedforward summing network $\{Z_1, Z_{2n}\}$ supplements output-stage performance and reduces the differentiated-error signal and, hence the dynamic range requirements of the amplifier input stage. For the *current-dumping* amplifier, increasing Z_1 requires an increase in the inductance of L_2 to maintain balance. However, with DCDF,

the output impedance need not increase as the generator $sL_2 I_0$ shown in Fig. 11*d* is independent of the balance condition. Fig. 12 shows an equivalent amplifier with output voltage derived feedback and DCDF.

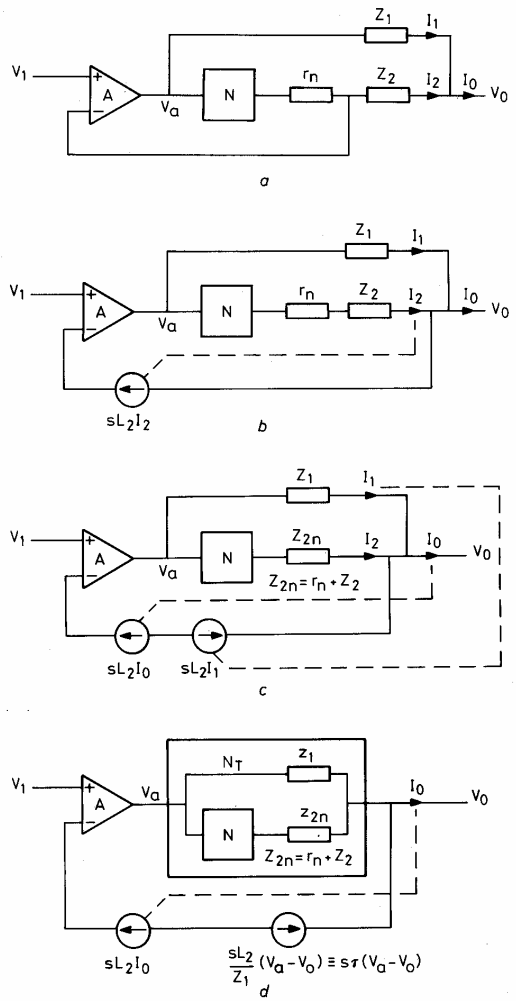


Fig. 11 *Transformation from current dumping to DCDF*
a Current dumping topology with unity-gain feedback
b Current dumping amplifier with differential-current sensor
c Current dumping amplifier with differential-current sensing of both output and feedforward currents
d Current dumping amplifier configured with DCDF

5.2 DCDF enhancement of current dumping

The class *A* pre-output stage of a current-dumping amplifier supplies current to both the output transistors and the feedforward impedance Z_1, where, because these currents are generally nonlinear functions of the input signal, distortion occurs when the class *A* amplifier has a finite output impedance Z_a. To reduce this distortion, Fig. 13 shows a DCDF enhanced current-dumping amplifier that senses both I_1 (the current in Z_1) and I_n (the nonlinear input current to the output stage *N*), where the closed-loop gain can be expressed as

$$\frac{V_0}{V_1} = \frac{A}{1+A} + \frac{\begin{matrix} sA(L_n I_n + L_1 I_1) - A(I_0 - I_1)Z_2 \\ -(I_1 + I_n)Z_a - I_1 Z_1 \end{matrix}}{(1+A)V_1} \quad (6a)$$

where, putting $Z_2 = sL_2$ and equating terms in I_1 and I_n, the balance conditions are stated as

$$A = \frac{Z_a + Z_1}{s(L_1 + L_2)} = \frac{Z_a}{sL_n} \quad (6b)$$

that is

$$L_n = (L_1 + L_2)\frac{Z_a}{Z_a + Z_1} \quad (6c)$$

Although, in this scheme, both I_n and I_1 are sensed, in the Walker–Albinson amplifier, $L_1 = 0$ because of the

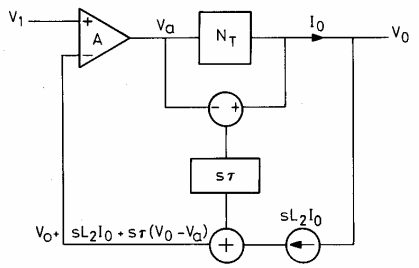


Fig. 12 *Differential-error and output-voltage derived feedback amplifier*

inclusion of Z_2; thus only a single low-current sensor for I_n is required. The technique is also applicable to the transconductance amplifier of Section 4.3, where, comparing *A* and Z_a with the core cells g_1, g_2 and using a Norton–Thevenin transformation,

$$A = g_1\left(r_0 - \frac{1}{g_2} + \frac{1}{sC_0}\right) \quad (6d)$$

$$Z_a = \frac{1}{g_2} \quad (6e)$$

If $g_2 r_0 = 1$, then $A = g_1/(sC_0)$, and the balance conditions follow as

$$L_1 + L_2 = \frac{C_0}{g_1}\left(Z_1 + \frac{1}{g_2}\right) \quad (6f)$$

$$L_n = \frac{C_0}{g_1 g_2} \quad (6g)$$

Providing the core of the transconductance cell g_2 has a broad bandwidth, balance accommodates both output-

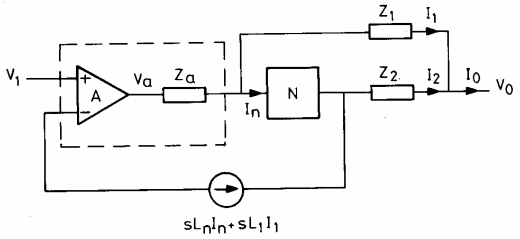


Fig. 13 *Current dumping amplifier with DCDF enhancement*

stage nonlinearity and nonlinear loading on the forward amplifier.

6 Verification of DCDF using nonlinear transient analysis

To validate the distortion reduction of a dual-sensor DCDF amplifier, a nonlinear transient analysis was per-

formed on the Fig. 9 transconductance topology using the nonlinear model described in Fig. 14, where bias voltages B_1 and B_2 determine the output-stage bias current.

class C (0 mA bias), where DCDF alignment is most sensitive; however, under optimum balance, low-level distortion cannot be observed in Fig. 15 for 100% correction.

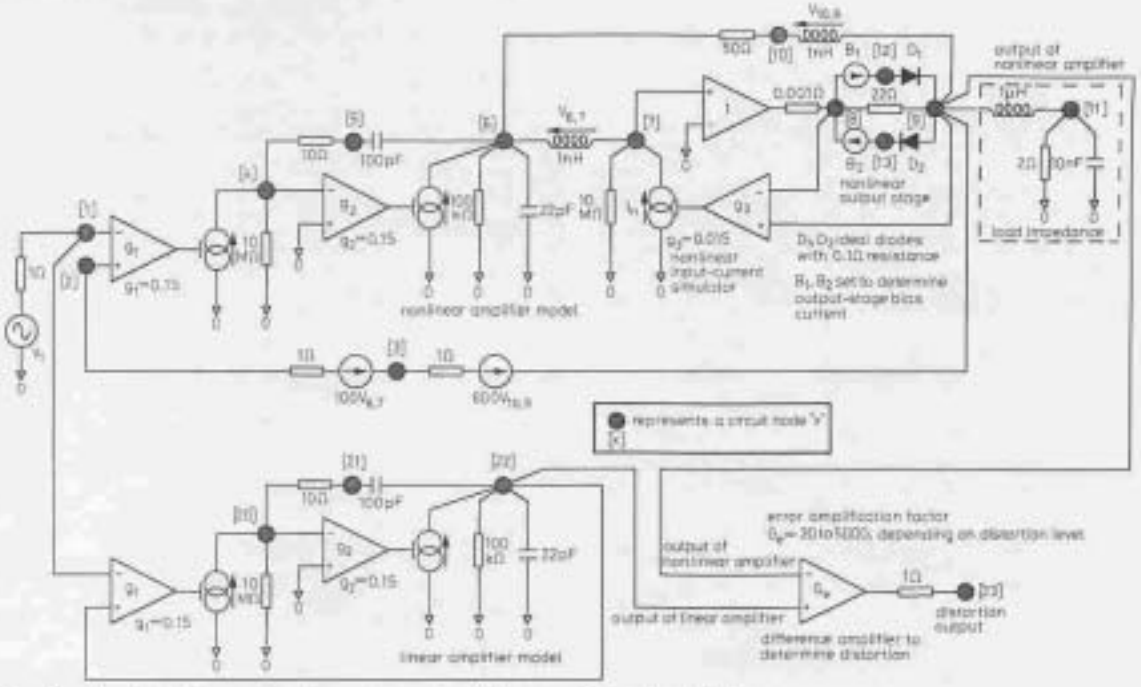


Fig. 14 *Simulation of linear and nonlinear amplifiers with a difference amplifier to determine distortion*

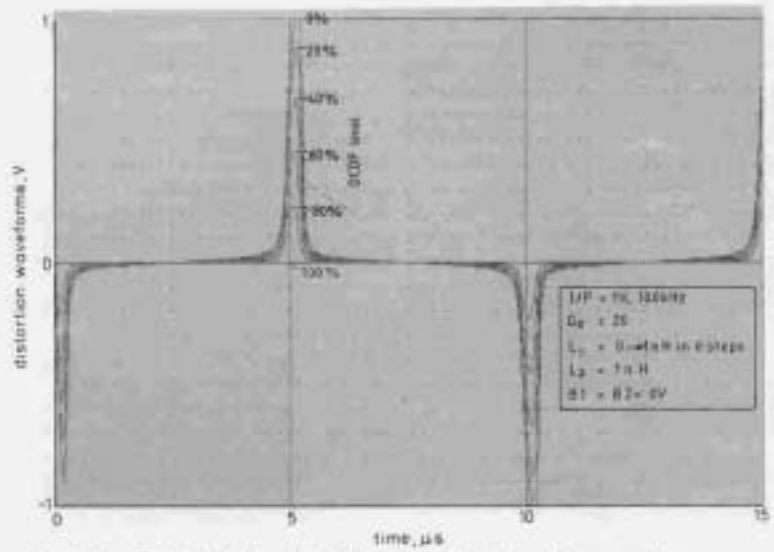


Fig. 15 *Distortion waveforms for class C biased amplifier with 6 levels of DCDF and optimum loading correction*

Three sets of distortion traces are presented in Figs. 15, 16 and 17 and relate to the data in Table 1. Component values are defined in Fig. 14, and all time-domain distortion waveforms correspond to a 1 V amplitude and a 100 kHz sinusoidal input signal (chosen to accentuate low-level crossover distortion), with a load impedance of 2 Ω//10 nF. Significant impulsive distortion results under

Under class *AB* (47 mA bias), distortion is reduced where the error traces in Figs. 16 and 17 confirm the balance conditions deduced in Section 4. However, for optimum distortion correction again, both DCDF loops must be aligned, where, for example, if difference amplifier gain $G_e = 5000$ is selected, no distortion is observed on the 100% trace in Fig. 17.

Table 1: Amplifier and evaluation data

	Output bias current	Time axis	Error amplifier gain	Input DCDF
Fig. 15	0 mA	0–15 μs	20	Yes
Fig. 16	47 mA	0–30 μs	420	No
Fig. 17	47 mA	0–30 μs	5000	Yes

7 *m*th-order nested DCDF error-correction amplifier

DCDF can be extended to systems of order m that incorporate m first-order amplifiers and m DCDF correction paths using m feedforward resistors. An mth-order DCDF system is shown in Fig. 18. The inner amplifier loop (designated by a gain N_1) is identical to that discussed in Sections 4.1 and 4.2 and includes dual correction for both output-stage voltage transfer and current-transfer nonlinearities. A second amplifier with a forward-path gain A_2, in association with N_1, now forms a second feedback loop, where an additional feedforward resistor R_2 and differential-current transformer realise the second correction path. It is assumed unnecessary here to incorporate current correction as the input current to amplifier A_1 is generally both linear and negligible, and hence only one transformer is required. This method of nesting correction amplifiers can then be systematically extended to a composite mth-order topology.

The closed-loop transfer function of the mth-order amplifier and its associated error function with respect to nonlinear gain N_0 can be derived as follows:

Consider the inner amplifier loop formed by A_1, N_0, R_1, Z_{01}, L_1 and L_n, which is also depicted in Fig. 8 where $A = A_1$, $N = N_0$ and $Z_0 = Z_{01}$. To obtain an expression

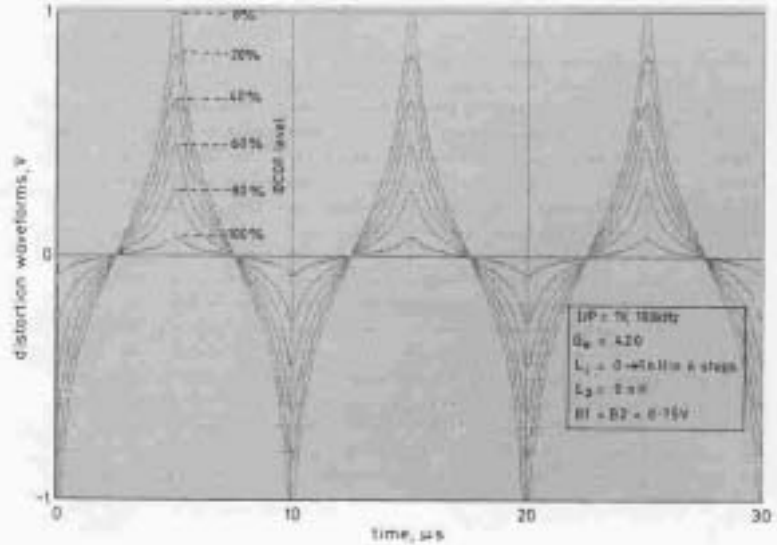


Fig. 16 *Distortion waveforms for class AB biased amplifiers with 6 levels of DCDF and no loading correction*

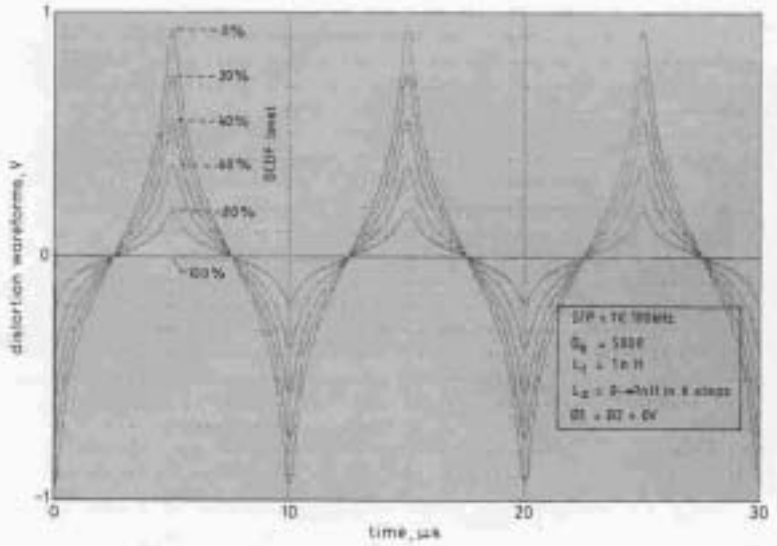


Fig. 17 *Distortion waveforms for class AB biased amplifier with optimum voltage-transfer correction and loading correction swept over 6 levels*

for I_n in terms of the output voltage V_0, the output-stage N_0 is given a nonlinear input impedance Z_{n0} (which is partially output-load-dependent), where $I_n = V_0/(N_0 Z_{n0})$.

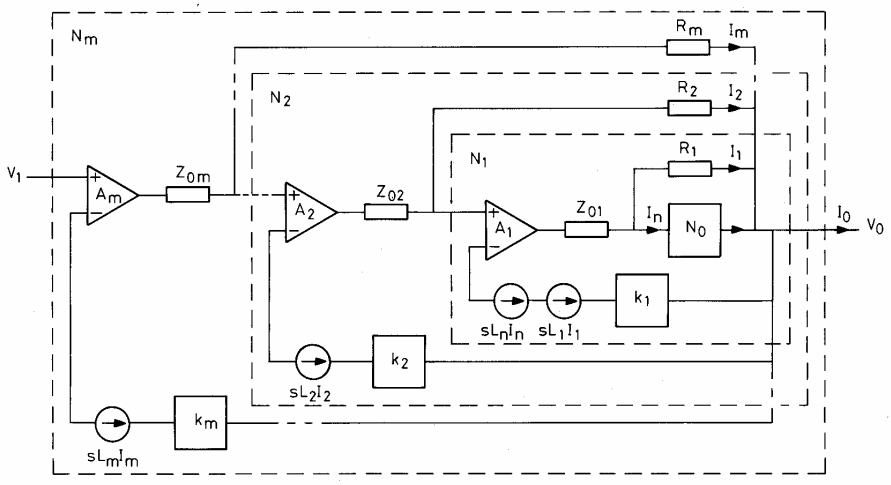


Fig. 18 *General mth-order, nested DCDF error-correction amplifier*

The transfer function of the inner amplifier loop $N_1 = V_0/V_1$ then follows as

$$\frac{V_0}{V_1} = \frac{A_1}{1 + k_1 A_1} + \frac{1}{V_1} \frac{sA_1(L_n I_n + L_1 I_1) - (I_1 + I_n)Z_{01} - I_1 R_1}{1 + k_1 A_1}$$

from which, balance equations λ_n, λ_m, and a factor Q_m can be identified as follows:

$$\lambda_n = \frac{1}{Z_{n0} N_0}(Z_{01} - sL_n A_1) \tag{7a}$$

$$\lambda_m = 1 + \frac{Z_{0m} - sL_m A_m}{R_m} \tag{7b}$$

$$Q_m = 1 + A_m(k_m - 1) \tag{7c}$$

An iterative expression for the closed-loop transfer function N_m in terms of N_{m-1} is

$$N_m = \left[1 + \frac{Q_m + \lambda_n + \lambda_m(N_{m-1}^{-1} - 1)}{A_m}\right]^{-1} \tag{7d}$$

However, except for amplifier N_1, the input impedances $Z_{n(m-1)}$ of the remaining amplifiers in the nest are large, and therefore, for $m > 1$, corresponding terms $\lambda_n \approx 0$. Hence, applying eqn. 7*d*, the closed-loop gain N_m follows as

$$N_m = \left\{1 + \frac{Q_m}{A_m} + \sum_{r=2}^{m}\left[\frac{Q_{r-1}}{A_{r-1}} \prod_{r}^{m} \frac{\lambda_r}{A_r}\right] + (N_0^{-1} - 1)\prod_{r=1}^{m}\frac{\lambda_r}{A_r} + \frac{\lambda_n}{A_1}\prod_{r=2}^{m}\frac{\lambda_r}{A_r}\right\}^{-1} \tag{7e}$$

To express the overall sensitivity to imbalance and loop gain, the error function E_m is

$$E_m = \frac{(1 - N_0^{-1}) - \lambda_n \prod_{r=2}^{m} \lambda_r}{N_m \prod_{r=1}^{m} A_r} \tag{7f}$$

where the target transfer function is given by N_m for $N_0 = 1$ and $\lambda_n = 0$. Eqn. 7*f* reveals that the dependence upon N_0 and the input-current of the output stage are reduced by both the product of the m forward-path amplifier gains $A_1 A_2 A_3 \cdots A_m$ and the product of the m transmission zeros involving the balance condition lambda functions. Also, assuming optimal alignment of the m balance conditions, eqn. 7*e* reduces to $N_m = N_{opt}$, where

$$N_{opt} = \frac{A_m}{1 + k_m A_m} \approx \frac{1}{k_m} \tag{7g}$$

where, in this idealised case, stability is dependent only upon A_m and k_m.

The advantage of this new structure over earlier nested-amplifier proposals [16] is the ability to generate m transmission zeros in the nonlinear error function, which is further desensitised by the product of the amplifier gains $A_1 A_2 A_3 \cdots A_m$. Also, the multiple DCDF loops assist in the design of a stable amplifier system where, under optimal balance, the closed-loop gain takes the form $A_m/(1 + k_m A_m)$; that is, for an ideal system, only one amplifier determines stability, although practical circuitry may lead to a more complicated result.

8 Conclusions

A method for implementing an efficient means of error correction has been described that uses a differential-current transformer where the transfer functions of transformer and forward-path amplifier are complementary. Both voltage-transfer and transconductance pre-output-stage amplifiers can be accommodated, and a second DCDF loop can correct for output-stage nonlinear current gain.

Error-correction performance and the expressions for optimum balance were confirmed using nonlinear transient analysis of a stylised, nonlinear amplifier model. However, unlike the *current-dumping* topology, the balance state is independent of the feedback factor, as the floating-current transformers can be connected directly to the input stage. This also increases their resonant frequency by allowing a reduction in the number of secondary turns.

DCDF and current-dumping feedforward error-correction were compared, and, with DCDF, the level of feedforward addition was shown to be uncritical of the balance state. However, feedforward can reduce the magnitude of DCDF sensed error signals resulting from both crossover distortion and limited output-stage high-frequency gain.

The DCDF technique was extended to an *m*th-order nested-feedback topology using *m* feedforward resistors in association with *m* differential-current transformers. This system reduces further sensitivity to output-stage nonlinearity and suggests an alternative means of controlling closed-loop stability such that, under optimal alignment, only the outermost amplifier and feedback path of the nest appear in the expression for closed-loop gain.

It is expected that the dual DCDF correction system can be introduced into existing circuits with only minor modifications to topology, although the potential for an *m*th-order loop should be of interest in very low-distortion applications. The modest-cost overhead compared with other more circuit-intensive proposals [13, 15] should also prove attractive, especially as the floating-circuit elements are passive and, in association with appropriate screening, are independent of power supply-induced distortion.

9 References

1 HAWKSFORD, M.J.: 'Distortion correction circuits for audio amplifiers', *JAES*, 1981, **29**, pp. 503–510

2 HAWKSFORD, M.O.J.: 'Reduction of transistor slope impedance dependent distortion in large-signal amplifiers', *JAES*, 1988, **36**, pp. 213–222

3 LIPSHITZ, S., and VANDERKOOY, J.: 'Is zero distortion possible with feedback?'. Presented at 76th Convention of Audio Eng. Soc., Preprint 2170 (F-4), 8th–11th October 1984, New York

4 WALKER, P.J., and ALBINSON, M.P.: 'Current dumping audio amplifier'. Presented at 50th Convention of Audio Eng. Soc., London, UK, 4th–7th March 1975

5 WALKER, P.J.: 'Current dumping audio power amplifier', *Wireless World*, 1975, **81**, pp. 560–582

6 VANDERKOOY, J., and LIPSHITZ, S.P.: 'Current dumping — does it really work?', 1978, **84**, (1510), pp. 38–40

7 VANDERKOOY, J., and LIPSHITZ, S.P.: 'Feedforward error correction in power amplifiers', *JAES*, 1980, **28**, pp. 2–16

8 McLOUGHLIN, M.: 'Current dumping review — 1', *Wireless World*, 1983, **89**, (1572), pp. 39–43

9 McLOUGHLIN, M.: 'Current dumping review — 2', *Wireless World*, 1983, **89**, (1573), pp. 35–41

10 WALKER, P.: 'Current dumping', *Wireless World*, 1983, **89**, (1575), p. 49

11 BAXANDALL, P.J.: 'Current dumping', *Wireless World*, 1983, **89**, (1575), pp. 49–50

12 VANDERKOOY, J., and LIPSHITZ, S.P.: 'Current dumping review', *Wireless World*, 1984, **90**, (1577), p. 49

13 HAWKSFORD, M.J.: 'Power amplifier output stage design incorporating error feedback correction with current dumping enhancement'. Presented at 74th Convention of AES, Preprint 1993 (B-4), 8th–12th October 1983, New York

14 ALLINSON, N.M., and WELLINGHAM, J.: 'Distortion reduction in frequency-dependent feedback-feedforward amplifiers', *Int. J. Electron.*, 1985, **59**, pp. 667–683

15 HAWKSFORD, M.J.: 'Distortion correction in audio power amplifiers', *JAES*, 1981, **29**, pp. 27–30

16 CHERRY, E.M.: 'A new result in negative-feedback theory, and its application to audio power amplifiers', *IEEE Trans.*, 1978, **CTA-6**, pp. 265–288

Quad-input current-mode asymmetric cell (CMAC) with error correction applications in single-ended and balanced audio amplifiers

M.O.Hawksford

Indexing terms: Audio amplifiers, Quad-input asymmetric cell, Cell topology

Abstract: An amplifier topology is introduced that functions principally in a current-steering mode and offers wide bandwidth and low distortion. The cell can accommodate an additional difference port and thus finds application in a range of audio amplifier incorporating error correction. The cell topology and its properties are described and methods of error correction are reviewed where it is shown that the new topology can be used in both single-ended and balanced power amplifiers.

1 Introduction

It is now widely accepted, among audio circuit designers, that simplicity in the choice of topology is desirable, providing a high standard of objective performance is maintained within the operating envelope of the system; however, this strategy requires the identification of techniques that use active devices in their most linear operating mode. Current-steering circuits [1] offer wide bandwidth as the charging of parasitic capacitance, with associated wide voltage swings on multiple circuit nodes, is reduced. For similar reasons, such cells can also yield a useful reduction in nonlinear distortion providing appropriate cascode circuitry [2] is used at circuit nodes where wide voltage swings are encountered.

A critical factor affecting audio amplifier performance relates to the signal currents that flow in ground lines and power supplies as these can both create direct errors and induce errors through interstage interaction. It is, therefore, desirable to identify circuits that minimise and localise such currents so as to enable signal currents to flow in well-defined closed paths that ideally exclude the power supply and ground line. With signals of modest level ($\sim 2V_{RMS}$) it is straightforward to achieve low distortion. However, for higher-level signals, such as encountered in power amplifiers, this objective is more problematic because of transistor parameter modulation and output-stage nonlinearity.

IEE Proceedings online no. 19960149

Paper first received 2nd December 1994 and in revised form 20th October 1995

The author is with the Centre for Audio Research and Engineering, Department of Electronic Systems Engineering, University of Essex, Wivenhoe Park, Colchester C04 3SQ, UK

To address these problems, this paper reviews some of the principles of error correction [3–6] and shows how a new topology, defined as a current-mode asymmetric amplifier cell (CMAC), can be adapted for operation in such amplifiers by providing four input nodes, two of which function as an error-feedback port. This tutorial element describes a unified approach to error correction in both single-ended and balanced audio power amplifiers. An analysis of nonlinear distortion is presented and it is shown how a local error-correcting feedforward path within a CMAC can enhance its internal linearity.

2 Elementary CMAC topologies and their properties

Two widely adopted elementary amplifier cells [7] are shown in Figs. 1 and 2 that achieve inverting and noninverting gain, respectively. However, in Fig. 3 a modified topology is depicted that incorporates properties of both the earlier cells and forms the kernel of a CMAC. In this basic configuration it is assumed that no loading is applied to the amplifier output port. Therefore, the signal current flows directly between collector and emitter in a closed path that can exclude the power supply. Consequently, the input current, i_{in}, must correspond to the base current, i_b, of the transistor and results in an input impedance, Z_{in}, given by

$$Z_{in} = (R_1 + r_e)(1 + \beta) \quad (1)$$

where β is the transistor collector-base current gain and $r_e = \partial V_{BE}/\partial I_E$. In applications requiring a low voltage gain, linearity improves as $R_1 >> r_e$, where in the limit of $A_v = 1$, $R_1 = \infty$ and output and input are connected only by R_2. It follows that any signal current drawn from the output must draw additional current directly from the input which lowers the input impedance, as in this simple example the input (applied at the junction of R_1 and R_2) is unbuffered and input and output are linked by R_2. Consequently, in practical circuits either the input or the output should be buffered and the use of complementary circuitry to reduce distortion and facilitate DC biasing is desirable. To illustrate how this can be achieved, a circuit example is given in Fig. 4 incorporating a complementary, compound cell formed using transistors $T_1 - T_6$ while T_7 and T_8 form a complementary grounded-base stage. Transistors T_5 and T_6 limit the collector–emitter voltage of transistors T_3 and T_4, allowing them enhanced bias stability and freedom from large signal swings that can cause distortion through transistor parametric modulation. The distor-

tion performance of the compound, complementary cell is studied in Section 3.

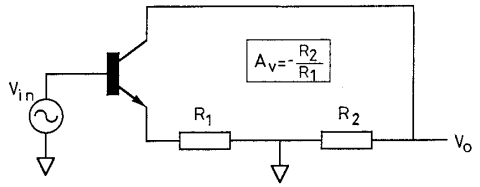


Fig. 1 *Common-emitter amplifier*

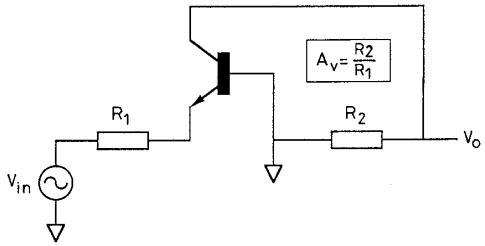


Fig. 2 *Grounded-base amplifier*

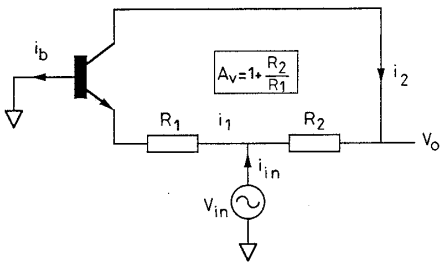


Fig. 3 *Current-mode asymetric cell*

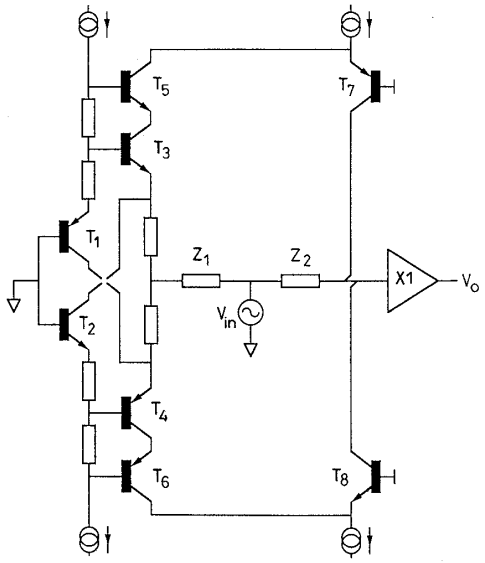


Fig. 4 *CMAC with complementary cell, output cascode and unity-gain buffer*

The circuit of Fig. 4 also can have a relatively high input impedance where to calculate the input signal current all current paths via transistor bases must be summed. However, to make the cell more tolerant to real circuit operating conditions some modifications are recommended:

(i) input-stage common-collector buffer

(ii) output-stage common-collector buffer

(iii) integral current mirror within the cell.

Recommendations (i) and (ii) are self evident; however, the incorporation of a current mirror offers a number of improvements where the basic topology is shown in Fig. 5. An input transistor T_1 acts as a unity-gain buffer stage but where the collector current i_x is mirrored with a gain m and applied to the emitter of T_2.

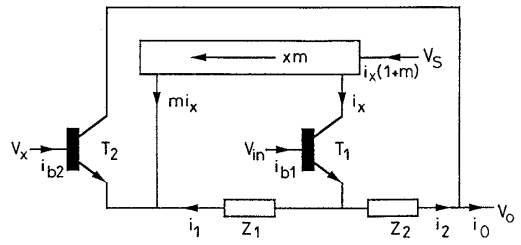


Fig. 5 *Current mirror enhanced CMAC*

The current mirror offers a useful performance tolerance to output load current i_o. In this application it is assumed that the input impedance of the current mirror is low and is referenced to the power supply voltage while the output impedance is high thus acting as a current source to the emitter of T_2. Assume initially $V_{in} = 0$V and $m = 0$, then, because of the low impedance at the emitter of T_1, most of the output current i_o flows in Z_2 yielding an output impedance $Z_o \approx Z_2$. However, for $m > 0$ and assuming V_x, $V_{in} = 0$V, then $i_x \approx i_2$ giving $i_o = i_2 (1 + m)$, whereby

$$Z_o = \frac{Z_2}{1+m} \tag{2}$$

Because the current i_x is reduced from i_o to $i_o/(1+m)$, the input impedance is increased by $(1+m)$ and any tendency for modulation of r_{e1} (i.e. for T_1, $r_{e1} = \partial V_{BE1}/\partial I_{E1}$) is correspondingly reduced, thus lowering distortion contribution from this stage.

The current mirror engenders a tolerance to loading and simultaneously aids linearisation of T_1. The value of m is uncritical in a basic CMAC (in practice being constrained by DC biasing and stability considerations) although a high value will result in both a lower output resistance and a higher input impedance. Observe that the signal current i_1 is unaffected by the choice of m although, as m is increased, a greater fraction of the output current now flows in T_2 contributing to the distortion generated by this transistor. Hence, distortion performance in a CMAC cell remains critical upon the stage represented by T_2, requiring appropriate selection of bias current and topology, where an example is given in Section 5 (see Fig. 16). However, as described later in this Section, when extra error correction input ports are included, m becomes a critical parameter that requires an accurate definition.

A current mirror results in a flow of signal current via the power supply which is contrary to the conceptual aim of a CMAC cell. However, the magnitude of this current is determined principally by the output current i_o as the only other component results from the small base current of T_2. In most CMAC applications, it is envisaged that the output current will be reduced by incorporating an output buffer stage (for example, an emitter follower), thus signal current in the current mirror is lower aiding linearity in this stage. Such a strategy is followed in Section 4.

2.1 Quad-input node CMAC

An extension to the CMAC is to use the low-impedance nodes at the emitters of T_1 and T_2 as current input ports enabling the incorporation of error correction [6] as described in Section 4. The modified cell is shown in Fig. 6.

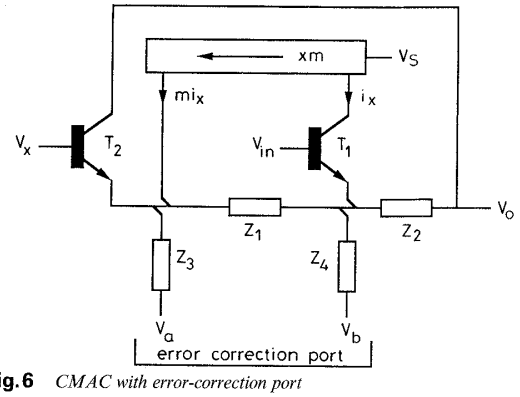


Fig. 6 *CMAC with error-correction port*

Expressing V_o as $f(V_{in}, V_x, m, Z_1, Z_2, Z_3, Z_4, V_a, V_b)$, then

$$V_o = \frac{Z_2}{1+m}\left[\frac{V_a}{Z_3} - m\frac{V_b}{Z_4}\right] - V_x\left[\frac{Z_2}{Z_1} + \left(\frac{1}{1+m}\right)\frac{Z_2}{Z_3}\right] + V_{in}\left[1 + \frac{Z_2}{Z_1} + \left(\frac{m}{1+m}\right)\frac{Z_2}{Z_4}\right]$$

To force V_o to be dependent upon $(V_a - V_b)$, make

$$\frac{Z_2}{1+m}\left[\frac{V_a}{Z_3} - m\frac{V_b}{Z_4}\right] = V_a - V_b$$

that is, expressing Z_3, Z_4 in terms of Z_2,

$$Z_3 = \left(\frac{1}{1+m}\right)Z_2 \tag{3}$$

$$Z_4 = \left(\frac{m}{1+m}\right)Z_2 \tag{4}$$

whereby

$$V_o = (V_a - V_b) - V_x\left(1 + \frac{Z_2}{Z_1}\right) + V_{in}\left(2 + \frac{Z_2}{Z_1}\right) \tag{5}$$

The CMAC with an error-correction port (with unity weighting, see '1' on symbol) is represented by the symbol shown in Fig. 7, which forms the amplifier cell used in the example power amplifier circuits incorporating error correction.

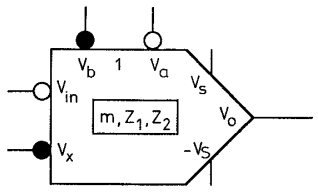


Fig. 7 *Symbol for CMAC with single error-correction port*
○——— noninverting input
•——— inverting input

3 Distortion performance of complementary amplifier cell

Consider the complementary cell shown in Fig. 8 as a subcircuit within a CMAC where the transistors are biased with two voltage generators V_B. Each emitter resistor is R_E and the input signal current, i, divides symmetrically between the two transistors such that for T_1, $I_{E1} = I_B - i$ and for T_2, $I_{E1} = I_B + i$. Because of nonlinear transistor action, the common-mode bias current I_B is also a function of input current having a quiescent value I_{BQ} when $i = 0$.

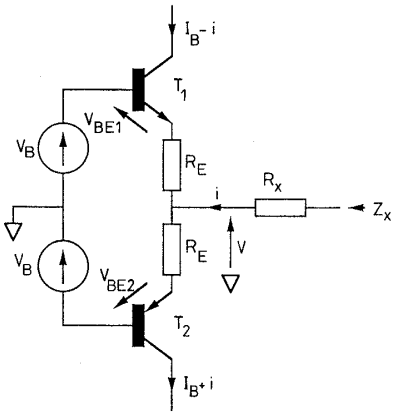


Fig. 8 *Elementary complementary emitter-coupled cell*

With reference to Fig. 8 the input impedance Z_{in} observed at the junction of the two emitter resistors R_E is established as follows:

$$v + (I_B - 0.5i)R_E + V_{BE1} - V_B = 0 \tag{6}$$

$$v - (I_B + 0.5i)R_E + V_{BE2} + V_B = 0 \tag{7}$$

Addition of eqns. 6 and 7 gives

$$v = 0.5[iR_E - V_{BE1} - V_{BE2}]$$

whereby

$$Z_{in} = \frac{\partial v}{\partial i} = 0.5\left[R_E - \left(\frac{\partial V_{BE1}}{\partial i} + \frac{\partial V_{BE2}}{\partial i}\right)\right]$$

Defining $\alpha = kT/q$, putting $V_{BE1} = \alpha \ln[(I_B - 0.5i)/I_s]$ and $V_{BE2} = -\alpha \ln[(I_B + 0.5i)/I_s]$ and differentiating V_{BE1}, V_{BE2} with respect to i, then

$$\frac{\partial V_{BE1}}{\partial i} = \frac{\alpha}{I_B - 0.5i}\left(\frac{\partial I_B}{\partial i} - 0.5\right)$$

$$\frac{\partial V_{BE2}}{\partial i} = \frac{-\alpha}{I_B + 0.5i}\left(\frac{\partial I_B}{\partial i} + 0.5\right)$$

Substituting the differentials in the expression for Z_{in} and simplifying,

$$Z_{in} = 0.5R_E + \frac{0.5\alpha}{I_B^2 - 0.25i^2}\left(I_B - i\frac{\partial I_B}{\partial i}\right) \tag{8}$$

Similarly, subtracting eqns. 6 and 7 and substituting for V_{BE1}, V_{BE2},

$$2I_BR_E + \alpha \ln\left(\frac{I_B - 0.5i}{I_s}\right) + \alpha \ln\left(\frac{I_B + 0.5i}{I_s}\right) = 2V_B$$

Under quiescent conditions $i = 0$, $I_B = I_{BQ}$, whereby

$$I_{BQ}R_E + \alpha \ln\left(\frac{I_{BQ}}{I_s}\right) = V_B$$

thus eliminating V_B and putting expression in terms of I_{BQ},

$$(I_B - I_{BQ})R_E + \frac{\alpha}{2}\ln\left(\frac{I_B^2 - 0.25i^2}{I_{BQ}^2}\right) = 0 \tag{9}$$

The differential $\partial I_B/\partial i$ can now be evaluated as,

$$\frac{\partial I_B}{\partial i} = \frac{0.25\alpha i}{R_E(I_B^2 - 0.25i^2) + \alpha I_B} \tag{10}$$

which enables the input impedance Z_{in} to be expressed as,

$$Z_{in} = \frac{R_E}{2} + \frac{\alpha}{2}\left(\frac{I_B R_E + \alpha}{R_E(I_B^2 - 0.25i^2) + \alpha I_B}\right) \tag{11}$$

For the special case where $R_E = 0$, then

$$I_B = (I_{BQ}^2 + 0.25i^2)^{0.5} \tag{12}$$

$$\left.\frac{\partial I_B}{\partial i}\right|_{R_E=0} = \frac{i}{4I_B} \tag{13}$$

$$Z_{in}|_{R_E=0} = \frac{\alpha}{2I_B} = \frac{\alpha}{2(I_{BQ}^2 + 0.25i^2)^{0.5}} \tag{14}$$

To determine the error in the general function for the total input impedance Z_x including the series resistor R_x as shown in Fig. 8, assign the nonlinear component of input resistance as $z(i)$, then

$$Z_x = R_x + Z_{in}(i = 0) + z(i) \tag{15}$$

Hence, defining a distortion factor $DZ(i) = z(i)/(R_x + Z_{in}(i = 0))$, then

$$DZ(i) = \frac{\alpha}{I_{BQ}(2R_x + R_E) + \alpha}\gamma(i) \tag{16}$$

where

$$\gamma(i) = \frac{(I_B R_E + \alpha)(I_{BQ} - I_B) + 0.25R_E i^2}{R_E(I_B^2 - 0.25i^2) + \alpha I_B} \tag{17}$$

The overall distortion level is reduced by the premultiplier of $\gamma(i)$ defined in eqn. 16 which is a function of I_{BQ}, R_x and R_E. However, for optimum selection of R_E $DZ(i)$ should be constant against i where the nonlinear function $\gamma(i)$ as defined by eqns. 17, 12 is plotted as a function of i with R_E as parameter and is shown in Fig. 9. In this example I_{BQ} = 25mA and R_E ranges from 0.02 to 1.92Ω, showing a minimum distortion factor for R_E = 0.62Ω.

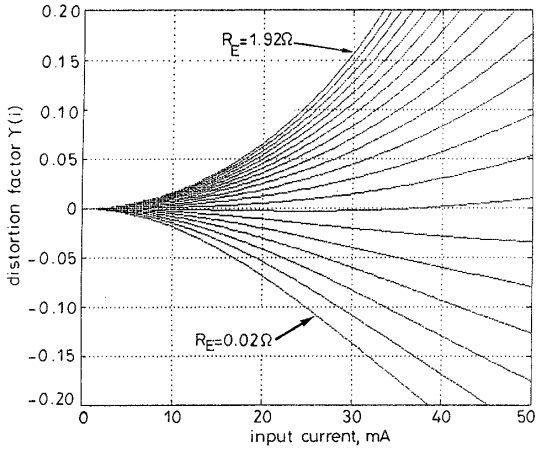


Fig. 9 *Normalised distortion characteristics for complementary emitter-coupled cell*
Circuit data: I_{BQ} = 25mA; ΔR_E = 0.1Ω; input current 0–50mA

4 Principles of distortion reduction and the application of CMAC in single-ended and balanced power amplifier

The cardinal mechanisms within an error-correction system are negative feedback, error feedback and error feedforward, where these techniques can be compounded with linearisation procedures as described in Sections 5 and 6. There has been considerable technical debate on methods of error correction [3–6] so we review only the salient features.

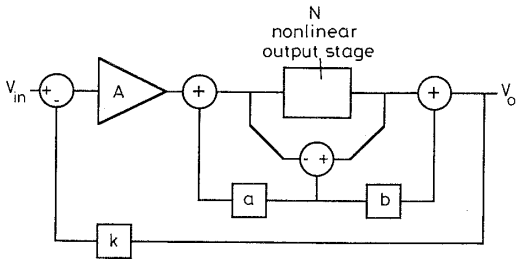


Fig. 10 *Power amplifier incorporating feedback and feedforward error correction*

For a single-ended amplifier, error correction can be applied locally to an amplifier stage as shown in Fig. 10, where N represents the nonlinear transfer function of the output stage of a power amplifier.

The closed-loop transfer function $G(f)$ of this feedback/feedforward system is given by

$$G(f) = \frac{A\left(\frac{N + bN - b}{1 - aN + a}\right)}{1 + kA\left(\frac{N + bN - b}{1 - aN + a}\right)} \tag{18}$$

To reveal the error-correction properties, define an error transfer function $E(f)$ normalised for N = 1, where

$$E(f) = \frac{G(f)}{G(f)_{N=1}} - 1$$

and

$$E(f) = \frac{(N - 1)(1 + a + b)}{1 + (kbA - a)(N - 1) + kAN} \tag{19}$$

The closed-loop gain is independent of N for

$$a + b + 1 = 0 \tag{20}$$

This expression is the balance condition of the error correction amplifier where three factors contribute to reduced distortion:

(i) making $N \approx 1$

(ii) the balance condition ($a + b + 1 = 0$) introduces a zero in $E(f)$

(iii) a large value of the product $\{k\ a\ N\}$ reduces sensitivity of $E(f)$ to variation in N although we note the limitation cited by Lipshitz and Vanderkooy [8].

The topology of Fig. 10 is only one variant but is sufficient to demonstrate the principle of compound error feedback/error feedforward [3, 9] where two special cases emerge:

(i) $a = -1$, $b = 0$ pure error feedback around N

(ii) $a = 0$, $b = -1$ pure error feedforward around N.

In practice (i) is bounded by stability constraints where, if

$$a = \frac{-1}{1 + s\tau} \tag{21}$$

then, for an optimum balance condition and complete distortion cancellation,

$$b = \frac{-s\tau}{1+s\tau} \tag{22}$$

This result demonstrates how a combination of error feedback and error feedforward can theoretically achieve total distortion cancellation even when the error feedback loop has finite bandwidth. This approach can be shown to be the basis of the Walker [10, 11] 'current-dumping' amplifier, although additional observations on the role of differential-current derived feedback (DCDF) were made in an earlier study [12].

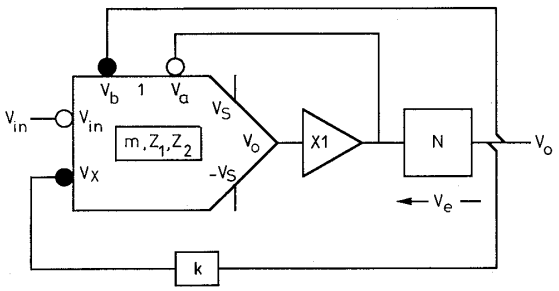


Fig. 11 *Basis error correction amplifier using CMAC*

A CMAC can be used directly within an error-feedback system, where adopting the symbology of Fig. 7, a basic single-ended amplifier is shown in Fig. 11. Alternatively, a balanced amplifier can use a CMAC with error correction where error injection can be derived either as shown in Fig. 12 or by crosscoupling between two differentially driven power amplifier, as in Fig. 13. Research by Sandman also offers additional insight into this process [13, 14].

However, observation of the crosscoupled, balanced amplifier in Fig. 13 reveals that even if N_1 and N_2 exhibit zero error, there remains a large common-mode signal applied to each error input port of the two CMACs. By combining the two configurations of error correction and introducing an additional error correction input to the CMAC as shown in Fig. 14, the common-mode signal can be cancelled contributing to lower distortion. The modified balanced power amplifier is shown in Fig. 15, where the error input ports for each CMAC have a weighting of 0.5 and the feedback factor of output to input is k.

The balance condition and operation of the power amplifier is verified as follows:

$$V_{o1} + V_{e1} = +V_{in}\left(2+\frac{Z_2}{Z_1}\right) - kV_{o1}\left(1+\frac{Z_2}{Z_1}\right) + 0.5(V_{e1}-V_{e2})$$

$$V_{o2} + V_{e2} = -V_{in}\left(2+\frac{Z_2}{Z_1}\right) - kV_{o2}\left(1+\frac{Z_2}{Z_1}\right) - 0.5(V_{e1}-V_{e2})$$

Thus, by subtraction

$$(V_{o1}-V_{o2}) + (V_{e1}-V_{e2}) = 2V_{in}\left(2+\frac{Z_2}{Z_1}\right) - k(V_{o1}-V_{o2})\left(1+\frac{Z_2}{Z_1}\right) + (V_{e1}-V_{e2})$$

that is, the error voltages cancel, whereby

$$\frac{V_{o1}-V_{o2}}{V_{in}} = \frac{2\left(2+\frac{Z_2}{Z_1}\right)}{1+k\left(1+\frac{Z_2}{Z_1}\right)} \tag{23}$$

This Section has demonstrated how error correction can be used in both single-ended and balanced amplifiers where the underlying principle is error feedforward-feedback. Even when the power amplifiers are in a crosscoupled balanced configuration each has to function simultaneously both in a feedforward mode and recursively.

5 Error-feedforward linearisation of CMAC

A local error-feedforward correction loop [4] formed around transistor $T_{2a,b}$ can be used to sense the instantaneous base-emitter voltage and inject a correction current, as shown in Fig. 16. Effectively transistors $T_{3a,b}$ and $T_{4a,b}$ form a complementary differential amplifier which in concert with the two complementary current mirrors each of current gain m, allow a correction current $(m+1)V_e/Z_e$ to flow in Z_2.

Assuming no loss of current via the output port or transistor bases then

$$V_o - V_{in} = Z_2\left(i_1 + (m+1)\frac{V_e}{Z_e}\right)$$

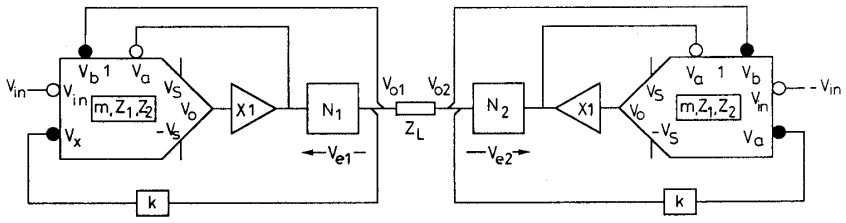


Fig. 12 *Balanced amplifier using 'normal' error correction*

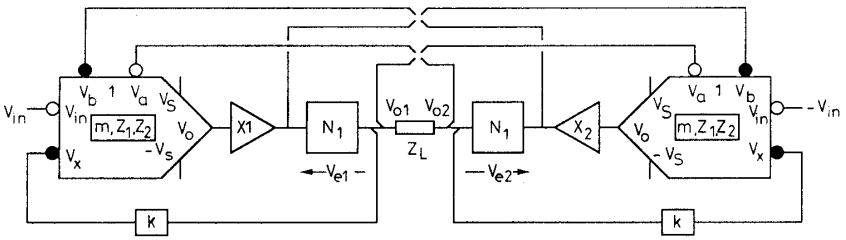


Fig. 13 *Balanced amplifier using 'crosscoupled' error correction*

and

$$V_{in} - V_x = i_1 Z_1 + V_e$$

Eliminating i_1,

$$V_o = V_{in}\left(1 + \frac{Z_2}{Z_1}\right) - V_x \frac{Z_2}{Z_1} + V_e Z_2 \left(\frac{m+1}{Z_e} - \frac{1}{Z_1}\right)$$

Hence, if

$$Z_e = (m+1)Z_1 \tag{24}$$

then dependence upon the nonlinear V_{BE} of $T_{2a,b}$ is minimised.

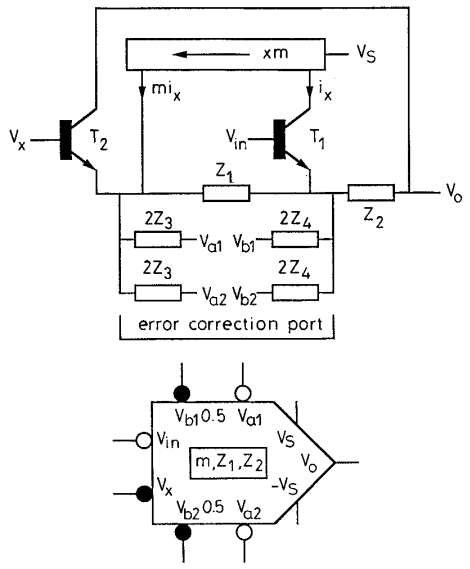


Fig. 14 *CMAC with dual error-correction ports*

In practice, a cell using the complementary configuration already exhibits low distortion as described in Section 3, however the addition of feedforward-error correction acts to further reduce distortion for applications requiring high linearity. Such correction is of course in addition to the error correction applied to the output stage.

6 Differential current-derived feedback enhanced amplifier with CMAC

Differential current-derived feedback (DCDF) [12] can also be applied to the CMAC as a means of error correction as illustrated in Fig. 17. DCDF is a method of using a current transformer to sense the time differential, of an error current that is associated with a nonlinear amplifier stage which is then used as an element within the feedback signal appearing in a first-order amplifier loop. Optimal balance of this additional feedback loop can achieve distortion cancellation. In Fig. 17 the error voltage V_e is associated with the nonlinear stage N where the derived error voltage (at the secondary of the current transformer) sL_eV_e/R_e is applied to the input of the CMAC. Analysis is straightforward and follows directly as,

$$V_o + V_e = V_{in}\left(1 + \frac{Z_2}{Z_1}\right) - \left(V_x - sL_e\frac{V_e}{R_e}\right)\frac{Z_2}{Z_1}$$

where if

$$1 = s\frac{L_e}{R_e} \times \frac{Z_2}{Z_1} \tag{25}$$

then the transfer function is independent of V_e. This is readily achieved by putting $Z_1 = R_1$ and $Z_2 = 1/sC_1$, yielding $L_e = R_e(R_1C_1)$.

7 Conclusions

This paper has considered techniques for error correction in audio power amplifiers using both single ended and balanced configurations. The principal method of error correction using feedback and feedforward was described and various power amplifier configurations considered.

An asymmetric current-steering cell was introduced designated a CMAC and its properties described. It was shown how the basic cell could be enhanced with a local feedback path using a current mirror. The CMAC offers simplicity yet through current steering and the use of unity-gain buffers, wide bandwidth and low distortion is inherent. The CMAC employs transistors in a linear configuration where the inherent properties of complementary circuitry and the operation of transistors with near constant base-collector voltages has been exploited. Where large voltage swings occur, grounded base or enhanced cascode stages [2] are recommended.

A property of the CMAC is that in the principal signal path (excluding current mirror) a single current loop is formed where effectively there is only one stage of amplification. Also it is evident that there is a direct path from input to output which is advantageous in low-gain amplifiers in terms of bandwidth and distortion. The current mirror yields a reduction in output impedance of the CMAC, but itself only has to handle signal components due to output loading and not the main signal. As such the mirror should achieve excellent linearity and can be considered as a secondary signal path operating only on demand, even if a signal is being amplified. Indeed, with appropriate design, the current mirror can function as the bias network for the main amplifier stage.

A well defined current path that does not demand signal current from the power supply can offer a similar advantage to a shunt regulator by localising the signal-current path. This enhances performance by reducing interstage coupling and the effects of signal

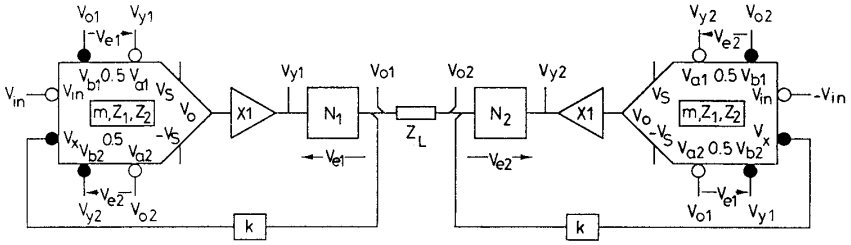


Fig. 15 *Balanced amplifier with dual error input CMAC to cancel common-mode component in error signal*

currents within the ground connection. As such the need for star grounding is less critical. The performance advantage of this stage should be compared with a two-stage amplifier that uses an inverting current mirror in the second stage, such amplifiers demand signal current from the power supply thus requiring enhanced regulation, also signal currents can circulate over wider areas of circuitry resulting in unpredictable interstage coupling.

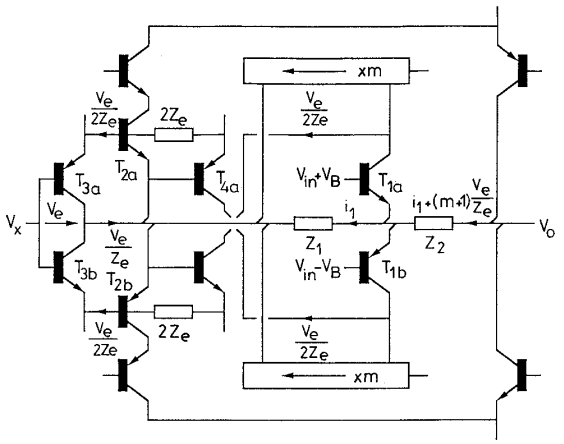


Fig. 16 *Error-feedforward linearisation of CMA*

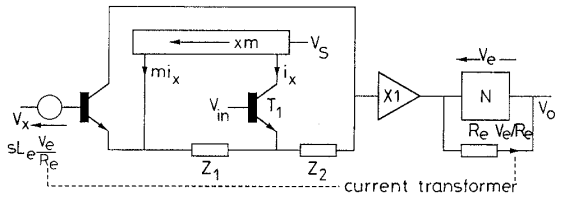


Fig. 17 *Single-ended CMAC with DCDF*

The principle of error feedforward/feedback was described and the application of the CMAC in various amplifier configurations illustrated. It was shown how the CMAC could include a difference port to accommodate the correction of output-stage distortion. Both single ended and balanced configurations were shown where the latter offered the advantage of common-mode signal cancellation in the CMAC thus improving linearity under large signal conditions.

8 References

1 ALEXANDER, M.: 'A current-feedback audio power amplifier', 88th AES convention, March 1990, preprint 290 (D-5)
2 HAWKSFORD, M.O.J.: 'Reduction of transistor slope impedance dependent distortion in large-signal amplifiers', *J. Audio Eng. Soc.*, 1988, **36**, (4), pp. 213–222
3 VANDERKOOY, J., and LIPSHITZ, S.P.: 'Feedforward error correction in power amplifiers', *J. Audio Eng. Soc.*, 1980, **28**, (1/2), pp. 2–16
4 HAWKSFORD, M.O.J.: 'Distortion correction circuits for audio amplifiers', *J. Audio Eng. Soc.*, 1981, **29**, (7/8), pp. 503–510
5 HAWKSFORD, M.J.: 'Power amplifier output stage design incorporating error feedback correction with *current dumping* enhancement', 74th convention of the AES, New York, 8–12 October 1983, preprint 1993 (B-4)
6 HAWKSFORD, M.J.: 'Distortion correction in audio power amplifiers', *J. Audio Eng. Soc.*, 1981, **29**, (1/2), pp. 27–30
7 RITCHIE, G.R.: 'Transistor circuit techniques: discrete and integrated (3rd edn.)' (Chapman & Hall, London, 1993)
8 LIPSCHITZ, S., and VANDERKOOY, J.: 'Is zero distortion possible with feedback?', 76th convention of the Audio Engineering Society, New York, 8–11 October 1984, preprint 2170 (F-4)
9 ALLINSON, N.M., and WELLINGHAM, J.: 'Distortion reduction in frequency-dependent feedback-feedforward amplifiers', *Int. J. Electron.*, 1985, **59**, (6), pp. 667–683
10 WALKER, P.J., and ALBINSON, M.P.: 'Current dumping audio amplifier', 50th convention of the Audio Engineering Society, London, 4–7 March 1975
11 VANDERKOOY, J., and LIPSHITZ, S.P.: 'Current dumping – does it really work?', *Wireless World*, 1978, **84**, (1510), pp. 38–40
12 HAWKSFORD, M.O.J.: 'Differential-current derived feedback in error-correcting audio amplifier applications', *IEE Proc., Circuits Devices Syst.*, 1994, **141**, (3), pp. 227–236
13 SANDMAN, A.M.: 'Errors – a positive approach', PhD thesis, City University, London, Nov. 1989
14 SANDMAN, A.M.: 'Who designed this?', *Electron. World*, September 1991, pp. 788–790

POWER AMPLIFIER OUTPUT STAGE DESIGN INCORPORATING ERROR FEEDBACK CORRECTION WITH CURRENT DUMPING ENHANCEMENT

1993 (B-4)

M. J. Hawksford
University of Essex
Essex, England

Presented at the 74th Convention 1983 October 8-12 New York

AN AUDIO ENGINEERING SOCIETY PREPRINT

POWER AMPLIFIER OUTPUT STAGE DESIGN INCORPORATING ERROR FEEDBACK CORRECTION WITH CURRENT DUMPING ENHANCEMENT

Dr. M. J. Hawksford, Department of Electrical Engineering Science, University of Essex, Colchester, Essex, England.

0. INTRODUCTION

The design of high performance power amplifiers has devolved into two quasi-orthogonal schools. One school exploits the use of high levels of negative feedback which in general is applied in multiple loops following the work of Cherry[1] while the second philosophy attempts to minimise overall feedback and directs attention to the optimisation of each amplifier cell using distributive local feedback and corrective strategies[2,3].

Clearly, both philosophies offer merit and are each capable of good performance where no doubt purists will attempt to draw comparisons using canonical equivalents. In this paper we address some of the problems of the low feedback school as this offers an interesting challenge in the design of low distortion amplifiers especially in applications where transistors are exercised over a wide range of their dynamic characteristics. The momentum for this work has been spurred by the belief that amplifier performance should be near aperiodic and include only a minimum of extra energy storage devices to dictate the target transfer function.

A principle advantage of distributive feedback is that the active device characteristics can determine the stability of each stage without the addition of external compensation. It is also straightforward to configure low-level signal circuitry which exhibits excellent linearity together with well behaved non-linear overload with fast recovery. An earlier paper [2] reviewed some of the advantages of low feedback amplifiers and more recent work by Ottala [4,5] has discussed the interaction of distortion and the inevitable loop delay. It is probably worth reiterating that with pure negative feedback one is attempting to correct distortion retrospectively which is a dubious process at best. It is of concern that attempts to analyse combinations of amplifier nonlinearity and non-linear band limiting mechanisms with real signals prove extremely difficult. Consequently a design approach that circumvents some of the analytical problems may well lead to a more satisfactory conclusion.

In this paper we discuss further the application of error correction to the power amplifier output cell where it is shown that the technique can reduce both voltage transfer and current transfer distortion. Since

distortion correction requires the optimisation of a balance condition to minimise distortion then the sensitivity of adjustment is fundamental to the success of the technique. The paper will also discuss methods of de-sensitising the balance condition by the expedience of current dumping [6] and local feedback where effectively multiple zeros can be generated in the error function.

The work concludes by introducing circuit topologies that aid system realisation of the enhanced strategies as well as emphasising some of the finer design detail.

1. NON LINEAR DISTORTION IN THE FOLLOWER OUTPUT CELL

Bipolar junction transistors (BJT) are fundamentally transconductance devices exhibiting a high output resistance. It is only when they are configured in the classical follower configuration illustrated in Figure 1.1 that the output impedance is low as a result of near unity gain feedback.

The follower circuit will always exhibit a gain <1 when driving a finite load impedance (excluding near instability) thus there will be a finite error voltage V_E between input and output. A further error signal will also result when the driving impedance is finite and this is compounded by signal dependency of the current gain.

It is possible to synthesise approximate distortion models though exact analysis is complex and a strategy for designing a reliable cascaded compensation network proves elusive. The problem results from the transconductance of the transistors being both emitter current dependent and temperature dependent. The thermal problem is of particular concern. It can be shown that the error voltage V_E is both dependent upon the output current I_o as well as the absolute temperature of each transistor, being particularly sensitive to the differential junction temperature. Consequently the error signal will in part include non-linear dependence on transistor thermal time constants, a problem which is aggravated when transistors dissipate high transient power under signal excitation.

In class AB amplifiers the magnitude of V_E can be further increased when one or other output devices turns off though this can be reduced by non-switching configurations [8]. It is also worth noting that even under class A operation V_E is finite though the error should tend to a linear function of the output current I_o.

The remainder of this paper addresses the problem of compensating for

the error signal V_E where we commence by summarising the following observations:-

(a) Correction is applied locally to the output stage thus minimising further intermodulation of output cell distortion and signal in stages that share a component of output distortion through feedback.

(b) Correction is only applied when $V_E \neq 0$, it therefore follows the spirit of the low feedback school.

(c) The nature of V_E including thermal dependence within bounds commensurate with error amplifier distortion, is unimportant.

(d) The technique can be used within an overall feedback amplifier to reduce transient phase modulation due to output stage non linearity.

(e) The technique, when optimised, decouples non-linear load impedances from the amplifier driving stage even if modest overall feedback is applied.

(f) The correction signals do not flow through the power supply thus aiding suppression of power supply induced distortion and simplifying the error signal path.

2. ERROR DISTORTION CORRECTION WITH FEEDBACK ENHANCEMENT

The basic topology of an error feedback correction loop has been the subject of previous reports [2,3]. One potential disadvantage of the technique is the degradation in performance that occurs through misalignment of the balance condition for minimum distortion, this can result from poor design, component aging or dynamic variation of the balance condition with signal i.e. secondary distortion within the error amplifier.

In this section we explore the sensitivity of the correction loop to alignment errors and show through the expedience of a second, local error feedback loop that the sensitivity can be reduced.

The system diagram of error feedback which is a subset of the more general feedforward-feedback structure [2] is shown in Figure 2.1 where N represents the voltage amplification of the output cell wher $N \cong 1$.

It has been shown [2] that the voltage amplification A (see Fig. 2.1) is given by

$$A = \frac{N}{(1-a) + aN} \qquad \ldots 2.1$$

where the balance condition is defined as a = 1: making A = 1.

It is constructive to decompose A into its optimum value and an error gain component A_E, where for a near unity gain amplifier,

$$A = 1 + A_E \qquad \ldots 2.2$$

Defining an error function E,

$$E = \frac{A_E}{A} = -(1-a)\left[\frac{1-N}{N}\right] \qquad \ldots 2.3$$

Differentiating E with respect to a and N, the sensitivity to misalignment and distortion can be explored, i.e.

$$\frac{\partial E}{\partial a} = \frac{1-N}{N} \qquad \ldots 2.4$$

$$\frac{\partial E}{\partial N} = \frac{1-a}{N^2} \qquad \ldots 2.5$$

We follow similar procedures in later analyses where equations 2.3, 2.4 and 2.5 form benchmarks.

Equation 2.4 and 2.5 both reveal significant sensitivity to error amplifier gain a. Consequently, moderate misalignment of a will incur a distortion penalty if $N \neq 1$, together with a secondary source of distortion if a undergoes dynamic modulation by the error signal. However, we note the favourable condition that the error amplifier loading factor[2] is minimised as $N \to 1$.

Consider next the modified structure shown in Figure 2.2 where a second error-difference loop is configured to augment the basic correction loop. We will show in Sections 3, 4 and 5 that the second loop can be implemented readily within the output stage circuitry. We also emphasise the intention that the level of feedback applied with the second loop is of modest level thus allowing wide near-aperiodic performance, with the ultimate performance bound being determined by inherent transistor characteristics.

In observing the structure of Figure 2.2, it is important to note the subtle distinctions between the error feedback loop and the secondary feedback loop. In the first loop, the correction is added prior to the distorting stage N, while in the second loop the error signal is added between the input and output nodes, it does not therefore introduce a further error function zero but simply desensitises the voltage gain from displacement in N and a.

Levels of secondary distortion are also reduced using the technique of Figure 2.2 as neither amplifier a nor amplifier b are exercised when N = 1, a condition that should be observed in the design of the output stage.

The voltage amplification A_f of the enhanced topology can be shown to be,

$$A_f = \frac{N(1+b)}{\{(1-a) + N(a+b)\}} \qquad \ldots 2.6$$

where b is the differential gain of the secondary feedback loop. Following a similar decomposition of A_f as illustrated by equations 2.2 and 2.3, we derive a modified error function E_f, where

$$E_f = - \frac{(1-a)(1-N)}{(1+b)(N)} \qquad \ldots 2.7$$

The sensitivities of E_f with respect to a and N follow as,

$$\frac{\partial E_f}{\partial a} = \frac{(1-N)}{(1+b)N} \qquad \ldots 2.8$$

$$\frac{\partial E_f}{\partial N} = \frac{(1-a)}{(1+b)N^2} \qquad \ldots 2.9$$

Compared with the results of the basic correction system a reduction in sensitivity by a factor $(1+b)^{-1}$ is realised both for dynamic variation of N with signal and for secondary distortion due to modulation of a.

3. ERROR FEEDBACK DISTORTION CORRECTION WITH CURRENT DUMPING ENHANCEMENT

The previous section showed at the system level an enhancement over pure error feedback by introducing a secondary local error feedback loop. However, earlier papers [6,7] have cited a method of error feedforward correction colloquially designated "current dumping". This technique has been pioneered by Peter Walker and has generated much technical discussion [9]. A similar derivative was later implemented by Sansui [10].

Here we combine the technique of enhanced error feedback correction and current dumping to realise a further reduction in sensitivity to system misalignment. To clarify the operation of the circuit techniques discussed in Section 5 we introduce the enhanced system using a semi-circuit

presentation as illustrated in Figure 3.1.

The error feedback circuitry uses a transconductance amplifier (g_m) in association with R_1 while transistor T_1 operates as a unity gain follower with its collector current controlling the two-output port current mirror. The current mirror sources two currents: a feedback enhancement current mI, following Section 2 and a current dumping correction current nI.

We proceed by defining parameters p, q, r where initially transistor T_1 is assumed to have zero base emitter voltage, i.e.

$$p = \frac{nR_o}{(1+m)R_2} \qquad \ldots 3.1$$

$$q = g_m R_1 \qquad \ldots 3.2$$

$$r = \frac{mR_1}{(1+m)R_2} \qquad \ldots 3.3$$

Hence assuming an output load impedance Z_L, the voltage transfer ratio A_d is given by,

$$A_d\left[1 + \frac{R_o}{Z_L}\right] = \left[\frac{1+r + p(1-q)\left(\frac{1-N}{N}\right)}{1+r + (1-q)\left(\frac{1-N}{N}\right)}\right] \qquad \ldots 3.4$$

Expression 3.4 reveals if either p or q (or both) = 1, then A_d is independent of N and the output impedance is R_o. Also misalignment or non-linear error in p and q is desensitised by r if r > 0.

Again following the technique of Section 1 (equations 2.2, 2.3) we define an error function E_{fd} for the system both with feedback and current dumping enhancement, i.e. if

$$E_{fd} = \frac{A_d\left[1+\frac{R_o}{Z_L}\right] - 1}{A_d\left[1 + \frac{R_o}{Z_L}\right]} \qquad \ldots 3.5$$

then,
$$E_{fd} = -\left[\frac{(1-p)(1-q)\left(\frac{1-N}{N}\right)}{(1+r) + p(1-q)\left(\frac{1-N}{N}\right)}\right] \qquad \ldots 3.6$$

Note that in specifying equation 3.5, the term $\left[1 + \frac{R_o}{Z_L}\right]$ was included in the expression to avoid the complication of output attenuation due to R_o, which is a linear effect dependent upon the load impedance Z_L.

The sensitivities of E_{fd} with respect to p, q and N follow as,

$$\frac{\partial E_{fd}}{\partial p} = \frac{\left(\frac{1-q}{1+r}\right)\left(\frac{1-N}{N}\right)\left\{1 + \left(\frac{1-q}{1+r}\right)\left(\frac{1-N}{N}\right)\right\}}{\left\{1 + p\left(\frac{1-q}{1+r}\right)\left(\frac{1-N}{N}\right)\right\}^2} \quad \ldots 3.7$$

$$\frac{\partial E_{fd}}{\partial q} = \frac{\left(\frac{1-p}{1+r}\right)\left(\frac{1-N}{N}\right)}{\left\{1 + p\left(\frac{1-q}{1+r}\right)\left(\frac{1-N}{N}\right)\right\}^2} \quad \ldots 3.8$$

$$\frac{\partial E_{fd}}{\partial N} = \frac{\left(\frac{1-q}{N^2}\right)\left(\frac{1-p}{1+r}\right)}{\left\{1 + p\left(\frac{1-q}{1+r}\right)\left(\frac{1-N}{N}\right)\right\}^2} \quad \ldots 3.9$$

The error function E_{fd} expressed by equation 3.6 shows the amplifier to exhibit three zeros in the domain of p, q and N which are mutually orthogonal as indicated by equation 3.1 and 3.2. The desensitisation due to (1+r) should also be observed. These results indicate the level of enhancement possible and allow a prediction of overall system non linearity to be made.

However, the analysis assumed the non-linear output stage N to exhibit an infinite input impedance. Since in practice this will not be the case, we explore in the next section methods of accommodating modulation of a finite input impedance within the overall system structure.

4. MINIMISATION OF CURRENT TRANSFER NON LINEARITY

Since the voltage transfer ratio of the output cell N will distort it follows that the current transfer ratio will also be non linear. In Figure 4.1, the non linear stage N is shown with non linear input resistance R_N.

Let us assume that the amplifier produces an undistorted output V_x (after error correction for example), it follows that the input current I_x is given by

$$I_x = \frac{V_x}{(NR_N)} \qquad \ldots 4.1$$

Consequently distortion in I_x follows from both N and R_N. In the previous sections we assumed $R_N \rightarrow \infty$, in this section techniques are developed which enable finite, non-linear variations in R_N to be accommodated.

The technique is based upon an idea first cited in an earlier paper [2], though only a basic discussion was given. In essence a small series resistor senses the input current to N, then part of this error signal is used within the compensation loop.

The simplified system topology is shown in Figure 4.2 where the voltage V_E includes both a component of the error voltage across N as well as the error voltage across the input current sensing resistor R_3.

The configuration also includes current dumping enhancement though to allow orthogonality in the error function zeros, two current mirrors are used. Non orthogonality implies that the input error current I_x would appear in the current dumping correction current nI_1, thus increasing the loading factor on this stage and leading to a more complex interactive balance condition.

The error signal V_E is defined with reference to Figure 4.2 as,

$$V_E = I_x R_3 + kV_x\left(\frac{1-N}{N}\right) \qquad \ldots 4.2$$

where $$k = \frac{R_y}{(R_x+R_y)} < 1 \qquad \ldots 4.3$$

Assuming transistors T_1 and T_2 to have zero base emitter voltage and defining parameters w, x, y, z where

$$w = kg_m R_1 \qquad \ldots 4.4$$

$$x = \frac{R_1}{R_3}\left\{\frac{k}{(1+m)(1-k)}\right\} \qquad \ldots 4.5$$

$$y = n\frac{R_o}{R_4} \qquad \ldots 4.6$$

$$z = \frac{mR_1}{(1+m)R_2} \qquad \ldots 4.7$$

then analysing the circuit of Figure 4.2 where V_x is the output of N, it follows that

$$B = \frac{V_x}{V} = \frac{(1 + z)}{(w+z) + \frac{1}{N}\left\{(1-w) + \frac{R_3}{R_N}\left\{(1-x) - \left(\frac{w-x}{k}\right)\right\}\right\}} \quad \ldots 4.8$$

If now we include the current dumping correction loop, then

$$\left(1 + \frac{R_o}{Z_L}\right)\frac{V_o}{V_{in}} = y + (1-y)\,B \quad \ldots 4.9$$

where once more following the form of equation 3.5 and defining the error function E_{fo} for the present structure then

$$E_{fo} = \frac{-\left[(1-w)\left(\frac{1-N}{N}\right) + \frac{R_3}{(NR_N)}\left\{(1-x) - \left(\frac{w-x}{k}\right)\right\}\right](1-y)}{(1+z) + y\left[(1-w)\left(\frac{1-N}{N}\right) + \frac{R_3}{(NR_N)}\left\{(1-x) - \left(\frac{w-x}{k}\right)\right\}\right]} \quad \ldots 4.10$$

Expression 4.10 shows a strong resemblance to equation 3.6 but includes the non-linear input resistance R_N. In this configuration however, there are three balance conditions to be observed where for optimum performance $w = x = y = 1$ and $z > 0$.

We note that the expressions for the balance conditions yield orthogonality between w, y and N though w and x are linked. The latter effects only cancellation of R_N and can be desensitised by the normal procedure of making $R_N >> R_3$. However, the total error is desensitised by making $z > 0$.

When the system of Figure 4.2 is close to optimum, equation 4.8 reveals $B \cong 1$ even though N exhibits both non-linear voltage transfer and non-linear current transfer, consequently the loading factor of transistor T_1 and the associated current mirror is minimised hence reducing secondary distortion. This result is important and is not inherent in conventional current dumping where the low power class A amplifier must always provide a proportion of the output current, at least with the Quad[6] configuration.

However, one problem area in configuring the second correction loop as pure current dumping is reflected in the choice of R_4 (Figure 4.2)

which can be evaluated using equation 4.6 with y = 1.

In practice R_o is small ($\cong 0.1\Omega$) to minimise the output impedance of the amplifier, which implies a low value of R_4 or a high value of n. This latter requirement has ramifications in the choice of quiescent current of the feedforward correction amplifier both with respect to its magnitude and offset stability. However, feedforward correction is particularly applicable at high frequency where feedback error correction will fail due to bandwidth constraints.

We therefore propose to modify the structure of Figure 4.2 to include both feedback and feedforward error correction where the feedforward loop will become dominant at high frequency. The new topology is shown in Figure 4.3. In this sense the parallel with the Quad approach should be observed and the similarity is acknowledged [6]. However, note the distinction, the error signal is measured directly across the output stage, thus correction is only applied when there is a voltage difference between input and output stage. In the Quad configuration, the pre-output stage amplifier is included within the differencing amplifier, thus true error feedback is not implemented as the error signal is effectively added between the differencing nodes (i.e. input and output of the dumping transistors).

In Figure 4.3, transistor T_1 and resistor R_4 yield a current I_1 which is proportional to the total output stage error voltage V_{ET}. A fraction of this current $n_1 I_1$ is then fed back to the input stage where it is combined with the input signal current $g_1 V_1$.

Thus we proceed by establishing V_o as a function of the new input V_1 but including the contribution from the secondary error feedback-feedforward loop shown in Figure 4.3.

Neglecting the small base current of R_1, then

$$\frac{V_o}{V_1} = \frac{g_1 Z_i}{\left[1+\frac{Z_o}{Z_i}\right]} \cdot \left[\frac{BR_4 + n_2 Z_o(1-B)}{R_4 - n_1 Z_i(1-B)}\right] \qquad \ldots 4.11$$

Defining an error function E_{f1} and again following the form of equations 2.2, 2.3 then,

$$E_{f1} \cong \frac{-(1-B)\left[1 - \frac{(n_1 Z_i + n_2 Z_o)}{R_4}\right]}{\left[1 - \frac{n_1 Z_i}{R_4}(1-B)\right]} \qquad \ldots 4.12$$

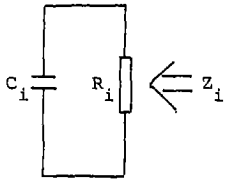
where B is the gain of the enhanced output stage and is defined by eqn. 4.8.

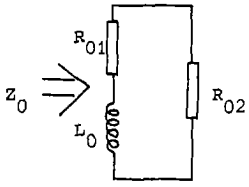
Equation 4.12 shows the balance condition of the secondary loop to be given by,

$$R_4 = (n_1 Z_i + n_2 Z_o) \qquad \ldots 4.13$$

Following Quad [6] we acknowledge that Z_o can be made inductive which in turn can be compensated for by Z_i being partly capacitive. Using these components a transition region between error feedback and error feedforward becomes possible which is advantageous for both correcting high frequency distortion and for maintaining a low output impedance.

As an example, suitable choices for Z_i and Z_o are illustrated below,





whereby given $n_1 n_2$, R_i and R_4

$$R_{01} = \frac{R_4(R_4 - n_1 R_i)}{n_1 n_2 R_i} \qquad \ldots 4.14$$

$$R_{02} = \frac{R_4}{n_2} \qquad \ldots 4.15$$

$$L_o = \frac{C_i R_4^{\,2}}{n_1 n_2} \qquad \ldots 4.16$$

This section has illustrated how a unity gain follower output stage may be linearised by using an enhanced error correction topology. We can assess the effectiveness of this strategy by noting the cluster of zeros that appear in the error function E_{f1}, eqn. 4.12, where orthogonality is advantageous both from design optimisation and from the effects of secondary distortion due to non linearity in the correction amplifiers. In particular, with the secondary error feedback-feedforward structure of Figure 4.3, the low loading factor of the optimised B amplifier should be observed together with the ability to limit the bandwidth of the error feedback loop for good stability yet extend the high frequency correction using the wide band feedforward loop.

In the following sections we illustrate how these techniques could be implemented and suggest their possible application in high efficiency amplifier configurations.

5. CIRCUIT TOPOLOGY OF UNITY GAIN POWER OUTPUT CELL

Several derivatives of error correction topologies have been introduced in the previous sections. Here we attempt to translate these ideas into basic circuitry and to discuss further techniques for enhancing performance.

The target is to implement a system based upon Figure 4.2, together with the enhancement of Figure 4.3. We will proceed by discussing an implementation of Figure 4.2 that excludes for clarity the current dumping feedforward correction loop (T_1,R_4,R_o,n).

Transistors T_2 and T_3 form a Darlington buffer to reduce loading on the pre-output stage and to supply drive current via R_1 and R_3 to the Darlington output transistors T_7. Consequently transistors T_2 supply significant current to the bases of T_7 and are a source of distortion. To minimise V_{BE}-I_e non linearity of T_2, the transistors are enclosed within a local feedforward correction loop formed by T_3 and T_4 where T_3 seconds as the first stage of the Darlington. However, we note that transistors T_2 and T_3 are enclosed within a local negative feedback loop using the(Xm)current mirrors, this further linearises T_2 and reduces the effective source impedance presented to the output transistors T_7. The resistor R_1 allows the error correction signal derived from the output cell to be superimposed on the input signal.

The main output transistors are shown operating at low collector base voltage which remains approximately constant due to the transistors T_8. This mode of driving decouples supply rail variation reduces power dissipation and minimises bias drift and thermally related distortions in the output transistors T_7. Since T_7 are operated at low collector base voltage, fast output transistors may be selected which can be placed in close physical proximity to the drive circuitry.

The output transistors are biased using transistors T_6 in an "amplified diode" configuration where the base emitter voltage of these two transistors forms the reference voltage. However, transistors T_6 together with T_5 also form the error amplifier for the output stage where their collector currents are circulated through resistors R_1 to add in the error correction signal. The advantage of this technique is that the error signal does not demand extra current from T_2, which was a feature of an earlier circuit [2]. Distortion in the driver stage is therefore reduced as well as keeping the signal paths both simple and local. It should be noted that the inclusion

of resistors R_3 allows current transfer non linearity to be compensated as discussed in Section 4. To enable the stability of the error correction feedback loop to be maintained, an impedance Z_C (parallel combination of inductance/resistance) is introduced into the emitter circuit of the T_5, T_6 difference amplifier. The problem of stability within an error feedback loop is given consideration in the Appendix.

Finally, we observe a further local feedback loop from the output node to emitters of T_2 via resistors R_2, this represents the enhancement discussed in Section 2 and is illustrated in the simplified diagram of Figure 3.1. The resistors R_2 also form a small but welcome feedforward signal path directly across the output stage.

To complete the output cell using a similar structure to the system of Figure 4.3, we include the secondary distortion correction enhancement scheme illustrated in simplified form in Figure 5.2, this includes the circuitry of Figure 5.1 though detail has been omitted for clarity.

Inspection of Figure 5.2 and Figure 4.3 will reveal the similarity where the current gains of the mirrors follow as,

$$n_1 = \frac{R_{11}R_{14}}{R_{12}(R_{11}+R_{13}+R_{14})} \quad \ldots 5.1$$

$$n_2 = \frac{R_{14}(R_{11}+R_{13})}{R_{15}(R_{11}+R_{13}+R_{14})} \quad \ldots 5.2$$

It is possible to increase the efficiency of the output stage by using an array of multiple supply rails and a form of Class C cascodé as illustrated in Figure 5.1 (transistors T_9). This effectively reduces the maximum collector base voltage of each output transistor thus minimising both secondary breakdown and power dissipation. The technique was first reported in [11] and later employed by Carver in the "cube" amplifier. Since the cascode connection appears outside the correction loop, transient distortion, when devices commutate should be controlled.

6. CONCLUSIONS

This report has presented extensions of the techniques of error correction in analogue power amplifiers where it was shown that local feedback with a secondary orthogonal correction loop can significantly desensitise the amplifier to misalignment.

A method of reducing current transfer distortion was cited. This method effectively suppresses the non linear image of the loudspeaker impedance as seen through the output cell and therefore removes the loudspeaker from the global amplifier circuit. It is important to note that this suppression was readily implemented within the voltage transfer correction loop with minimal cost penalty.

The report further emphasised the link between error feedback and error feedforward where the deployment of the two procedures was frequency selective: feedback being dominant at low frequency with a gentle cross-over to feedforward at high frequency. With correct choice of linkage the error correction capabilities become aperiodic with minimal degradation due to bandlimitation of the feedback correction loop.

Analysis introduced the error function description of an error correction amplifier and showed directly the potential impairment due to malalignment of the balance condition. The positive advantages of multiple zeros in the error function could then be assessed together with the importance of orthogonality both from system design and system optimisation criteria.

Finally, introductory circuitry was presented which described methods of realisation. The circuits included enhanced distortion correction strategy together with further local correction to enhance linearity. Methods of improving efficiency and decoupling supply rail variations were also briefly discussed.

It is hoped the paper will further stimulate interest in the application of corrective procedures. The circuits are not intended as definitive only as vehicles for illustration and discussion. The author suggest these techniques form a realistic alternative to the high feedback school and significantly extend what can be achieved by simple local negative feedback.

7. REFERENCES

1. Cherry, E.M., 'A new result in negative feedback theory and its application to audio power amplifiers', Int. J., Circuit Theory Appl., Vol. 6, pp. 265-288, July 1978.

2. Hawksford, M.J., 'Distortion correction in audio power amplifiers', J.AES, Vol. 29, No. 1/2, Jan/Feb 1981.

3. Hawksford, M.J., 'Distortion correction circuits for audio amplifiers' J.AES, Vol. 29, No. 7/8, July/August 1981.

4. Otala, M., 'Feedback-generated phase modulation in audio amplifiers', presented at the 65th Convention of the AES, London, March 1975, preprint 1576.

5. Otala, M., 'Phase modulation and intermodulation in feedback audio amplifiers', presented at the 68th Convention of the AES, May 1981, preprint 1751.

6. Walker, P.J., and Albinson, M.P., 'Current dumping audio amplifier', presented at the 50th Convention of the AES, London, March 1975.

7. Walker, P.J., 'Current dumping audio power amplifier', Wireless World, Vol. 81, pp 560-562, Dec. 1975.

8. Tanaka, S., 'New biasing circuit for class B operation', J.AES, Vol. 29, No. 3, pp 148-152, March 1981.

9. Vanderkooy, J., and Lipshitz, S. P., 'Feedforward error correction in power amplifiers', J.AES, Vol. 28, pp 2-16, Jan/Feb 1980.

10. Takahashi, S., and Tanaka, S., 'Design and construction of a feedforward error-correction amplifier', J.AES, Vol. 29, No. 1/2, Jan/Feb 1981.

11. Thornton, R.D., Linvill, J.G., Chenette, E.R., Ablin, H.L., Harris, J.N., Boothroyd, A.R., Willis, J., and Searle, C.L., 'Handbook of basic transistor circuits and measurements', Semiconductor electronics education committee/Vol. 7, John Wiley and Sons inc., pp 26-32, 1966.

Appendix. Stability analysis of error feedback including equivalent topologies

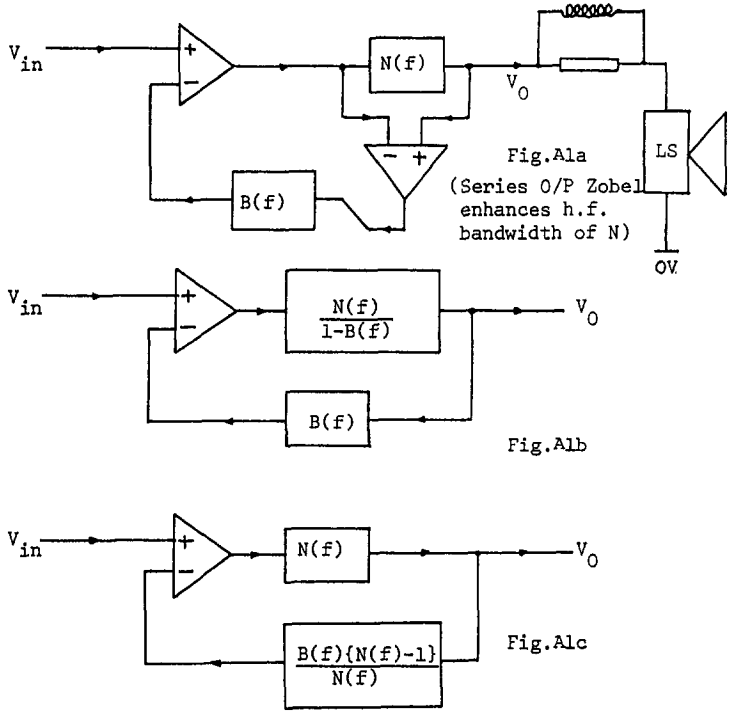


Figure A1 - Equivalent error feedback configurations

By analysis the closed loop gain G(f) and loop gains, $H_1(f)$, $H_2(f)$, follow, where:

$$G(f) = \frac{N(f)}{\{1-B(f)\}+B(f)N(f)} \quad \ldots\ldots A1$$

$$H_1(f) = \frac{N(f)B(f)}{B(f)-1} \quad \ldots\ldots A2$$

$$H_2(f) = B(f)[1-N(f)] \quad \ldots\ldots A3$$

It is interesting to compare the system in Figure A1b with Figure A1c. In the former we ideally require infinite gain in the forward path ($B(f) \approx 1$) whereas in the latter we have zero gain in the feedback path ($N(f) \approx 1$), yet both generate the same closed loop gain. On an initial consideration Figure A1b appears impractical whereas Figure A1c would appear eminently acceptable.

However, if we test for instability by equating the loop gain to unity then both equations A2 and A3 result in the condition:

$$B(f)[1-N(f)] = 1 \qquad \ldots\ldots \text{A4}$$

(for oscillation).

The output stage N(f) is, within the context of a power output stage, a near unity gain amplifier (e.g. a complementary emitter follower output stage with local negative feedback through emitter degeneration). However, the transfer function of N(f) will depend upon the load impedance (loudspeaker). For good control of the closed loop gain it is desirable to maximise the bandwidth of N(f). Basically this will be limited by the f_T of the output transistors, however, by use of a series output Zobel network see Fig.A1a, (and/or feedforward across N(f)) the bandwidth can be extended such that to a good approximation,

$$N(f) = \frac{N_0(1 + j\frac{f}{f_3})}{(1 + j\frac{f}{f_1})(1 + j\frac{f}{f_2})} \qquad \ldots\ldots \text{A5}$$

The error amplifier can then be designed using wide bandwidth circuitry with local negative feedback and compensation (see circuit diagram in Fig.5.1) such that B(f) is a well defined first-order network where:

$$B(f) = \frac{B_0}{(1 + j\frac{f}{f_4})} \qquad \ldots\ldots \text{A6}$$

Generally $f_1 \approx f_T$ where f_3, $f_4 > f_1$. We also note that B_0 is close to unity, though in analysis account must be taken for values ranging, say, from 0.8 to 1.2 for stability and amplitude peaking under conditions of poor adjustment. From equations A4, A5 and A6 we deduce the following criteria for loop stability:

Let f_n be the natural frequency of oscillation at unity loop gain;
N_L the lower bound to N_0,
N_u the upper bound to N_0.

1. For dc instability ($f_n = 0$)

$$N_L = (\frac{B_0-1}{B_0}) \qquad \ldots\ldots \text{A7}$$

2. For ac instability ($f_n > 0$)

$$f_n = \{f_1 f_2 (1+N_0 B_0 \frac{f_4}{f_3}) + f_4(f_1+f_2)(1-B_0)\}^{\frac{1}{2}} \qquad \ldots\ldots \text{A8}$$

$$N_u = \left[\frac{(\frac{f_1+f_2}{B_0 f_4})\left[1+\frac{f_4^2}{f_1 f_2}(1-B_0)(1-B_0)+(\frac{f_1+f_2}{f_4})\right]}{\left[1-\frac{1}{f_3}\{f_4(1-B_0)+(f_1+f_2)\}\right]}\right] \qquad \ldots\ldots \text{A9}$$

Conditions for stability:

Condition A $\quad N_u > N_L$

Condition B $\quad N_L < N_0 < N_u$

Condition C $\quad$ If $f_4 \to \infty$ then $f_3 < \left[\frac{N_0 B_0}{(B_0-1)(\frac{1}{f_1}+\frac{1}{f_2})}\right]$, $(B_0 > 1)$

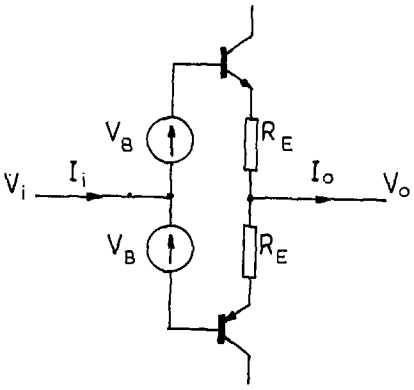
In practice, N_0 is just less than 1 (although it can be non-linear) and $(1-B_0) = 0$ to minimise distortion. For optimum performance it is important to maximise f_3 and yet keep high frequency peaking in the closed loop gain to within tolerable bounds. As an example, the following results are computed for a simulated example where:

$f_1 = 1\text{MHz}$, $f_2 = 20\text{MHz}$, $f_3 = 5\text{MHz}$

Table of Results ($P(N_0=x)$ => max. gain at $N_0 = x$).

f_4 MHz	$B_0=1$ ($N_L=0$)			$B_0=1.2$ ($N_L=0.17$)		
	$f_n(N_0=1)$ MHz	$P(N_0=0.8)$ dB (approx)	$P(N_0=1)$ dB (approx)	$f_n(N_0=1)$ MHz	$P(N_0=0.8)$ dB (approx)	$P(N_0=1)$ dB (approx)
1	4.90	1.1	1.9	4.54	2.3	2.9
5	6.32	2.8	3.2	4.80	8.2	7.0
10	7.75	3.1	3.2	5.10	14.5	11.2
20	10.00	3.1	3.0	5.66	21	18.5
40	13.42	2.9	2.7	6.63	21	17.8

The results suggest that providing the bandwidth of N(f) is maximised then stability problems are minimal (practical circuit designs also corroborate this conclusion).



V_i , input voltage
V_o , output voltage
I_i , input current
I_o , output current
V_B , bias voltage

Fig. 1-1 Follower configurations using complementory transistors

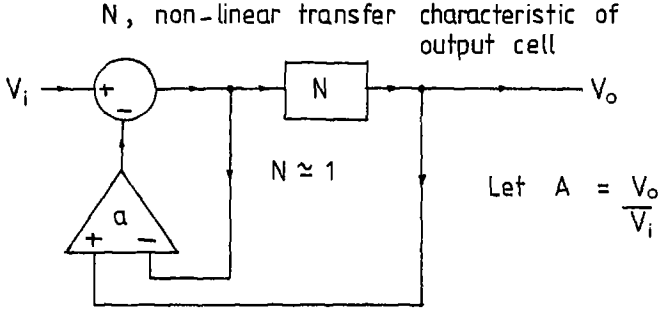


Fig. 2-1 Basic error feedback correction system

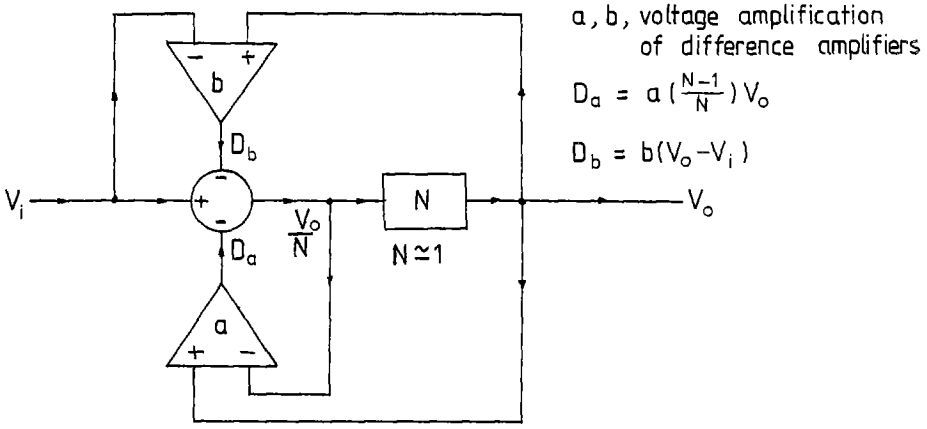


Fig. 2-2 Error correction feedback enhanced by secondary local feedback

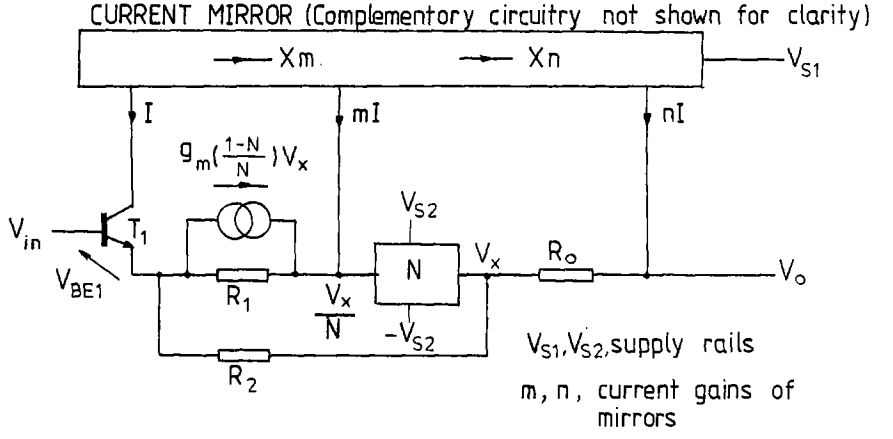
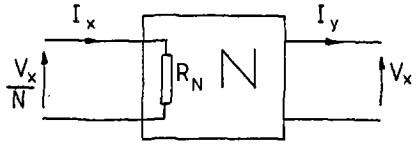


Fig. 3-1 Semi-circuit format of error feedback with current dumping enhancement



N , non-linear voltage transfer
R_N, non-linear input resistance
I_x, input current
I_y, output current
V_x, output voltage

Fig 4-1 Non-linear output cell

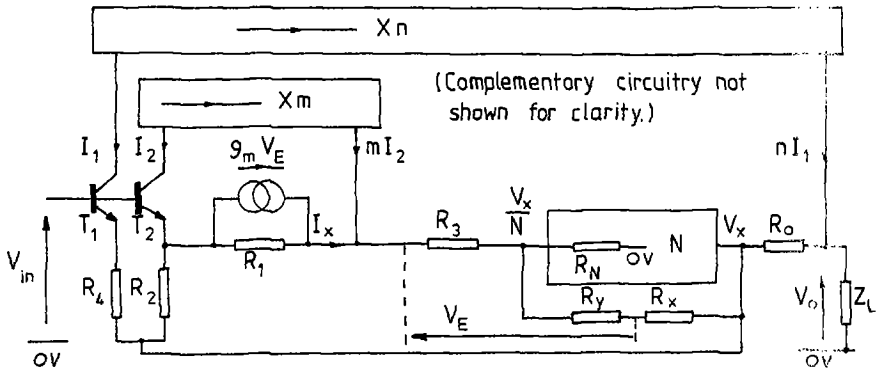


Fig. 4-2 Enhanced distortion correction with input error current compensation and current dumping

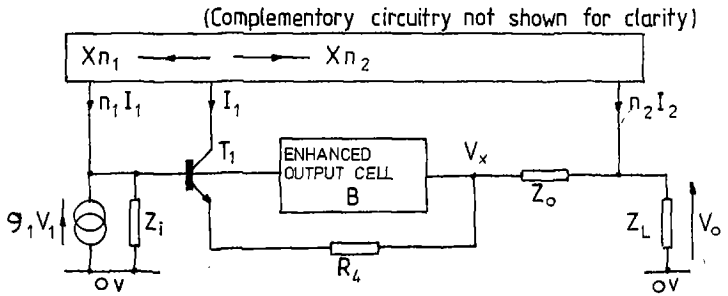


Fig. 4-3 Enhanced distortion correction using a secondary feedforward - feedback error correction loop

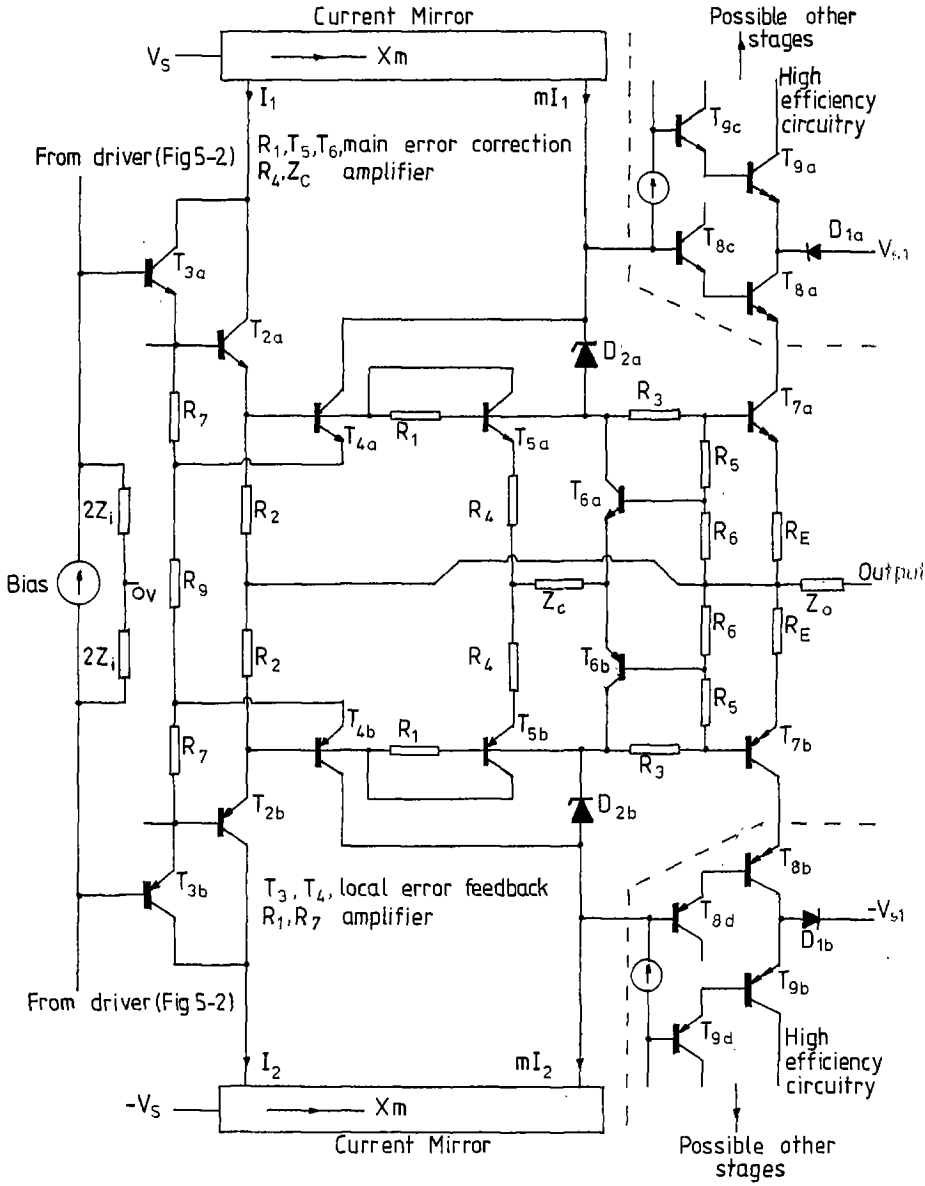


Fig. 5-1 Output cell with enhanced error feedback, input current compensation and high efficiency cascode circuitry

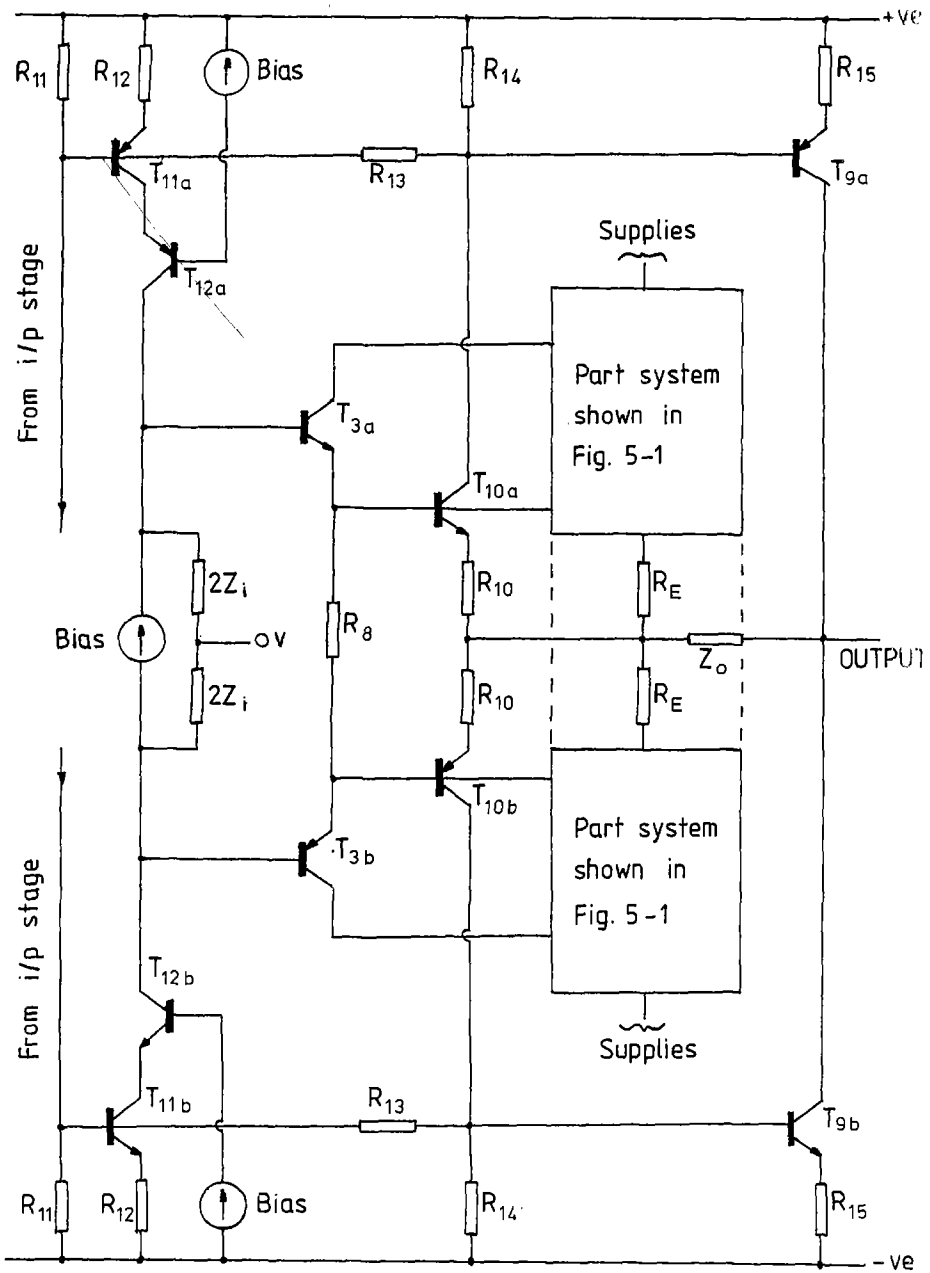


Fig. 5-2 System of Fig. 5-1 with further feedback - feedforward enhancement re Fig. 4-3

PONTOON AMPLIFIER CONSTRUCTIONS
INCORPORATING ERROR-FEEDBACK
LOCATION OF FLOATING POWER SUPPLIES

2247 (A-14)

Malcolm John Hawksford
University of Essex
Department of Electrical Engineering Science
Colchester, Essex, UK

Presented at the 78th Convention 1985 May 3-6 Anaheim

AN AUDIO ENGINEERING SOCIETY PREPRINT

PONTOON AMPLIFIER CONSTRUCTIONS INCORPORATING ERROR-FEEDBACK LOCATION OF FLOATING POWER SUPPLIES

Dr M.J. Hawksford

Department of Electrical Engineering Science
University of Essex
Wivenhoe Park
Colchester
Essex CO4 3SQ
England

ABSTRACT

A redefinition of the role of the line amplifier and power amplifier is forwarded as a more optimum construction for both minimising power supply related error signals and improving system transparency, within an economic framework. A methodology offering a quasi-bridge and a symmetric, differential bridge structure incorporating floating supplies (pontoon amplifier) complements this new proposal, and a novel topology using error-feedback location of the floating supplies is presented.

This paper was prepared as an on-going study on amplifier systems for:

FAMCO
26 Rue Saint-Louis-en-L'Ile
75004 Paris
France

0 Introduction

Traditional power amplifiers are configured to operate using either single or dual power supplies, with a node referenced to signal ground. Usually an amplifier is designed to exhibit an input sensitivity in the region 500mV to 2V, to attain a full output voltage across the loudspeaker. We therefore have a situation where high sensitivity circuitry referenced to ground is combined with high power circuitry where, for example, ground line contamination, power supply injection and induced interference from current carrying conductors bounds system transparency.

Audition of power amplifier circuits reveal significant differences in performance, especially where demanding loads (loudspeaker impedance) are encountered. Examination of the more successful power amplifiers reveals subjective performance to exhibit a greater correlation with factors related to power supply and ground rail design, than to absolute choice of amplifier topology, providing linear and non-linear distortion performance is within acceptable bounds.

In this paper, we take to task the conventional power amplifier construct and proffer a philosophy based upon unity-gain cells as being a more optimum and rational approach. We examine also power supply methodology and, in particular, review the application of bridge amplifiers and floating supplies as a method of reducing power supply injected distortion and ground rail contamination.

1 Unity-gain, power buffer and line amplifier functions

The development of compact disk (CD) has seen disk players offering peak output signals typically in the region of 2V, a value significantly greater than that offered by many preamplifiers. Indeed, it is argued that the traditional preamplifier is obsolete with CD players, where we observe a proliferation of "passive preamplifiers" as a more optimum interface between CD player and power amplifier. This approach is, in principle, most welcome as the reduction of redundant circuitry only enhances overall system transparency.

Since the traditional line amplifier in audio reproduction equipment no longer offers a useful function within the primary signal channel, it is reasonable to redefine its role and the move towards a construct that realises a more optimum performance. The methodology proposed for this achievement is as follows:

* high-gain, low-level circuitry either referenced or partially referenced to ground should be removed from the main power amplifier system

* power amplifier is designed as a unity-gain, power cell with circuitry floating at signal level, where supply structures are optimised for minimum error-signal induction

* supplies are of floating design, again to minimise error-signal induction

* line amplifier role is redefined to include the voltage amplification previously supported by the power amplifier. The line amplifier is now a voltage amplifier with high output voltage swing and is designed to drive only a unity-gain, power buffer

* line amplifier is geographically remote from the power buffer and has local optimised supplies

* unity-gain power buffer includes local error correction feedback, possibly with error reduction feedback (1,2) to reduce sensitivity to misalignment of balance condition(s)

  (no circuitry is referenced to ground; it floats with the signal)

* high voltage line amplifier is not embedded in an overall feedback path from power output cell; its function is orthogonal and its operation mode class A

* symmetric, differential, unity-gain power buffers driven by a differential line amplifier, the preferred construct; an amplifier construction we designate "a pontoon amplifier".

Efficiency targets generally dictate that output power cells operate in class AB; consequently the currents drawn by the output and driver transistors are a gross distortion of the input signal, often resembling a half-wave rectifier function (3). The bandwidth of these currents is therefore much in excess of the audio band and represents gross distortion. Clearly, any induction to high sensitivity circuitry, either by magentic coupling, ground line contamination or non-linear modulation of supplies, will cause an error signal to be embedded within the primary signal. Such error will be more problematic under complex transient excitation with difficult loads than the more traditional steady-state evaluation test functions.

It is suggested that the principle advantage of class A (and to a lesser extent, the non-switching, soft distortion cells (4)), is that the currents within the amplifier are now a more linear function of the input signal, without rapid switching edges. Thus, we see primarily linear error rather than transient non-linear error, generated within class AB.

Consequently, for a transparent class AB output amplifier, the supply must be isolated from the more sensitive voltage amplifier, and we must seek to minimise ground rail contamination and magnetic induction. Also, for an optimum interface, the input impedance of the power buffer should be high, constant and not represent a reflection of the loudspeaker or be distorted by non-linear currents within the output cell.

[In principle, as an intermediate step in optimising an audio system, a conventional power amplifier forms the line amplifier function (where linear, class A operation is assured through minimal load currents), while a unity-gain, power buffer, in geographical isolation, drives the loudspeakers. Ideally, the power buffer is placed in close proximity to the loudspeaker system to minimise cable/loudspeaker related error signals.]

Fig.1.1 illustrates a basic system layout shown for CD analogue disk and tuner applications (by way of example).

Our principle objective is to present a design philosophy for a unity-gain, power buffer based system, by drawing on previous work on error feedback distortion reduction (2) and, in particular, exploring power supply structures that minimise contamination of ground rail robustness and of primary signal by using a pontoon amplifier construction.

2 Bridge amplifier construction

A detailed and excellent paper by Greiter (5) discussed several topologies for realising a bridge amplifier using transistor devices and there has since been many adaptations and developments.

Before proceding to introduce a quasi-bridge, floating topology with error feedback supply location, we will, as a preamble, revise some of the principle advantages of a conventional bridge amplifier, where Fig.2-1 shows two possible system layouts, together with an indication of supply currents and PSU loading.

Basically, a bridge amplifier consists of two differentially driven power amplifiers connected with the load (Z_L) across the two outputs, thus each amplifier has a load impedance of $Z_L/2$. The main advantage of this scheme is that the output voltage is doubled for a given supply rail voltage. However, closer examination reveals other advantages and follow an earlier communication (6,7). These factors are now reappraised in brief:

2.1 Reduction of slew rate and TID

System examination reveals that for a given output voltage level, the slew rate of each amplifier is one-half that of a single-ended amplifier producing the same output level. This factor significantly enhances TID performance (8).

2.2 Ground rail contamination

Since large load currents are not referenced to ground, ground rail contamination is greatly reduced. This factor is significant, as supply currents (see comments in Section 1) are often gross distortions of the primary input signal (3).

2.3 Secondary breakdown and output current demand

For a given output voltage swing, a lower supply voltage is required, thus secondary breakdown of BJT transistors is reduced, realising a greater on-demand current capability. Also, due to lower operating power levels and transistors being effectively in series, again, higher output currents are achievable. This point is important where complex loads are driven due to high transient current demand: a topic of contemporary interest (3). Thus, bridge structures represent a more optimal methodology of achieving higher output power levels than simply increasing supply voltage directly.

2.4 Reduction in common-mode supply induced distortion

Fig. 2-1 cites two approaches to bridge amplifier construction. Initially (a) appears the more attractive, as there is essentially lower current demand on each of the four supply rails (V_{S1}, V_{S2},

$-V_{S1}$, $-V_{S2}$). However, there is a greater advantage to be sought in using a common supply for both halves of the bridge amplifier, that is related to symmetrical supply loading and common-mode supply rejection.

Let us assume, for purposes of discussion, a sinusoidal load current drawn from two class B amplifiers. Fig. 2-1(a) reveals that the currents in each amplifier supply line are "half-wave rectified" and there is a time displacement of a half period between positive and negative supply loadings. Thus the supply voltages on opposing positive and negative supplies will always behave disimilarly.

However, when a single supply is used, that is common to both halves of the bridge, as shown in Fig. 2-1(b), symmetrical loading of the supply rails follows by default. Consequently, if we observe the positive and negative supply waveforms for any general input function, they are mirror images, and this includes the ripple component. Note that this is true for class B operation, even when the loading of supplies is grossly non-linear. As an example, Fig.2-2 illustrates an exaggerated possible supply voltages on the positive and negative rails.

Consequently, for a more optimum performance with this configuration, the two amplifiers should be designed to offer low but symmetrical supply rejection characteristics. Thus, if δ_+ represents injection from the positive rail and δ_- represents injection from the negative rail, assumed identical, for both amplifiers, then referring to Fig. 2-1(b)

$$V_{A2} = A\,V_{in} + (\delta_+ + \delta_-) \quad \ldots \quad 2.1$$

$$V_{B2} = -A\,V_{in} + (\delta_+ + \delta_-) \quad \ldots \quad 2.2$$

$$\therefore \quad V_O = V_{A2} - V_{B2} = 2A\,Vin \quad \ldots \quad 2.3$$

Examination of equations 2.1, 2.2 and 2.3 reveal that there are three methods by which rail injection is minimised:

(i) Design amplifiers to minimise δ_+, δ_- injection

(ii) symmetrical amplifiers and common supply make $\delta_+ \simeq -\delta_-$ (hence $\delta_+ + \delta_-$ minimised)

(iii) bridge configuration reduces finite $(\delta_+ + \delta_-)$ by common mode rejection, as output signal is differential.

From this discussion, we conclude an optimum bridge amplifier uses a common supply for both amplifier circuits and not, as may appear initially preferable, a separate supply for each half of the bridge.

2.5 Output impedance

A disadvantage of a bridge amplifier is an apparent doubling of output impedance. However, where error reduction feedback is used, a low and broad-band impedance characteristic results, which negates this criticism.

2.6 Theoretical enhancement in SNR

If the two halves of a bridge amplifier offer identical but uncorrelated noise levels, then at the bridge output port, the signal adds coherently, while the noise combines as a power addition, thus for a given individual amplifier excitation we see a theoretical improvement of 3dB in SNR.

3 Floating quasi-bridge construction: quasi-pontoon amplifier

3.1 Floating topology and supply location

This Section considers an extension to the traditional bridge amplifier (discussed in Section 2), that is particularly relevant to unity-gain, power amplifiers, where we apply and extend an amplifier construct originally proposed by Brady in a Wireless World article (10,11), and further developed by Takahashi and Tanaka (12).

Principle advantages of this new topology are the minimisation of circuitry in the critical signal path and the more efficient use of the PSU, together with the maintenance of single-ended operation. The system employs a floating supply and offers most of the advantages discussed in Section 2.

The primitive circuit topology is illustrated in Fig. 3-1, where the unity-gain cell is shown in elemental form as a complementary emitter follower, to describe the circuit function.

The supply (V_S), by default, is isolated from preceding stages of the amplifier. Also, changes in V_S inject minimal effect on the output (V_O), though imbalance in T_1 and T_2 may result in slight error. The PSU (V_S) is held in position by the two impedances Z_x, having preferably a moderate value at dc (to reduce supply loading) and a value $<<Z_L$ at signal frequencies.

However, to maximise the output voltage swing, the impedances Z_x must change dynamically to enable the output voltage to approach $\pm V_S$, thus matching the performance of a conventional bridge, yet allowing single-ended operation with the "cold end" of the loudspeaker and interconnect referenced to ground. This operation is important, as the loudspeaker voltage equals the input voltage, being defined by a single, unity-gain power buffer, unlike the conventional bridge, where two amplifiers appear in the main signal path, in series with the loudspeaker.

To achieve true floating operation of the PSU, the location of the supply must be accurately steered with respect to ground, to both maximise the output voltage swing and to evenly distribute the power dissipation throughout the available power transistors. These criteria are realised by an enhanced floating PSU topology using error-feedback supply location, where a circuit primitive (without transistor bias circuitry) is shown in Fig.3-2. We designate this construction: a quasi-pontoon amplifier.

The theoretical operation of the circuit, together with the balance condition for optimal supply location and minimum distortion are established in the following analysis:

3.1.1 Class A bias of T_1, T_2 and T_3, T_4 transistors

Assume the transistors T_1, T_2, T_3 and T_4 are biased into linear operation (Class A) and the supply V_S is symmetrically positioned about ground, but with an offset of ΔV_S (i.e. positive and negative rails)are

$$\left[\frac{V_S}{2} + \Delta V_S\right], \left[\frac{-V_S}{2} + \Delta V_S\right] \text{ respectively.}$$

The bases of T_1 and T_2 form the PSU location control input, V_{il}, where in this example

$$V_{il} = 0.5\ V_o$$

which is determined by the two resistor divider chain (R_o) placed between V_o and ground (see Fig.3-2). We note $V_o = V_{in}$.

The emitter potentials of transistor pairs T_1, T_2 and T_3, T_4 are assumed to match their respective base potentials, where the effect of transistor r_e s are accounted by inclusion in resistor R_3, which takes the designation R_3*.

The objective is to determine a relationship that sets $\Delta V_S = V_{il} = 0.5V_o$ (in this example).

Neglect base currents and define the base potential of transistors T_3, T_4 as V_x, (also see Fig.3-2 for definitions of I_o, I_1, I_2 and I_3).

$$\therefore \quad V_x = \frac{1}{2}\left[(0.5V_S + \Delta V_S) - (I_1 + I_o)R_1 \right.$$

$$\left. +(-0.5V_S + \Delta V_S) + (I_2 + I_o)R_1 \right]$$

i.e. $V_x = \Delta V_S - 0.5\ (I_1 - I_2)R_1$ ... 3.1

Observing the emitters of T_1, T_2 and T_3, T_4

$$0.5V_o - V_x = I_3\ R_3^* \quad \ldots \quad 3.2$$

where, $I_3 = I_1 - I_2$ ... 3.3

Hence, from equations 3.1, 3.2 and 3.3

$$\frac{2}{R_1}\ (\Delta V_S - V_x) = \frac{1}{R_3}\ (0.5V_o - V_x) \quad \ldots \quad 3.4$$

Examination of equation 3.4 reveals a balance condition, whereby the relationship is independent of V_x and ΔV_S is optimally related to V_o (also in V_{in} and more generally, V_{il}),

i.e. if, $R_1 = 2R_3^*$ ... 3.5

then, $\Delta V_S = 0.5V_o$ ... 3.6

or more generally, $\Delta V_S = V_{il}$ ... 3.7

3.1.2 Class B bias of T_1, T_2 and T_3, T_4 transistors

A similar analysis is presented in this sub-section, but here the location transistors are biased in Class B.

Assume transistors T_1, T_4 are conducting and both carry a current I_x, while transistors T_2, T_3 are turned off. The circuit then simplifies to that shown in Fig.3-3, where non-conducting transistors are omitted.

The analysis neglects base currents, but assigns base-emitter voltage V_{BE1}, V_{BE4} to transistors T_1, T_4 (where, $V_{BE1} \simeq -V_{BE4} \simeq 0.7V$).

Using Fig.3-3 as reference, the following network equations are derived:

$$V_{il} - (V_{BE1} - V_{BE4}) - V_x = I_x R_3 \quad \ldots 3.8$$

$$V_x + (0.5V_S - \Delta V_S) = I_y(R_1 + R_2) \quad \ldots 3.9$$

$$\left[(0.5V_S + \Delta V_S) - (I_x + I_y) R_1\right] - V_x = I_y R_2 \quad \ldots 3.10$$

Extracting I_x from equations 3.9, 3.10, hence from equation 3.8

$$(\Delta V_S - V_x) = \frac{R_1}{2R_3}\left[V_{il} - (V_{BE1} - V_{BE4}) - V_x\right]$$

from which the balance condition follows,

If, $R_1 = 2R_3$ ... 3.11

then, $\Delta V_S = V_{il} - (V_{BE1} - V_{BE4})$ ... 3.12

The same balance condition as for the Class A case results except for an offset voltage $(V_{BE1} - V_{BE4}) \simeq 1.4V$ in the location of the PSU about ground. At crossover, this produces a PSU transition of $\simeq 2.8V$, otherwise circuit operation exhibits similar characteristics to those described in sub-section 3.1.1. In selecting R_3, some compensation for transistor r_e should be made, even though r_e is now grossly non-linear.

The analyses reveal an important property of the location circuit: i.e. optimal operation does not depend upon the class of bias chosen for the operation of transistor pairs T_1, T_2 and T_3, T_4. Hence during dynamic drive, should the location amplifier be driven out of Class A operation, correct location still prevails with full error-feedback correction, with the provision that changes in r_e, in comparison to R_3 (R_3*), cause only a moderate misalignment of the balance condition.

Consequently, a simple and efficient circuit topology results that represents an optimal control for a floating PSU. In practice, the circuit illustrated in Fig.3-2 can operate without bias currents as discussed in sub-section 3.1.2, though some distortion in the relationship of equation 3.6 results at crossover. However, we observe the error-feedback location circuitry is outside the unity-gain, power buffer which accurately defines the loudspeaker drive voltage (V_o) to match V_{in}, thus secondary distortion is well controlled.

Further operational enhancement (particularly relevant under zero bias of T_1, T_2 and T_3, T_4) is achieved by using feedforward resistors R_{ff} (see Fig.3-2), which achieve a bypass of PT_3, PT_4 during their non-conducting state at crossover, thus allowing Class B operation, if desired. However, it is recommended for "high-end" systems that non-linear transient displacement in ΔV_s at crossover is minimised by introducing bias within the T_1, T_2, T_3, T_4 and PT_3, PT_4 transistors, as presented in sub-section 3.1.1. Also, there is advantage is seeking high current gain in PT_3, PT_4, as this both reduces T_3, T_4 drive requirements and improves balance condition sensitivity under non-optimal alignment of equation 3.5 (see section 3.2 for further details).

In practice, though not strictly recommended, a two-channel amplifier could employ a common mains transformer core, with two independent floating windings. In practice, magnetic and capacitive coupling is minimised as the PSUs are outside the unity-gain power buffers. Examination of this reconfiguration, compared with a standard dual supply (of $\pm V_s$) shared between the two channels, reveal a minimal cost penalty either on transformer or reservoir capacitance. There is also the attraction that the two supplies are independently assigned to each channel. This arrangement could well prove attractive for bi- and tri-amped active systems.

Therefore to summarise, the balance condition $R_1 = 2R_3$* ensures the output voltage range of the amplifier is maximised and the power and voltage distribution is evenly distributed throughout the power transistors and compares favourably with a conventional bridge construction. The minamalist circuitry within the location amplifier offers wide bandwidth and allows accurate tracking of ΔV_s with V_{in}, yet its function lies outside the single, unity-gain power buffer. In this respect, it is more elegant than the conventional bridge amplifier.

Finally as a postscript: if current limiting or fuse protection is required (both for amplifier and loudspeaker protection), it may be placed in series with the single floating supply (thus requiring only one fuse/protection circuit), where its operation is symmetrical and prevents chain failure of supplies/devices/loudspeakers, that is common in dual, centre tapped PSUs.

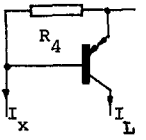
3.2 Sensitivity of supply location circuitry to loudspeaker load impedance and output transistor current gain.

This Section explores the sensitivity of the supply location circuitry under sub-optimal balance, as a function of loudspeaker load impedance and power transistor current gain.

Optimum power supply location depends upon the accuracy of the balance condition, expressed by equation 3.4. However, where sub-optimal balance occurs, errors are minimised by enhancing the current gain of the power transistors PT_3, PT_4 (shown in Fig.3-2).

The current gain of PT_3, PT_4 follows by considering these devices to be in parallel, together with the inclusion of R_4. The current gain β_o in the active state is then defined as

$$\beta_o = \frac{I_L}{I_x} = \frac{V_o}{|Z_L| I_x} \quad ,$$



PT_3, PT_4 in parallel

Using the approximation cited in Section 3.1 regarding T_1, T_2, T_3, T_4 ideallity and observing Fig.3-2, V_x is written

$$V_x = \frac{V_o}{2}\left[1 - \frac{2R_3^*}{\beta_o |Z_L|}\right] \qquad \ldots \quad 3.13$$

Define, $\alpha = \frac{R_1}{2R_3^*}$ ... 3.14

Then from equation 3.4, ΔV_S follows as

$$\Delta V_S = \frac{V_o}{2}\left[\alpha + (1-\alpha)\left[1 - \frac{2R_3^*}{\beta_o |Z_L|}\right]\right] \qquad \ldots \quad 3.15$$

The relative sensitivity relating ΔV_S to β_o is defined,

$$S^{\Delta V_S}_{\beta_o} = \frac{\beta_o}{\Delta V_S} \frac{\partial \Delta V_S}{\partial \beta_o} \qquad \ldots \quad 3.16$$

whereby,

$$S^{\Delta V_S}_{\beta_o} = \frac{(1-\alpha)\dfrac{2R_3^*}{\beta_o |Z_L|}}{\left[\alpha + (1-\alpha)\left[1 - \dfrac{2R_3^*}{\beta_o |Z_L|}\right]\right]} \qquad \ldots \quad 3.17$$

which is approximated as $\alpha \to 1$ to

$$\left.\frac{\Delta V_S}{S}\right|_{\beta_o} = \frac{(1 - \alpha)\ 2R_3}{\beta_o\ |Z_L|} \qquad \ldots \quad 3.18$$

Hence, in designing the floating power supply location circuitry, dependance on β_o for non-optimal balance is minimised by ensuring,

$$\beta_o >> \frac{2R_3^*}{|Z_L|} \qquad \ldots \quad 3.19$$

whereby, from equation 3.15,

$$\Delta V_S = 0.5 V_o \qquad \ldots \quad 3.20$$

i.e. ΔV_S tracks V_o and the balance condition is desensitised allowing equation 3.5 to become non-critical.

4 N-cell cascade of quasi-pontoon amplifiers

In this Section, N quasi-pontoon amplifier cells are interconnected in a cascade, to form a single amplifier exhibiting an N-fold increase in output voltage capability. However, since only a single, unity-gain power buffer defines the output voltage across the loudspeaker, there is only a distortion penalty commensurate with the increase in load current.

The quasi-pontoon amplifier, introduced in Fig.3-2, is here represented by the system block, shown in Fig.4-1, where the block has two, high-impedance inputs labelled I/P and V_{ip}, an output port V_o and a centre port ct, (same notation as Fig.3-2). Each cell has an individual, isolated, floating supply of nominal voltage, V_S.

The construction of the N-cell cascade is shown in Fig.4-2. Cell 1 forms the main signal amplifier, while cells 2 to N are driven by the R-uniform potential divider chain, such that V_o of cell (r + 1) drives ct of cell (r). Since all cell I/Ps are driven in synchronism, there are no accumulated time delay effects. Note, as in the earlier example of Fig.3-2, the supply location inputs $V_{ip}|^N_1$ are driven by the mean of V_o and ct, thus ensuring optimal local supply location for each cell; this is implemented by the R_o uniform potential divider chain. A design note in implementing this cascade, is that each cell must be capable of passing the peak load current, though no extra demand on device voltage rating is inherent.

5 Multi-cellular pontoon amplifier

The single and multi-cellular topologies described in Sections 3 and 4 are configured as single-ended amplifiers, with the loudspeaker 'cold-end' referenced to a star-ground to minimise ground contamination.

However, two differentially driven amplifiers may be combined to form a full-bridge amplifier, similar to Fig.2-1, that offers all the advantages cited in Section 2. A multi-cellular pontoon amplifier construction is illustrated in Fig.5-1 and is a direct extension of the cascade shown in Fig.4-2. Note how two, differentially driven, potential divider chains form optimum PSU location, both within individual cells and within the amplifier as a whole.

6 Mono-PSU pontoon amplifier

The preample in Section 2.4 revised the advantages of using a single PSU in a differential bridge amplifier. In this Section, we extend that discussion to the pontoon amplifier by combining a fully floating PSU with error feedback location, within a symmetrical differential bridge amplifier using unity-gain, power stages. The pontoon amplifier in elemental form is presented in Fig.6-1 and is an extension of the topology shown in Fig.3-2.

The PSU location circuitry senses the mean voltage between the two outputs of the unity-gain power buffers. The circuit operation is therefore tolerant of common mode signals.

A critical examination of the pontoon amplifier reveals the input bias currents of the unity-gain amplifiers are the only currents returned to ground. Consequently, the location transistors T_5, T_6 require only a low-power rating in order to compensate for this minimal offset current and attain correct dynamic placement of the PSU.

Under dynamic excitation, T_5, T_6 must also supply current through C_p (see Fig.6-1), the PSU mutual capacitance to ground. However, this current path is predominantly outside the main signal loop and the current is virtually zero as the PSU remains nearly static with respect to ground under symmetric differential drive. In practice, the screen of the transformer in the PSU is returned to the star-ground point of the amplifier, even though the currents in this line are minimised by the system configuration.

This present bridge proposal extends the philosophy discussed in this paper to a symmetrical, differential drive configuration, including the line amplifier, and represents an optimal system construct. The proposal minimises the signal residues injected from PSUs and should lead to an improved system transparency.

7 Transconductance power cell with fully floating supply

This Section presents a reconfiguration of the Fig.3-2 quasi-bridge topology, that enables the circuit to operate as a transconductance cell, for applications requiring both a high source impedance and high current delivery.

Central to the topology is a unity-gain cell with a high input impedance, where an elemental system is shown in Fig.7-1.

Since the supply is floating, the load current, I_L, must also flow through R_g. Hence,

$$I_L = \frac{V_{in}}{R_g} \quad \ldots \quad 7.1$$

whereby I_L is independent of V_s, providing the unity-gain buffer is optimised for supply rejection and offers high impedance current sourcing to its supply lines: a factor aided by the naturally high collector slope resistance of transistors.

In Fig.7-2, we outline an adaptation of the schematic of Fig.3-2, where the error feedback location circuitry is driven by

$$V_{il} = 0.5 \ [V_{in} - V_o] \quad \ldots \quad 7.2$$

to seek optimum placement of V_s, as V_o now depends intimately on both I_L and the load impedance Z_L,

i.e. $$V_o = I_L Z_L \quad \ldots \quad 7.3$$

In particular, the two sensing resistors R_o do not reduce the output impedance of the transconductance cell, as no current is returned to ground.

It is recommended that the location circuitry T_1, T_2, T_3, T_4 operates in Class A, with PT_3, PT_4 biased in Class AB. Also, feedforward resistors R_{ff} can be included to give partial bypass to the floating supply and provide continuity of current under crossover transitions. However, distortion is primarily a function of the unity-gain buffer and parallels the voltage drive configurations already discussed in detail.

Finally, in Fig.7-3, we present an extension to the single transconductance cell that parallels the system of Fig.4-2 and allows a greater output voltage across the load, by using a cascade of the elemental floating supply cell. The same notation as presented in Fig.4-1 is appropriate.

8 Conclusion

This paper has explored a redefinition of the relationship of line amplifier to power amplifier, which is particularly relevant with the advent of CD players, with high output signal levels (circa 2V). The proposal proffers the reconfiguration of the power amplifier as a separate unity-gain amplifier in cascade with an independent, high output voltage line amplifier, that exhibits only moderate current delivery, commensurate with driving the line stage+power amplifier interconnect. This approach seeks to eliminate redundant channel circuitry and enhance system performance, without incurring unnecessary cost penalties.

The orthogonal operation of line and output stage minimises signal corruption through power supply interaction, both by electromagnetic field coupling and directly, via common supply rail and ground line contamination. Also, a unity-gain power amplifier operates at a higher signal level, and is therefore less susceptible to supply rail induction.

This differs significantly from standard practice, where high sensitivity circuitry referenced to ground often share common supply rails, with the power output stage in close geographical proximity.

An adaptation of the classic bridge amplifier, the quasi-pontoon amplifier, was introduced that is well matched to the unity-gain proposal. The construct allows single-ended operation but configured with an independent, single, floating supply that does not require a centre tap. Use of a novel error feedback procedure enabled optimal location of the floating supply with respect to both output voltage and ground that also achieved a uniform power distribution. A secondary advantage of the scheme allows a greater PSU reservoir capacitance for a given spatial volume, as PSU voltage is reduced and no centre-tap is required. Extensions to the basic quasi-pontoon amplifier allowed in principle a cascade achieving any desirable output voltage, yet still maintaining optimum PSU placements within the chain, together with no accumulation of distortion. A fully floating, multi-cellular, differential output configuration was also presented.

The present paper has not discussed detailed circuitry for realising unity-gain, power buffer structures, as this represents on-going activities. However, several topologies and discussions that are suitable for adaptation have been presented in earlier work (1, 2). Also a final circuit depends upon target specification and designer component preferences. However, we reiterate the statement that, in the limit, amplifier performance is determined largely by the quality, construction integrity and topology of the PSU. In forwarding a reappraisal of the role of the line amplifier and power buffer, and configuring a more optimal bridge structure, the pontoon amplifier, with fully floating supplies, we seek to achieve an overall system configuration that offers subjective advantage, yet within an economic framework. Ultimately, symmetric, differential constructions with a fully floating supply and unity-gain power buffers, driven by differential line amplifier are recommended for an optimum system configuration.

Finally, we emphasise an important characteristic of the error correction supply location circuitry, where Section 3 reveals a desensitisation of the balance condition $R_1 = 2R_3^*$ by increasing the current gain of PT_3, PT_4 (Fig.3-2). In fact, the correction circuitry effectively compensates for the current gain of the power location transistors (under optimum balance) and this is true even if the current gains of PT_3, PT_4 are non-linear, a welcome circuit response that follows directly from a floating supply construct. Consequently, the proposed location circuit represents an ideal solution to the floating power supply configuration, where presented extensions of the principle enable a useful range of system applications to be explored together with all the advantages of bridge working.

References

1 Hawksford, M.J. 'Distortion correction in audio power amplifiers', JAES, Vol. 29, No. 1/2, pp.27-30, Jan/Feb. 1981

2 Hawksford, M.J. 'Power amplifier output stage design incorporating error feedback correction with current dumping enhancement', 74th AES Convention, New York, preprint 1993 (B-4), 1983

3 Cherry, E.M. 'A new distortion mechanism in Class B amplifiers', JAES, Vol. 29, No. 5, pp.327-328, May 1981

4 Tanaka, S. 'New biasing circuit for Class B operation', JAES, Vol. 29, No. 3, pp.148-152, March 1981

5 Greiter, O. 'Transistor amplifier output stages', Wireless World, Vol. 69, No. 1 to 4, Jan, Feb., March, April 1963.

6 Hawksford, M.J. Private communication to Editor, Hi-Fi Answers, Jan. 1983

7 Howard, K. 'A bridged version', Hi-Fi Answers, Haymarket Publications, May 1983

8 Otala, M. 'Transient distortion in transistorised audio power amplifiers', IEEE Trans. Audio Electroacourt., Vol. AU-18, pp.234-239, Sept. 1970

9 Martickainen, I. and Varta, A. 'About loudspeaker system impedance with transient drive', 71st Convention of AES, Montreux, preprint No. 1886 (D-3), March 1982

10 Brady, R.M. 'The floating bridge, new design principle for audio amplifiers', Wireless World, Vol. 86, No. 1537, pp.40-43

11 Brady, R.M. 'The floating bridge - 2, unconventional amplifier circuits for 15 and 200 watts', Wireless World, Vol. 86, No. 1537, pp.83-86, Oct. 1980

12 Takahashi, S. and Tanaka 'Fully balanced bridge', 72nd Convention of the Audio Engineering Society, Anaheim, CA, Oct. 1982, later: JAES, Vol. 32, No. 6, pp.415-421, June 1984.

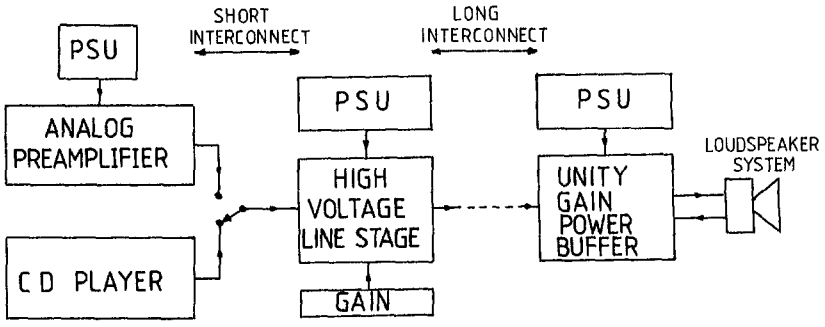
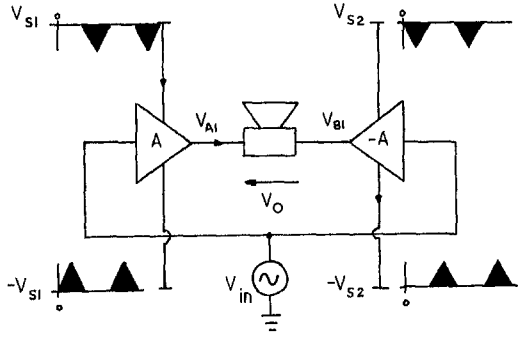
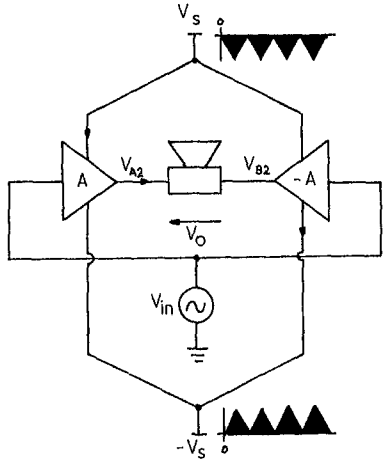


Fig. 1-1 Basic system of line stage and power buffer



(a) Bridge amplifier with seperate power supplies



(b) Bridge amplifier with common supply

Fig. 2-1 Bridge Amplifier

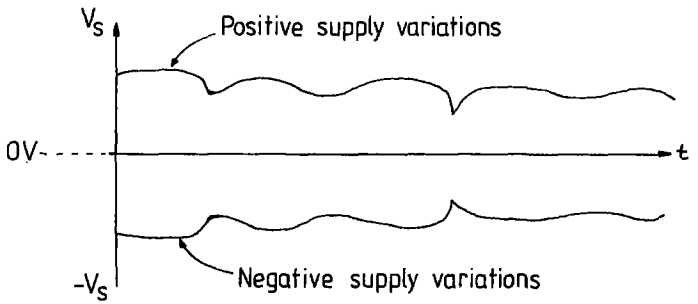


Fig. 2-2 Supply variation for bridge amplifier operating with common supplies (ref. Fig.2-1(b)).

(Note mirror image of supply variations)

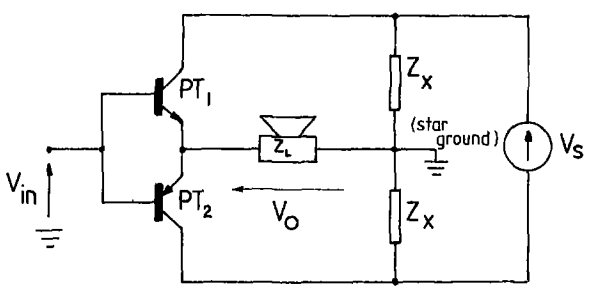


Fig. 3-1 Elemental floating power supply amplifier

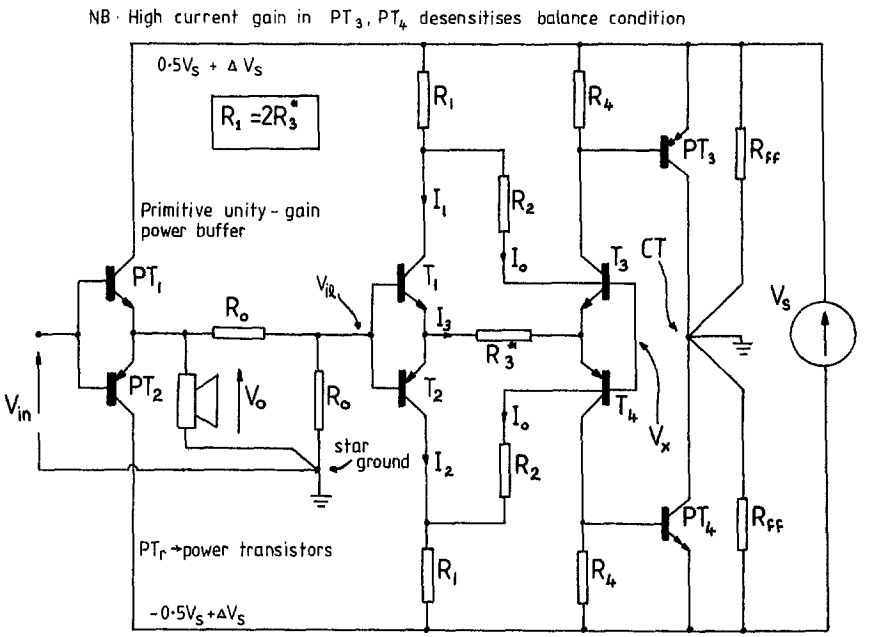


Fig. 3-2 Primitive quasi-pontoon amplifier with error feedback PSU location circuitry

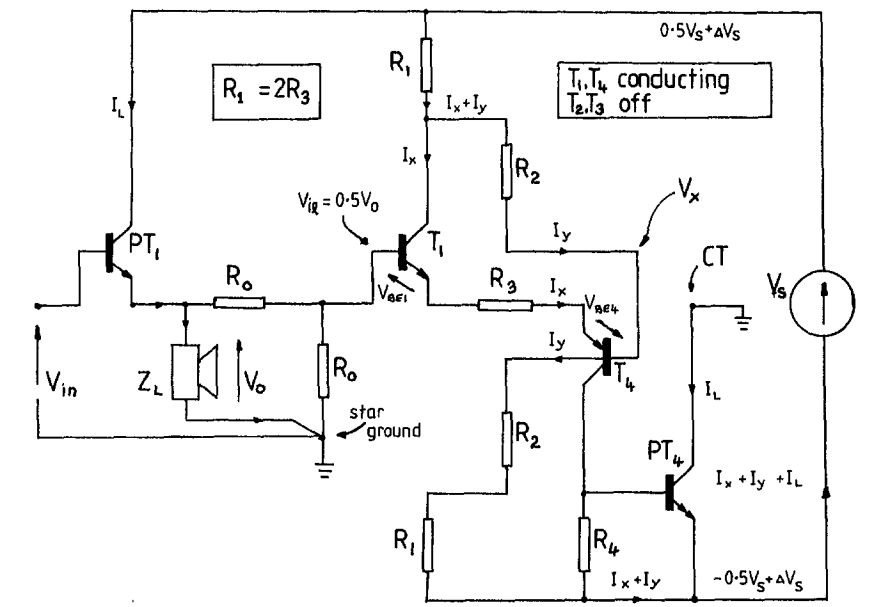


Fig. 3-3 Class B operation of location circuitry: T_1, T_4 conducting

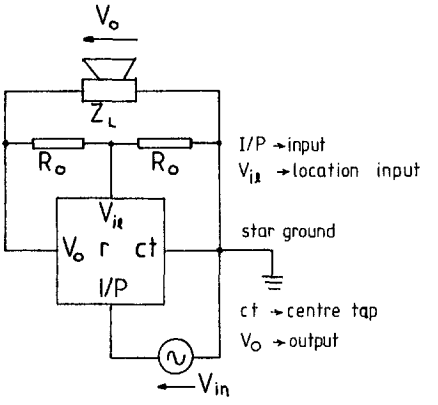


Fig. 4-1 System block representation of quasi-bridge cell

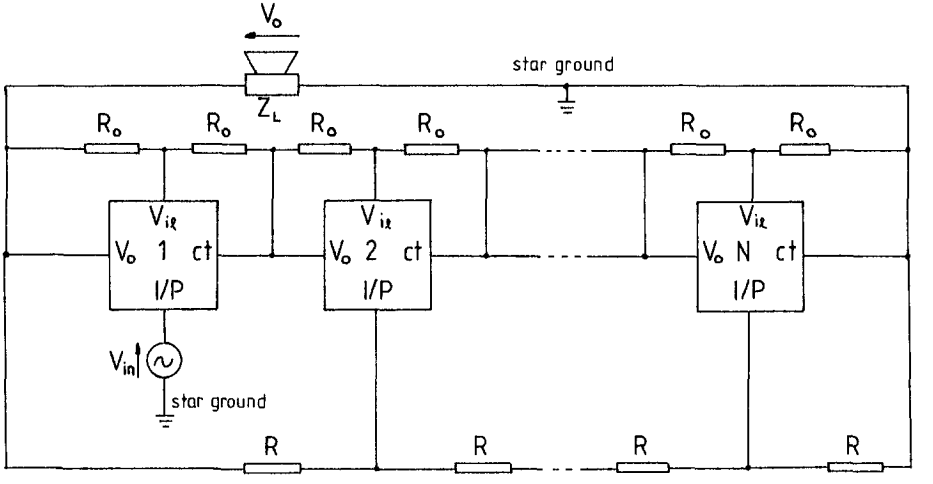


Fig. 4-2 Cascade of N unity-gain, quasi-pontoon amplifier cells with N floating PSU s

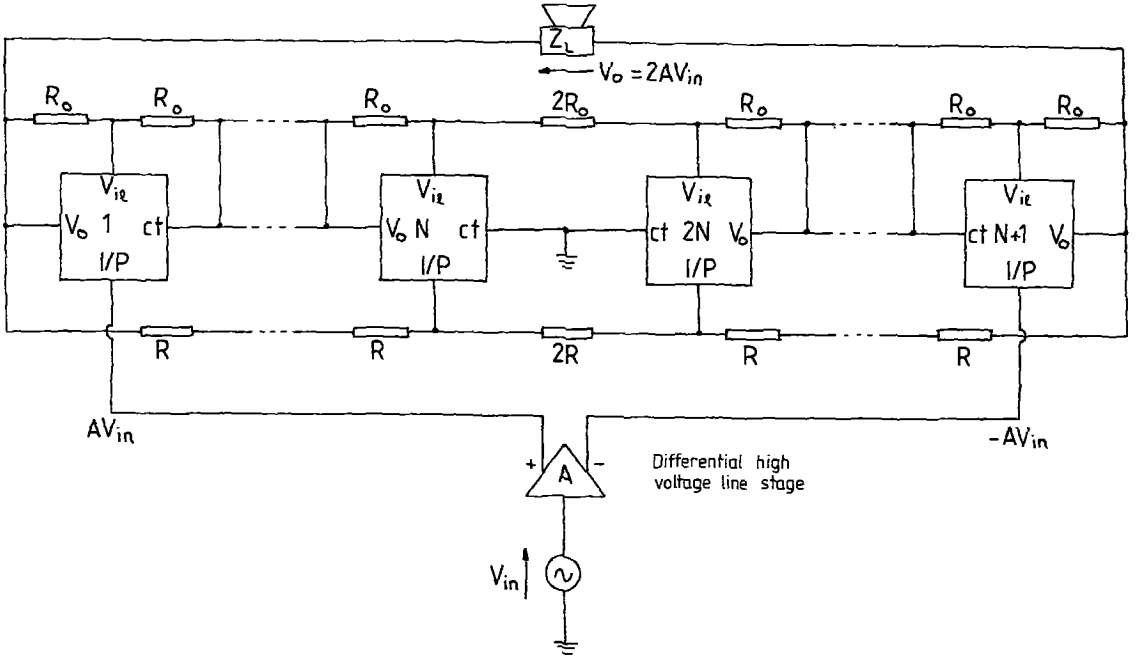


Fig. 5-1 Multi-cellular pontoon amplifier with 2N floating PSU s

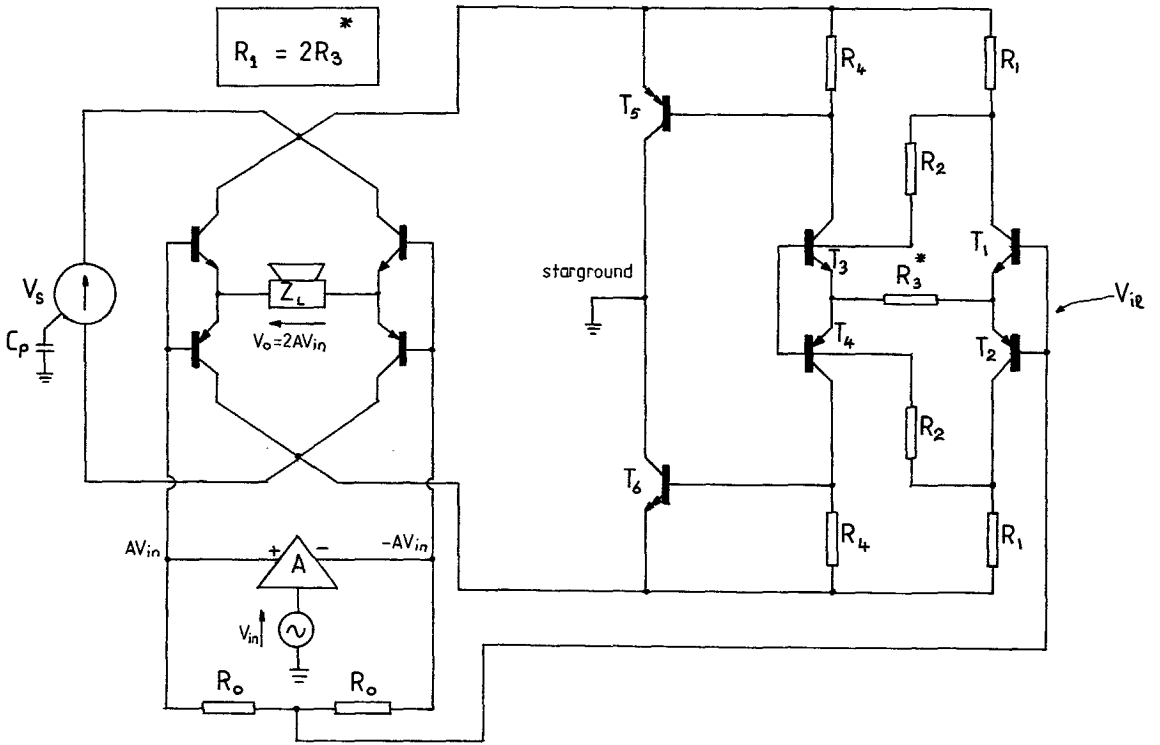


Fig. 6-1 Pontoon amplifier: fully floating, differential bridge amplifier with error correction supply location

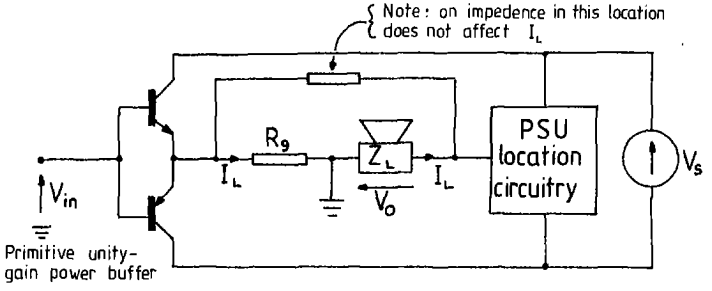


Fig. 7-1 Elemental transconductance gain cell with floating supply

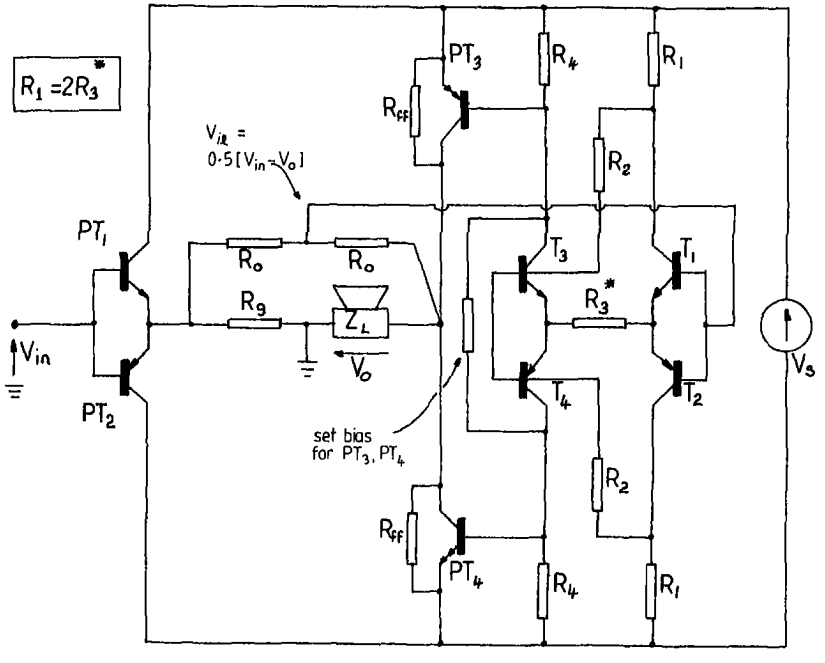


Fig. 7-2 Transconductance, unity-gain power cell using error-correction floating supply location referenced to V_o

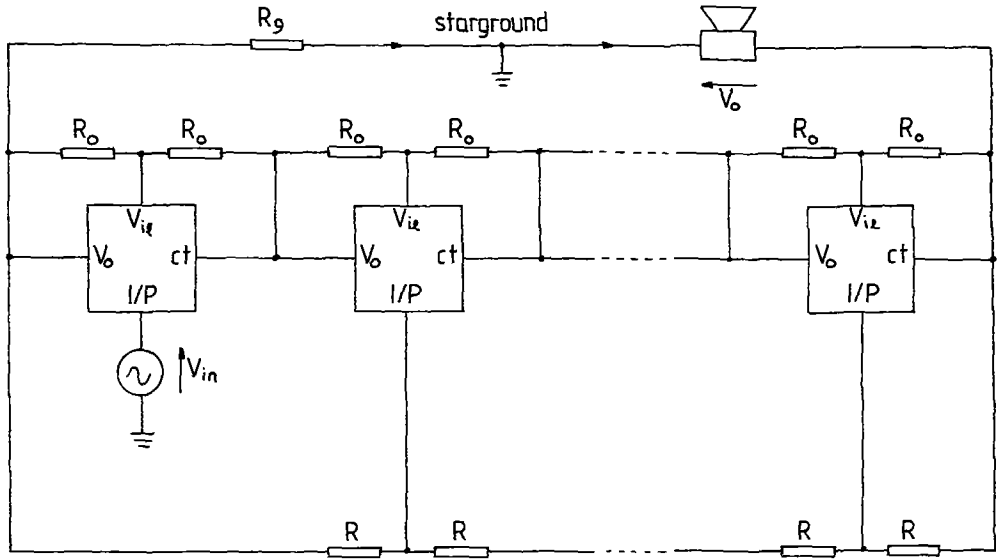


Fig. 7-3 N-cell cascade, transconductance amplifier with enhanced output voltage capability

Towards a Generalisation of Error Correction Amplifiers

M.O.J.Hawksford
Dept. Electronic Systems Engineering
University of Essex

Towards a generalisation of error correction amplifiers

0. Introduction

The "current dumping" amplifier was introduced at the 50th Convention of the AES (1) in 1975 by Peter Walker and Michael Albinson and represented a milestone in the evolution of analogue amplifers. Until that time, most transistor power amplifiers had been variations of directly biased output stages operating in either class A or class AB, together with overall negative feedback, to achieve acceptable linearity, although there were already several innovative developments, such as the work of Blomley (2)

Naturally, tube technology was well developed at this time, where device characteristics dictate rather different system topologies since tubes are essentially high voltage, low current devices and there is no "PNP" equivalent. Also, error reduction schemes were established for use with tube electronics well before the Walker amplifier. For example, a patent by Llewellyn (3) describes a method of error feedback distortion reduction, though simultaneous feedforward correction was not cited. Feedforward error correction however was originally described by the Black patent (4) and although it is used in high frequency circuits (5), feedforward had not, prior to 1975, found application as a correction procedure in audio power amplifiers. However, local feedforward within an overall feedback loop had been successfully applied in the tube circuitry of AR (6), where the cathodes of the output tubes effectively feed across the primary to secondary of the output transformer and directly couple to the load impedence, although no attempt was made to seek a topology capable of a balance condition equivalent to the Walker circuit (1).

Since 1975, there has been extended debate as to the virtues and fundamental principles of current dumping. Some researchers have proposed a balanced bridge analogue (7) to explain the distortion null, while others such as Nigel Allinson (8) have correctly recognised the combination of both feedback and feedforward within a common structure. While yet a further school (9, 10) has attempted to deny the existance of the mechanism of feedforward distortion correction preferrring what appears to be an impractical overall feedback loop (canonic form), that neglects the elegance of the original concept, which exhibits both a natural empathy with real device characteristics and a true error null.

In this paper, we re-affirm the existance of the "current dumping" principle and extend the comparative discussions by demonstrating equivalance with a more general composite error feedback, error feedforward model. It is not intended that this model invalidates other critical observations, rather that it complements them by taking an alternative stance, aimed primarily at linking earlier work on error feedback and error feedforward (11, 12, 13).

We then conclude the comparative discussion by proposing a structure common to control engineering, analogue computing and transient analysis from which many of the present day distortion correction systems can be derived. This re-inforces the foundation of error correction

and allows a vehicle for identifying new topologies that exhibit broad-band distortion correction. To demonstrate error correction, we use the transfer error function (14) as an indicator of both system performance and a means of identifying and classifying sets of system balance equation to achieve distortion nulling. This error function as well as expressing the system error, also allows the sensitivity of the balance equation (5) to be directly evaluated.

1. Primitive model of error feedback/feedforward

Error correction, rather than error reduction, implies there is a balance equation (or equations), which under optimal alignment exhibits a broad band distortion null rather than just a reduction in distorion. The apparent implication is that the canonic equivalent feedback (only) loop must allow infinte loop gain over a broad band of interest. However, for simple feedback, this theoretic requirement is impossible, an observation recently emphasised by Lipshitz and Vanderkooy (15), a paper forming a useful complementary discussion.

To achieve a theoretic broad-band distortion null at least one feedforward path that extends beyond the feedback loop is mandatory, to compensate for the limitations of any practical feedback loop that can be devised. We concur with Allison (8), Vanderkooy and Lipshitz (11) that this compenstion is a fundamental requirement, any system attempting to eliminate the feedforward path yet attain zero distortion is impractical. Of course, in making this observation we do not deny the low distortion achievments possible with feedback, we are considering the limiting case: what is achievable with feedback alone can, in principle, be enhanced with the inclusion of feedforward. However, in practical systems, device characteristics may well negate the performance advantages offered by a particular system philosphy, where there has already been much debate (16, 17, 18).

Let us commence by re-examining the error correction strategy proposed in an earlier paper (12), which offered both error feedback and error feedforward and which in combination enable a true distortion cancellation of the error arising from the non-linear output cell, N.

In this primitive model illustrated in Fig 1-1, the overall voltage gain for a given set of parameters {N, a, b} is A, where the corresponding target gain $A_t = 1$. Also, at any instant, the non-linear output cell is assumed to have an incremental gain N, where in this example for zero distortion $N \rightarrow 1$, which is compatible with the adoption of $A_t = 1$.

Hence observing the respective coefficients {a, b} in the error feedback and feedforward paths in the scheme of Fig 1-1, the overall transfer function A for non-optimal parameter alignment is,

$$A = \frac{N - b(N-1)}{1 + a(N-1)} \qquad 1\text{-}1$$

Defining the corresponding transfer error function (14, 19) E_1, as,

$$E_1 = \frac{A}{A_t} - 1 \qquad 1\text{-}2$$

and substituting for A, noting that in this case $A_t = 1$, then

$$E_1 = (N\text{-}1)\left[\frac{1 - (a+b)}{1 + a(N\text{-}1)}\right] \qquad 1\text{-}3$$

For this example, equation 1-3 shows that the error function E_1, tends to zero when either N=1 and/or more fundamentally, when the numerator contains the relationship,

$$a+b = 1 \qquad 1\text{-}4$$

which is the balance condition that enables a theoretic distortion null, providing equation 1-4 is maintained over an adequate bandwidth.

It is here that the need for a feedforward path that extends beyond the feedback loop is evident: the parameter 'a' is determined by the stability constraints of feedback, while the factor 'b' can be freely selected to yield a broadband distortion null as it is independent of factors affecting loop stability. We therefore conclude that both feedback and feedforward are complementary to achieving exact error correction as a = 1 cannot be attained over a broad band.

2. Conceptual equivalance of error feedback/feedforward to current dumping

Consider the re-configured schematic in Fig 2-1, noting its equivalence with Fig 1-1. This system is conceptually similar to the Walker amplifier (1) and the later derivative amplifier offered by Sansui (Super - ff) (20). At this stage this equivalance is less obvious as the bridge components of the current dumping amplifier are excluded, though the respective feedback and feedforward paths are identified; we will shortly extend the schematic to include the bridge that is more commonly associated with the Walker interpretation of current dumping.

The loop gain A_f of the negative feedback path in Figs 1-1, 2-1 is,

$$A_f = -N\frac{a}{(1 - a)} \qquad 2\text{-}1$$

where in this form, the term a/(1 -a) can represent the gain in the forward path.

If we observe the feedback loop in the original configuration of Fig. 1-1, the parameter 'a' can be selected as a first-order, low pass filter to achieve a stable loop, and is a typical for an error feedback loop (14).

$$\text{i.e. } a = \frac{1}{1 + j\omega\tau_a} \qquad 2\text{-}2$$

Observe how as $\omega \rightarrow 0$, $a \rightarrow 1$ which achieves optimum distortion correction at dc, while the time constant τ_a establishes the dominant loop break frequency, although unfortunately it also prevents optimum distortion correction by forcing $a < 1$ for $\omega > 0$.

It is common practice in feedback amplifiers, including the Walker amplifier, for the forward gain to take a form approximately to $A_o/(1 + j\omega A_o\tau_1)$, where A_o is the dc gain and $1/(2\pi\tau_1)$ is the gain-bandwidth product (ie the frequency at which the loop gain, for a first-order system, is unity), where the form of the function is illustrated in Fig. 2-2.

Hence, noting in Fig. 2-1 that the loop gain (excluding the output stage of incremental gain N as $N \approx 1$) has the equivalent form $a/(1-a)$, we can set

$$\frac{a}{1-a} = \frac{A_o}{1 + j\omega A_o\tau_1} \qquad 2\text{-}3$$

whereby,

$$a = \frac{A_o}{(1 + A_o)} \cdot \frac{1}{\left[1 + j\omega \dfrac{A_o\tau_1}{(1 + A_o)}\right]} \qquad 2\text{-}4$$

Hence, for the special case where $A_o \rightarrow \infty$, then $a \rightarrow 1/(1+j\omega\tau_1)$ and $A_p \rightarrow -N/(j\omega\tau_1)$, a result corresponding to equation 2-2 where $\tau_a = \tau_1$.

We observe in this example that in the canonic form, the loop filter is an integrator in cascade with the non-linear output cell N, a condition representative of negative feedback power amplifiers, where providing N is well behaved at high frequency, good stability margins can be achieved.

The optimum feedforward parameter 'b' to achieve broad band error correction is determined from equations 1-4 and 2-4 as

$$b = \frac{1 + j\omega A_o\tau_1}{(1 + A_o)\left(1 + j\omega \dfrac{A_o\tau_1}{1 + A_o}\right)} \qquad 2\text{-}5$$

which for $A_0 \rightarrow \infty$, simplifies to

$$b\Big|_{A_o \rightarrow \infty} = \frac{j\omega\tau_1}{1 + j\omega\tau_1} \qquad 2\text{-}6$$

In Fig 2-3, the Fig 2-1 structure is further modified to show how feedforward error addition can be performed by an output network L_O, R_O, while the forward gain a/(1-a) is represented by equivalent integrator transfer function (for $A_O \rightarrow \infty$) as $1/(j\omega\tau_1)$. Hence, comparing the coefficients b, (1-b) in Fig 2-1 with the L_O, R_O network of Fig 2-3, it follows that,

$$b = \frac{j\omega L_o/R_o}{1 + j\omega L_o/R_o} = \frac{j\omega\tau_1}{1 + j\omega\tau_1}$$

$$(1\text{-}b) = \frac{1}{1 + j\omega L_o/R_o} = \frac{1}{1 + j\omega\tau_1}$$

i.e. assuming the forward gain exhibits a first-order response, the error summation exactly matches the requirement for optimum distortion correction, providing

$$\tau_1 = \frac{L_o}{R_o} \qquad 2\text{–}7$$

Before completing the discussion of conceptual equivalence of the Fig 1-1 structure with the Walker amplifier, let us investigate how the system can be adapted to include overall gain. In Fig. 2-4, the scheme of Fig. 2-1 is reconfigured to include a feedback parameter k. However, by redefining the forward gain to be a/(1-a)k, the loop gain remains unaltered as do the equations derived for optimum balance. The final stage in our comparison with the Walker amplifier can now be made.

We have observed that for a large value of A_O, there is a near-optimum value for 'a' of unity, where this parameter [$A_O/(1+A_O) \rightarrow 1$ as $A_O \rightarrow \infty$] is very insensitive to changes in A_O, ie. at low frequency almost perfect error correction is achieved by using a large loop gain. We note also that although the forward path has a low 3dB break frequency ($\omega = 1/A_O\tau_1$), when translated to parameter 'a' by equation 2-4, this frequency is multiplied by a factor ($1 + A_O$), implying a break frequency tending to ($\omega = 1/\tau_1$). Therefore, an important circuit attribute of the Walker topology (1, 21) is that although the dc gain A_O maybe poorly defined, the terms a and τ_1, are more accurately specified, consequently, when selecting 'b' in the feedforward summing network, the balance is both predictable and stable.

In completing our comparison, we concur with Walker, by noting the forward gain requires a close adherence to an integrator transfer function. In principle, the integrator can be configured using a virtual-earth technique, where effectively the gain between X and Y, -a/(1 - a) in Fig 2-4, is formed using an amplifier incorporating a local , frequency selective feedback network, where the basic circuit is presented in Fig 2-5a. In Fig 2-5b, the more complete current dumping topology is shown, while in Fig 2-5c, the equivalent descrete circuit with associated capacitance of the Walker amplifier (1, 11) is illustrated, which can also yield an excellent approximation to the required integrator response. It follows from the circuits of Fig 2-5 a, b that the gain between nodes X and Y is,

$$\frac{-a}{(1-a)} = \frac{-1}{j\omega R_1 C_1}$$

$$\text{whereby, } a = \frac{1}{1 + j\omega R_1 C_1}$$

and corresponds to equation 2-4 where $\tau_1 = R_1 C_1$ and $A_o \rightarrow \infty$.

Hence using the balance condition stated in equation 1-4 and the result of equations 2-6, 2-7 the original Walker expression for balance follows,

$$\text{i.e. } R_1 C_1 = L_o / R_o \qquad 2\text{-}8$$

This Section has shown, that at a conceptual level, "current dumping" can be usefully compared with a composite error feedforward/error feedback structure. More important however, is the requirement for an error feedforward path that extends beyond the main feedback path to realise theoretic broad band distortion cancellation. The Walker amplifier would appear to be the first power amplifier to exploit this critical and fundamental requirement. Essentially it places the majority of the error reduction within the feedback path, which reduces sensitivity to imbalance and then simultaneously uses error feedforward for fine tuning the alignment of the balance condition at high frequency, and thus achieves true error correction - also of importance is that the error feedforward path does not require gain, allowing a passive summation network to be used (see Section 4 for further discussion). Consequently, the non-linear and time dispersive errors within the feedback loop that arise from output stage non-linearity are negated by feedforward.

3. Error feedforward/feedback correction with arbitrary gain

The correction topology decribed in Section 1 is restricted to a non-linear output cell that has a nominal gain of unity, where the balance condition in this primal example established an overall gain also of unity. However, in this Section, we generalise the error feedforward/feedback structure to a system with arbitary gain paramenters, where in general $N > 1$.

Since an accurate definition of overall gain is required, it is necessary to imbed a reference amplifier as shown in Fig 3-1, whose voltage gain R is, ideally equal to the overall target gain A_t. This strategy enables an output stage of incremental gain N, where N > 1, to be combined with error feedback/feedforward to fine tune the overall voltage gain, minimise distortion and thus achieve a better approximation to the target gain A_t.

With reference to Fig 3-1, the overall voltage gain A_h of the high gain system is,

$$A_h = \frac{N - b(N-R)}{1 + a(N-R)} \qquad 3\text{-}1$$

and the corresponding error function E_h is,

$$E_h = \frac{A_h}{A_t} - 1 \qquad 3\text{-}2$$

$$\text{i.e. } E_h = (N-A_t)\left[\frac{1 - (b + aA_t)\frac{(N - R)}{(N - A_t)}}{1 + a(N-R)}\right] \qquad 3\text{-}3$$

Examination of equation 3-3 shows that the error function is zero for the two sets of conditions,

$$\begin{cases} R = A_t & 3\text{-}4 \\ aA_t + b = 1 & 3\text{-}5 \end{cases}$$

$$\text{and/or, } N = A_t \qquad 3\text{-}6$$

Following a similar procedure to Section 1, we reconfigure Fig 3-1 to the structures shown in Fig 3-2 a,b. It is at once apparent that the first system is conceptually similar to the Walker (1) and Sansui (20) correction schemes, while the second is similar to Sandman's (22) error pick-off system. The disadvantage of the second system is of course the requirement for an extra amplifier with gain in the error feedforward path that is a function of the choice of coefficients in Fig 3-2(a), where for example: if $N \approx A_t$, a = 0.04, b = 0.20 then $aA_t = 0.80$ and $bR = bA_t = 4.00$, assuming $R = A_t$ and $aA_t + b = 1$.

4. Towards a more general error feedforward/feedback topology

A more general appraisal of distortion correction systems reveals that some systems can be decribed as an extension of the constant voltage class of filter structure, where under appropriate alignment, limited non-linearity within the filter topology is permissible. In fact, the Walker amplifier (1), Sandman's error pick-off (22) and feedback/feedforward topologies (8, 12) which all accommodate broad band distortion cancellation can be directly linked to the structure in Fig 4-1.

Although in Fig 4-1, the amplifier $A_1 \ldots A_n$ are shown as generalised transfer function, they

in turn can be decomposed into local feedback/feedforward structures in a similar way to that of Fig 4-1. For example, in the comparison with the Walker amplifier discussed in Section 1, local feedback around the forward amplifier yielded an approximation to an integrator response such that when the transfer function defining impedance of Fig 2-5a, b, c were included, a close equivalance to the bridge of the current dumping amplifier was observed.

We conclude this paper by demonstrating a procedure for establishing conditions of distortion cancellation for the more general feedback/feedforward topology of Fig 4-1.

The overall transfer function G_n for the n-stage structure of Fig 4-1 is given by,

$$G_n = \frac{b_o + b_1A_1 + b_2A_1A_2 + \ldots}{1 + a_1A_1 + a_2A_1A_2 + \ldots} \qquad 4\text{-}1$$

where assuming a target transfer function G_{tn} and defining an error function E_{Gn} similar to that in equation 1-2,

$$\text{then, } E_{Gn} = \frac{G_n}{G_{tn}} - 1$$

$$\text{i.e.} \quad E_{Gn} = \frac{(b_o - G_{tn}) + (b_1 - a_1G_{tn})A_1 + (b_2 - a_2G_{tn})A_1\ A_2 + \ldots}{G_{tn}(1 + a_1A_1 + a_2A_1A_2 + \ldots)} \qquad 4\text{-}2$$

Using equation 4-2, we can now procede to invent a range of systems for which $E_{Gn} = 0$:

Example of error correction systems

(i) Transfer function independent of $A_1, A_2, \ldots A_n$.

Observation of equation 4-2 reveals a set of (n +1) balance conditions which forces $E_{Gn} = 0$ under optimum alignment such that each of the (n + 1) equation does not include the amplifier gains $A_1, A_2, \ldots A_n$.

$$\left.\begin{array}{l} \text{i.e. } b_o = G_{tn} \\ \left[b_r = a_rG_{tn}\right]_{r=1}^{n} \end{array}\right] \qquad 4\text{-}3$$

The error function E_{Gn} of equation 4-2 also reveals that providing $\{A_1, A_2, \ldots A_n\} >> 1$, that significant reduction in sensitivity to balance misalignment can be attained. However, we note with caution that for $G_{tn} > 1$ then $b_o > 1$, which requires gain in the first

feedforward path. Of course, we could choose the coefficients

$$\{a_r, b_r\}_{r=1}^{n} = 0$$

though for an n-stage feedback amplifier there is advantage in using several local feedback paths.

(ii) Transfer function independant of A_2, A_3 ... A_n..

The problem in example (i) of $b_o > 1$ for $G_t > 1$ can be circumvented by regrouping the left hand terms in the numerator of equation 4-2,

i.e. write:

$$E_{Gn} = \frac{A_1\left[b_1 - \left\{G_{tn}\left(a_1 + \frac{1}{A_1}\right) - \frac{b_o}{A_1}\right\}\right] + (b_2 - a_2 G_{tn})A_1 A_2 + \ldots}{G_{tn}(1 + a_1 A_1 + a_2 A_1 A_2 + \ldots)}$$

whereby the balance equations become,

$$\left.\begin{aligned} b_1 &= G_{tn}\left(a_1 + \frac{1}{A_1}\right) - \frac{b_o}{A_1} \\ \text{and } \left[b_r &= a_r G_{tn}\right]_{r=2}^{n} \end{aligned}\right\} \quad 4\text{-}4$$

i.e. by including A_1 in the balance relationship, b_o can now be selected independently, allowing $b_o < 1$ or even $b_o = 0$ if desired.

(iii) Sandman's error-pick off distortion correction (22).

The system of error correction first attributed to Sandman (22) can be observed conceptually in the scheme of Fig 4-1 by simplifying the structure as follows:

Let, $b_2 = b_3 = \ldots = b_n = 0$

$a_2 = a_3 = \ldots = a_n = 0$

This simplification shown in Fig 4-2 reveals a single loop negative feedback amplifier where the input error voltage is amplified by b_o and then summed with a weighted contribution from the output of amplifier A_1. The overall transfer function of the Sandman scheme, G_S, and corresponding error function, E_S, follow from equation 4-1, 4-2 as,

$$G_s = \frac{b_o + b_1 A_1}{1 + a_1 A_1} \qquad 4\text{-}5$$

$$\text{and,} \quad E_s = \frac{(b_o - G_{t1}) + (b_1 - a_1 G_{t1}) A_1}{G_t (1 + a_1 A_1)} \qquad 4\text{-}6$$

where zero distortion in amplifier A_1 is achieved when

$$\left.\begin{aligned} b_o &= G_{t1} \\ b_1 &= a_1 G_{t1} \end{aligned}\right] \qquad 4\text{-}7$$

i.e. b_o determines the target gain G_{t1} and the condition for minimum distortion is set by $a_1 b_o = b_1$.

The principle disadvantage of this scheme is again the requirement of $b_o > 1$ for $G_{t1} > 1$, though since $G_{t1} < A_1$, the error amplifier b_o can be designed to exhibit a correspondingly lower distortion contribution than A_1, thus a useful performance enhancement is possible.

(iv) Current dumping error correction (1, 16)

Following the discussion of Section 1, we observe that the conceptual topology of the current dumping amplifier (1, 16) follows directly from Fig 4-1 when

$$b_o = 0, a_1 = 0$$

$$a_3 = a_4 = \ldots = a_n = 0$$

$$b_3 = b_4 = \ldots = b_n = 0$$

and the non-linear output cell N is represented within Fig 4-1 by the amplifier A_2. The transfer function G_w and corresponding error function E_w of the Walker current dumping topology then take the general form

$$G_w = \frac{b_1 A_1 + b_2 A_1 A_2}{1 + a_2 A_1 A_2} \qquad 4\text{-}8$$

$$E_w = \frac{(b_1 A_1 - G_{t2}) + (b_2 - a_2 G_{t2}) A_1 A_2}{G_{t2} (1 + a_2 A_1 A_2)} \qquad 4\text{-}9$$

In this particular example, the target transfer function G_{t2} can be derived from the overall transfer function G_w by putting $A_2 = 1$, a condition representative of zero distortion in the unity-gain output stage forming the current dumpers.

$$\text{i.e. } G_{t2} = G_w \Big|_{A_2 = 1} = \frac{(b_1 + b_2)A_1}{1 + a_2 A_1} \qquad 4.10$$

hence subsituting G_{t2} in the expression for E_w, from equation 4-9, and noting with passive error summation that $b_1 + b_2 = 1$,

$$E_w = \frac{(a_2 b_1 A_1 - b_2)(1 - A_2)}{1 + a_2 A_1 A_2} \qquad 4\text{-}11$$

i.e. zero distortion results when either $A_2 = 1$ (the optimum output stage gain in this example) or, more fundamentally,

$$\frac{b_2}{b_1} = a_2 A_1 \qquad 4\text{-}12$$

which is an alternative form of the balance condition for the current dumping amplifier, that is readily observed by reference to the analysis of Section 1.

In fact, the expression in equation 4-11 for E_w, succinctly describes most of the principle attributes of the basic Walker amplifier ie.

(a) zero distortion when $A_2 \rightarrow 1$

(b) further distortion reduction even with a finite loop gain, by using error feedforward.

(c) significant reduction of error by using negative feedback where the term $(1 + a_2 A_1 A_2)$ desensitives the balance condition to misalignment for $a_2 A_1 A_2 >> 1$.

(d) a true broad-band error null capability under optimal alignment of the balance condition.

(e) $\{a_2, b_1, b_2\} < 1$ if desired allowing passive error summation, though the parameters are in practice frequency dependant (1, 8, 16).

(v) Focus function error correction

Within the multi-loop feedback/feedforward structure of Fig 4-1, error correction can in principle be targeted individually onto any of the n amplifier stages A_1 to A_n to desensitise the overall transfer function to changes in incremental gain of the selected amplifier. This

ability to focus the error correction we define a focus function.

To illustrate the method of focusing error correction, consider an example of a n = 3 stage system where the second amplifier A_2 is selected for desensitisation. From the expression for the error function in equation 4-2, we may arrange the expression as follows:

$$E_{G3} = \frac{\left[(b_o + b_1A_1) - G_{t3}(1 + a_1A_1)\right] + A_1A_2\left[(b_2 + b_3A_3) - G_{t3}(a_2 + a_3A_3)\right]}{G_{t3}(1 + a_1A_1 + a_2A_1A_2 + a_3A_1A_2A_3)} \quad 4\text{-}13$$

Hence E_{G3} becomes zero and independant of A_2 when,

$$(b_o + b_1A_1) = G_{t3}(1 + a_1A_1)$$

and,

$$(b_2 + b_3A_3) = G_{t3}(a_2 + a_3A_3)$$

This method can readily be extended to the general case of n amplifiers, where two balance equations again result by appropriate factorisation of the numerator of the error function.

Since in a practical amplifier, the summation coefficients of the feedforward paths are more readily implemented using passive components, we put forward the opinion and thus concur with Walker that schemes should attempt to selct 'b' coefficents less than unity such that

$$\sum_{r=0}^{n} b_r = 1 \quad 4\text{-}14$$

where an appropriate method of passive implementation is shown in Fig 2-3. Hence if desired, this relationship can be considered as an extra constraint on the balance equation.

The error function described by equation 4-2, shows the advantage offered by multi-stage amplifier structures. Observation of the denominator reveals (to the right of the expression) high values of gain to be possible which greatly desensitize the error function to imbalance. Though a single high gain stage is permissible, multiple stages that distribute the gain and employ multiple feedback paths, allow greater degrees of freedom to attain stability under high loop gain (23) while simultaneous feedforward enables error correction to be focused on to selected critical stages to further enhance performance.

There is clearly an infinity of systems possible, where it is suggested that many of the proposals can be decomposed to a generalised feedforward/feedback structure as shown in Figure 4-1. This structure is therefore offered as a possible basis topology for error correction schemes, that do not depend upon dynamic gain modulation as corrective strategy (24).

5. Conclusion

This engineering report has attempted to establish a more general class of error correction

scheme, where we demonstate a structure, that can be configured at a conceptual level to realise many so called new systems.

We agree with Lipshitz and Vanderkooy that feedback alone cannot achieve a global error null and that at least one feedforward path is fundamental to this aim. Feedforward, embedded within a feedback structure enables compensation for finite loop gains, consequently we need not seek infinite loop gain to achieve a theoretic zero distortion.

In particular, the comparison of the "current dumping" amplifier to an error feedback/feedforward structure was a catalyst in the establishment of the more general model, especially where a re-interpretation of the closed-loop gain reveals essentially identical transfer functions.

However, although conceptual models are fundamental to a proper understanding of a system, ultimately it is the circuit realisation, layout and component choice that are paramount. The elegance of a system is represented both by the elegance of the circuit topology and its practical execution. Thus, although "current dumping" and other distortion correction strategies can be compared, the ultimate judgement should be at a circuit level, together with performance evaluation of the total system.

We can of course present multiple mapping from system concepts to circuit schematic, where new systems are invented. However, we must also recognise that ultimately, all systems are derivatives of feedback and feedforward, where the more recent developments have correctly combined these techniques into a common framework, rather than depending on the special cases of just feedback or just feedforward.

Of course, the proposal represented by the more general structure in Figure 4-1, will be familiar to those experienced in analogue computing techniques and transient analysis (25) where such systems are commonly used. Amplifier designers have side stepped this knowledge to a degree and probably concentrated more on the circuit aspects than the underlying philosophy.

The structure of Fig 4-1 allows in principle, many families of error correction schemes to be invented, that exhibit true theoretic error nulls. It also enables a desensitization of the balance conditions to be achieved, where this is readily observable using error function modelling. In particular, we can theorise on new topologies using multiple stages and multiple feedback paths to achieve closed loop stability, yet combined with feedforward to enable a greater reduction in the non-linear dispersive error inherent in feedback only systems. Also of academic interest is the means of focusing error correction on selected stages within the cascade, as discussed in Section 4 - (v).

6. References

1 Walker, P.J. and Albinson, M.P., 'Current dumping audio amplifier', presented at 50th Convention of the Audio Engineering Society, London, England, March 4-7, 1975

2 Blomley, P., 'New approach to class B amplifier design', *Wireless World,* pp. 57-61 and pp. 127-131, Feb/March 1971

3 Llewellyn, F.B., "Wave Translation Systems", US Patent 2,245,598, June 17, 1941

4 Black H.S., "Translating System", US Patent 1,686,792, October 9th 1928

5 Seidel H., Baurrier H.R. and Friedman A.N., "Error-controlled high power linear amplifiers at VHF", BSTJ, May-June 1968, pp. 651-722

6 AR Patent, US Patent No. 3,538,447

7 Vanderkooy J., and Lipshitz, S.P. 'Current dumping - does it really work?'. *Wireless World*, Vol. 84, No. 1510, pp. 38-40, June 1978

8 Allinson N., "Distortion reduction in frequency-dependent feedback-feedforward amplifiers", *Int. J. Electronics,*Vol. 59, No. 6, 1985, pp. 667-683

9 McLoughlin, M., 'Current dumping review - 1', *Wireless World,* Vol. 89, No. 1572, pp. 39-43, Sept. 1983

10 McLoughlin, M., ' Current dumping review - 2', *Wireless World,* Vol. 89, No. 1573, pp. 35-41, October 1983

11 Vanderkooy, J. and Lipshitz S.P. 'Feedforward error correction in power amplifiers'. *JAES*, Vol. 28, No. 1/2. pp. 2-16, Jan/Feb 1980

12 Hawksford, M.J., 'Distortion correction in audio power amplifiers', *JAES*, Vol. 29, No.1/2. pp.27-30, Jan/Feb 1981

13 Hawksford, M.J., 'Distortion correction circuits for audio amplifiers', *JAES*, Vol. 29, No. 7/8, pp. 503-510, July/August 1981

14 Hawksford, M.J., 'Power amplifier output stage design incorporating error feedback correction with current dumping enhancement', Presented at the 74th Convention of the AES, Preprint No. 1993 (B-4), October 8-12, 1983, NY

15 Lipshitz, S. and Vanderkooy, J., 'Is zero distortion possible with feedback?', Paper presented at 76th Convention of the Audio Engineering Society, Preprint No. 2170 (F-4), October 8-11, 1984, NY

16 Walker, P., 'Current Dumping', Letter to Editor, *Wireless World*, Vol. 89, No. 1575, p. 49, December 1983

17 Baxandall, P.J., 'Current Dumping', Letter to Editor, *Wireless World*, Vol. 89, No. 1575, pp.49-50, December 1983

18 Vanderkooy, J. and Lipshitz, S.P., 'Current Dumping Review', Letter to Editor, *Wireless World*, Vol. 90, No. 1577, p. 49, Feb. 1984

19 Hawksford M.J., 'The Essex Echo: Reflexion', *Hi-Fi News and Record Review*, Vol. 30, No. 12, pp.35-40, Dec. 1985

20 Takahashi, S. and Tanaka, S., ' Design and construction of a feedforward error-correction amplifier', *JAES*, Vol. 29, No. 1/2. pp. 31-47, Jan/Feb 1981

21 Walker, P.J., 'Current dumping audio power amplifier', *Wireless World*, Vol. 81,

pp.560-582, Dec. 1975

22 Sandman, A.M., 'Reducing distortion by error add-on', *Wireless World*, Vol. 79, pp. 32, Jan. 1973

23 Cherry, E.M., 'A new result in negative feedback theory and its application to audio power amplifiers', *Int. J. Circuit Theory Appl.*, Vol. 6, pp. 265-288, July 1978

24 Frey, D.R., 'A new technique for minimising distortion', *JAES*, Vol. 31, No. 5. pp. 320-325, May 1983

25 Mills, P.G.L. and Hawksford M.J., ' Transient analysis : a design tool in loudspeaker system engineering', 80th Convention of the AES, Montreux, March 1986, preprint 2338 (E8).

Dr M.J. Hawksford
Department of Electronic Systems Engineering
University of Essex
Wivenhoe Park
Colchester
Essex
England

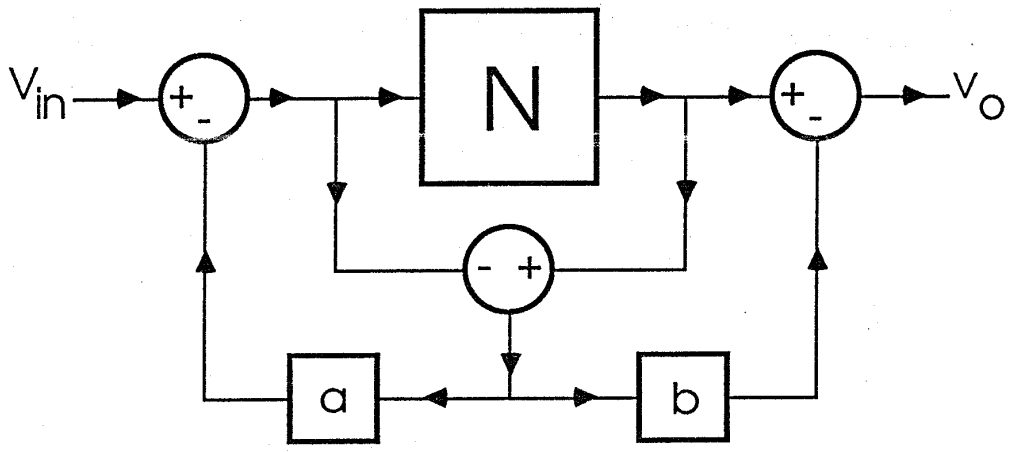


Fig. 1.1 Error feedback/feedforward distortion correction.

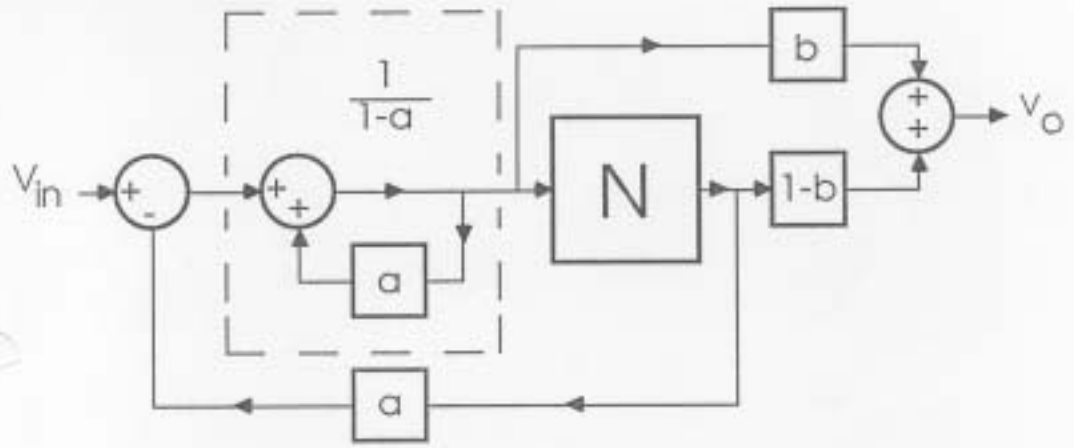


Fig . 2.1 Reconfiguration of feedback/feedforward correction scheme.

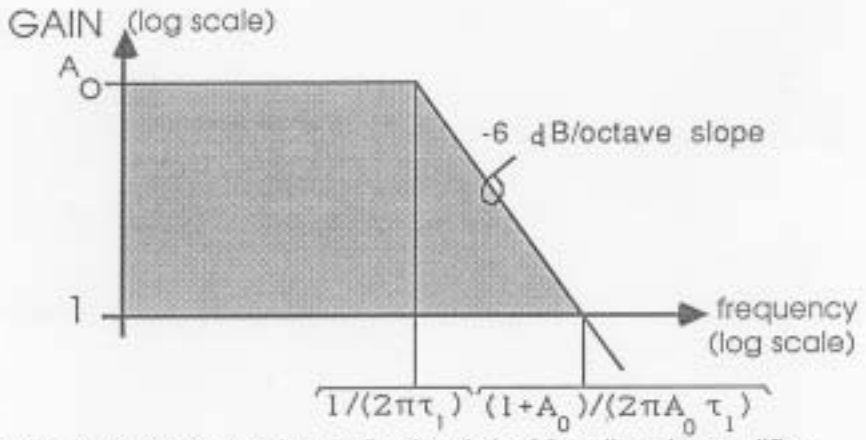


Fig. 2.2 Typical forward gain Bode plot of feedback amplifier.

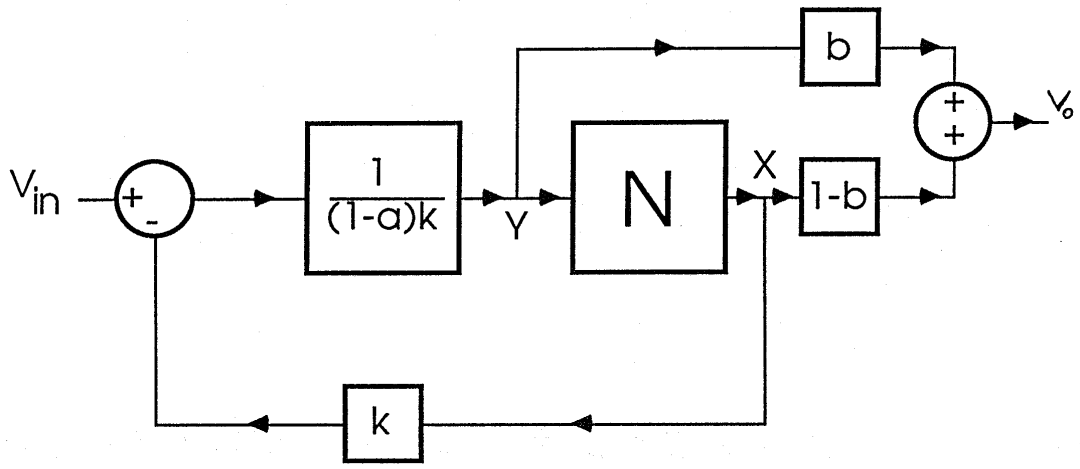


Fig. 2.4 Error correction amplifier with voltage gain.

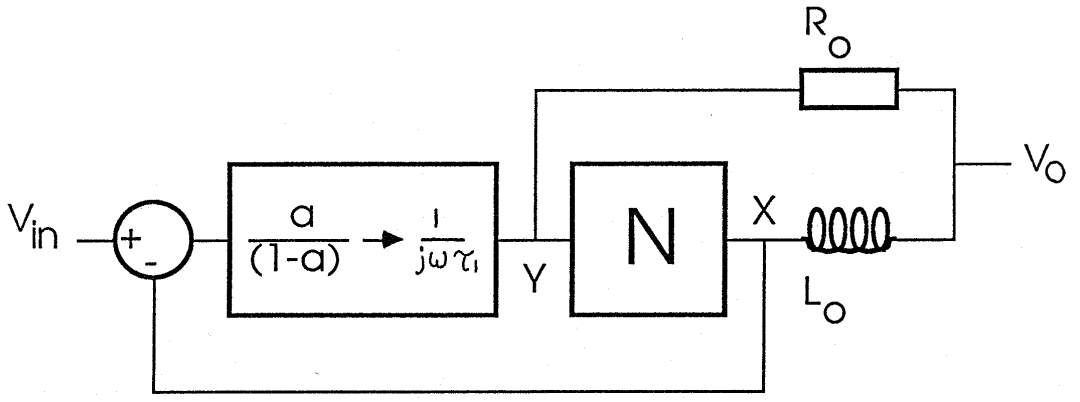


Fig. 2.3 Error summation using passive R , L_o network.

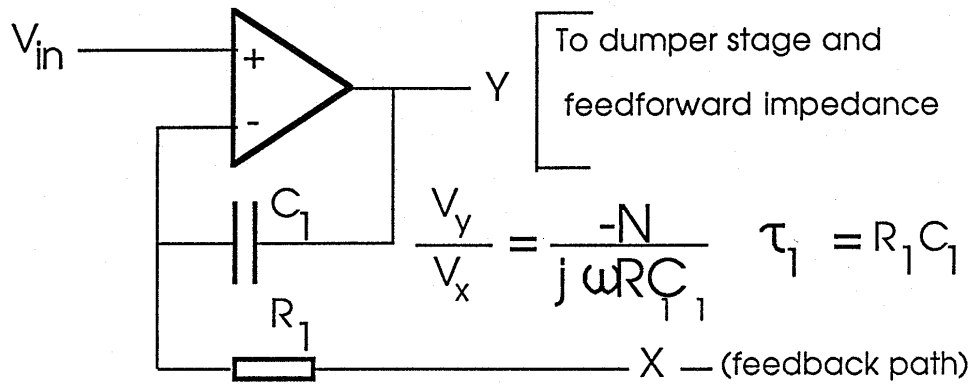


Fig. 2.5a Basic virtual-earth integrator in feedback path of current dumping amplifier.

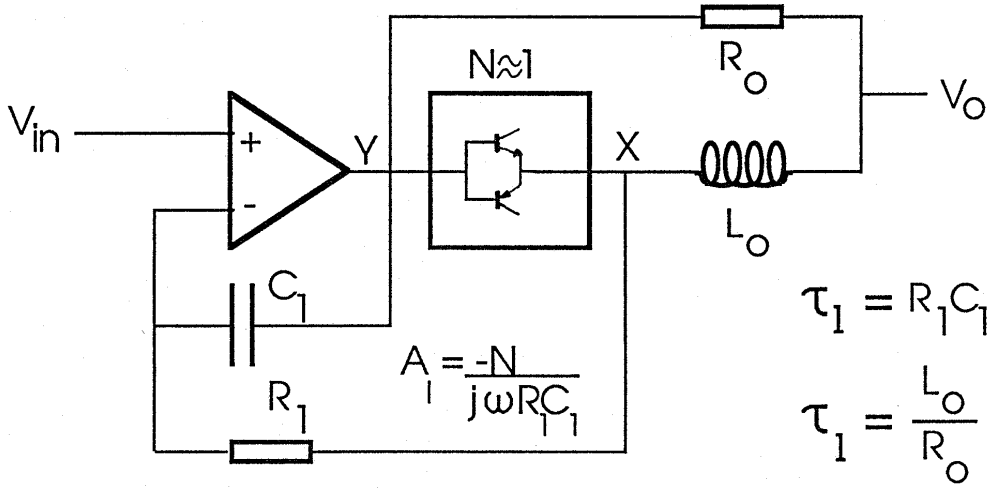
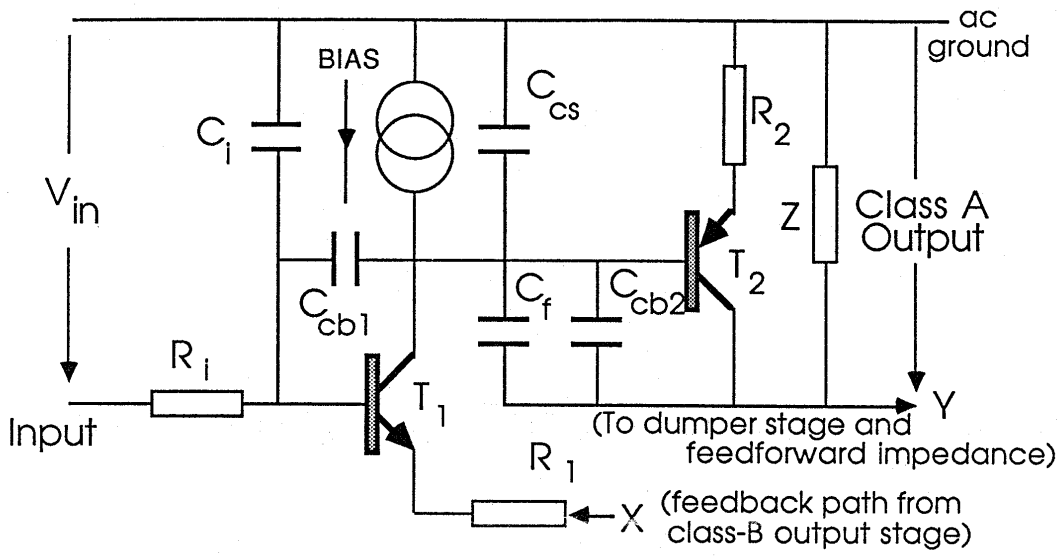


Fig 2.5b Elementary current dumping amplifier showing bridge components: R_1, C_1, loop integrator; R_0, L_0, output summation.



C_i input band limiting capacitor (in conjunction with R_i)

C_{cb1}, C_{cb2} collector-base capacitance of T_1 and T_2

C_{cs} effective capacitance of current source

C_f feedback capacitor

T_1 input transistor

T_2 composite emitter follower buffers and inverter stage

Z effective load impedance presented to the collector of T_2 due to class-B stage and biasing circuitry

Fig. 2.5c Simplified class-A amplifier topology emphasising transistor and feedback capacitors.
(see ref. 1 and 16).

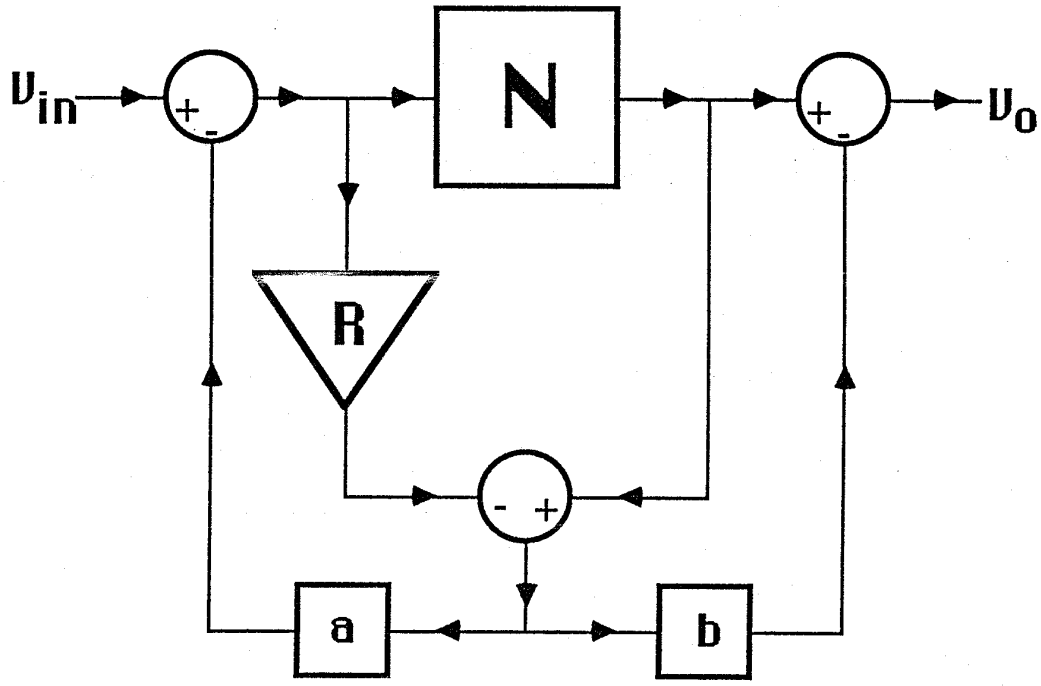


Fig. 3.1 Error feedback/feedforward with non-linear cell, exhibiting overall gain (ie. N>1).

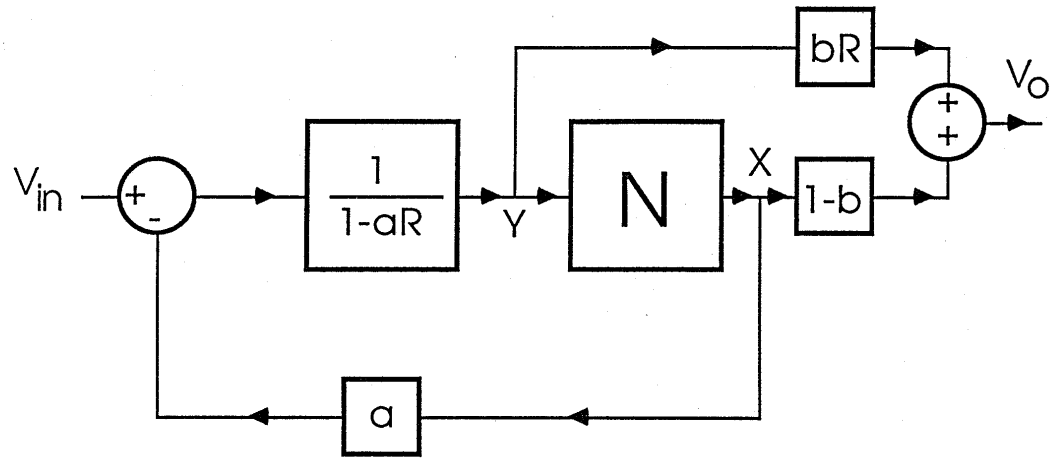


Fig. 3.2a Reconfiguration of Fig. 3.1 feedback/feedforward structure..

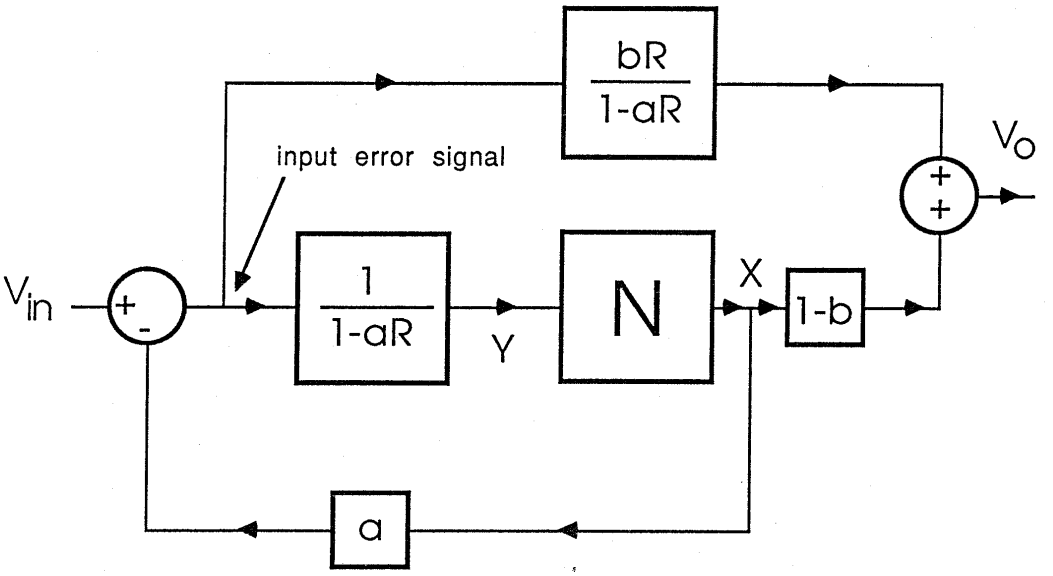


Fig. 3.2b As Fig. 3.3a but with feedforwakd path referred to input error signal.

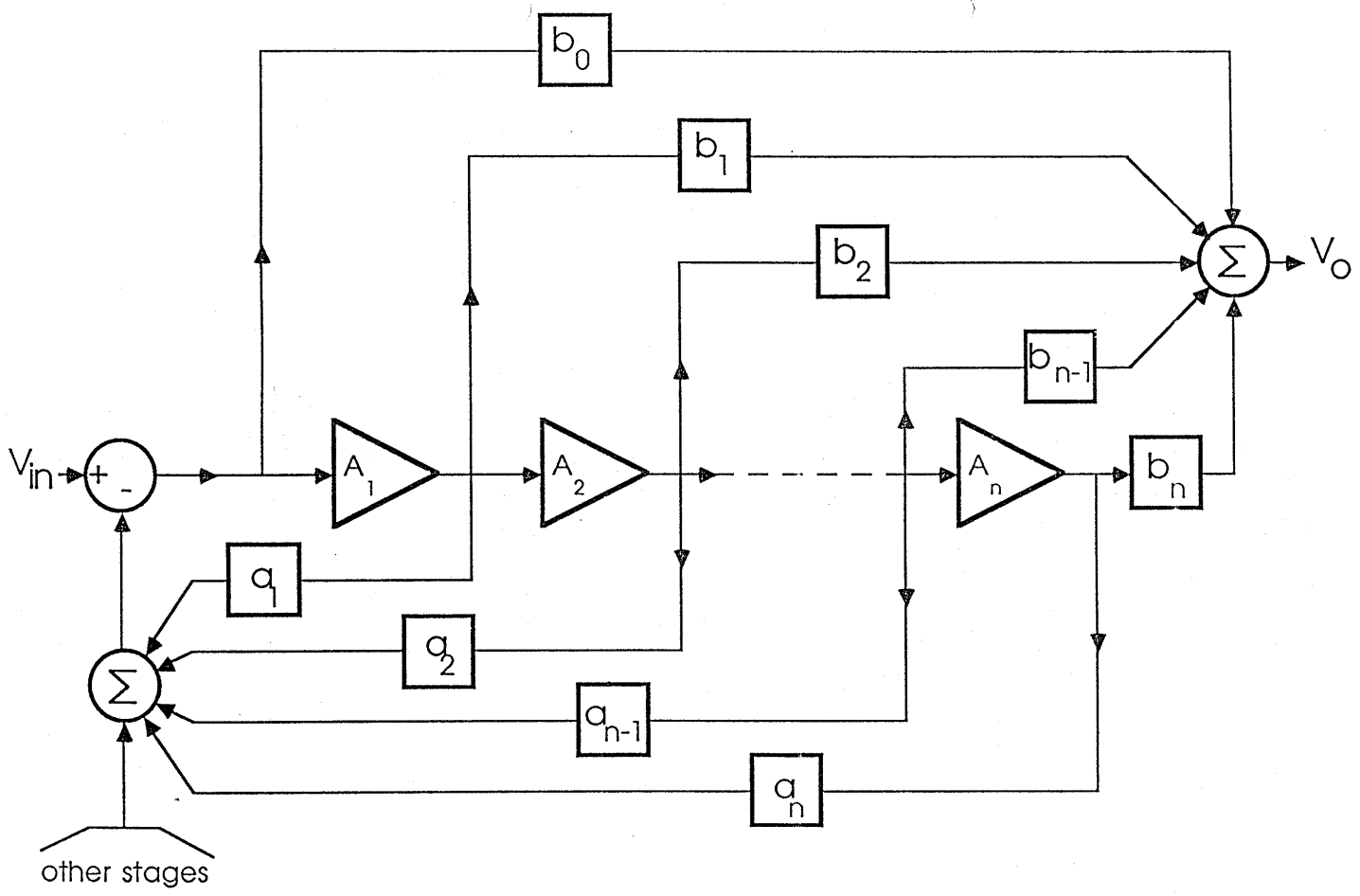


Fig. 4.1 N-stage generalised feedback/feedforward topology.

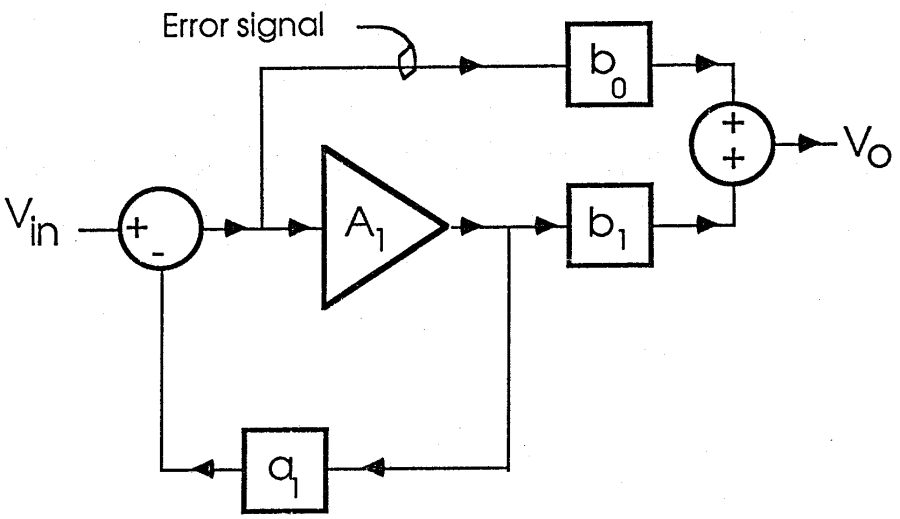


Fig. 4.2 Basic conceptual model of Sandman's error-pickoff (22).

Relationships Between Noise Shaping and Nested Differentiating Feedback Loops 3358 (B-4)

J. Vanderkooy
University of Waterloo, Department of Physics
Waterloo, Ontario, Canada

M.O.J. Hawksford
University of Essex, Department of Electronic Systems Engineering
Wivenhoe Park, Colchester, UK

Presented at the 93rd Convention 1992 October 1–4 San Francisco

AN AUDIO ENGINEERING SOCIETY PREPRINT

Relationships between noise shaping and nested differentiating feedback loops

J. Vanderkooy
Department of Physics
University of Waterloo, Ontario
M.O.J. Hawksford
Department of Electronic Systems Engineering
University of Essex, England

Abstract

We study the application of heavy feedback in two different topologies, namely multiple-order noise shaping and nested differentiating feedback loops. Both have similar loop gain and stability considerations, although the two approaches have different implied circuit environments and areas of application. In noise shaping, emphasis is placed on the integrator characteristics of each gain stage, whereas leaky integrators or flat-gain stages with high-frequency poles form the usual basis of the nested differentiating loop concept. This short didactic paper helps in understanding the application of large amounts of feedback to control noise or distortion at baseband frequencies.

1 Introduction

The concept of nesting differentiating feedback loops (NDFL) has been introduced and promoted by Cherry [1, 2]. Generally we think of analogue circuits for its application, and the basic idea is that for the stage that creates the most distortion (usually a power output stage), the encompassing differentiating feedback stabilises the loop at high frequencies while allowing increased feedback at lower frequencies to significantly reduce system distortion.

2 Noise shaping

Our approach to the application of large amounts of feedback has come from a digital perspective, namely that of noise-shaping, as typically applied to requantizers in one-bit ADCs and DACs; for reference, see Hawksford [3]. Consider the first and second-order noise shapers shown in Figure 1. Figure 1(a) shows two topologies in which the error of the quantizer Q is modified by a single delay for the first-order shaper (there must always be at least a one-sample delay) and the filter $H(z) = z^{-1}(2 - z^{-1})$ for the second-order shaper. In each case the filtered error is subtracted retrospectively from the input sequence to enable a partial correction for the quantizer error. Figure 1(b) shows the same topologies redrawn in a different form, while Figure 1(c) shows the transformation to a delta-sigma converter topology. The digital integrator blocks are shown in the insert, and it is clear that the order of the noise shaper can be increased by simply extending the number of integrator blocks, encompassing each of them in a feedback loop from the output. It is readily shown that these noise shapers have signal transfer functions of unity (except for the simple delays, which could be removed in some of the circuits) and the quantization error (normally considered as noise) is shaped by a response

$$R(z) = (1 - z^{-1})^N \qquad \ldots 1$$

where N is the order of the shaper. Such noise shaping structures have been much discussed in the literature as well [4].

3 Distortion shaping

Figure 2 illustrates a further progression of ideas. Figure 2(a) shows a third-order digital shaper for which the quantizer output-related error q[n] is explicitly drawn in, while Figure 2(b) shows an equivalent analogue circuit using real integrators, in which the quantizer is

replaced by an output stage.

We can interpret this circuit as a limiting digital case in which the sampling frequency has become infinite, hence removing all delays, and where the discontinuous output quantizer is replaced with a continuous but nonlinear analog output stage. If the output stage has gain G and the integrators have respective time constants τ_{1f}, τ_{2f} and τ_{3f} for the feedback paths, τ_{1s}, τ_{2s} and τ_{3s} for the signal paths, then the transfer function of the system can be written as

$$\left[\frac{\tau_{1s}\,\tau_{2s}\,\tau_{3s}\,s^3}{G} + \frac{\tau_{1s}\,\tau_{2s}\,\tau_{3s}\,s^2}{\tau_{3f}} + \frac{\tau_{1s}\,\tau_{2s}\,s}{\tau_{3f}} + \frac{\tau_{1s}}{\tau_{1f}}\right] Y(s) = X(s) + \frac{\tau_{1s}\,\tau_{2s}\,\tau_{3s}\,s^3}{G}\,Q(s) \qquad \ldots 2$$

for which the quantizer output-related error q[n] is explicitly drawn in.

We note that the τ parameters define both the stability criteria and signal frequency response of the circuit, and that the output error Q(s) is weighted by the cube of the signal frequency. Consequently, this analogue feedback circuit can be viewed as a "distortion" shaper, where the multiple integrators by virtue of their large low-frequency gain, reduce the effects of nonlinearity at lower frequencies.

In the circuit of Figure 2(c) the feedback paths to the last two integrators have been moved forward, thus each incorporating an extra integrator, and compensating differentiators $s\tau_{2s}$ and $s\tau_{3s}$ have been inserted to keep the transfer function unaltered. An extra amplifier with gain g = 1 has also been inserted to give the circuit a more usual configuration. It is this modified circuit which we wish to compare with the NDFL circuits of Cherry [1].

In this reconfiguration, the output stage is modelled with frequency-independent gain G, but the dashed bracket associates the third integrator with the output stage so that it can display real poles. It is also possible to think of each integrator as being "leaky", of form a/(b + s), more like the actual stages of an amplifier. We might associate the first integrator, if present, with an intermediate amplifier stage, the second integrator with the voltage-gain stage, and the third integrator and output block with a more realistic output stage.

4 Discussion

The circuit of Figure 2(c) can be identified directly with NDFL circuits, even though there may be differences of small detail. The nested loops may not all take their feedback signal directly from the output, for example, although there is then a difference in implied circuit environment. In Figure 2(b), each of the integrators (which could be leaky) are regarded as similar, but in Figure 2(c) the last integrator is considered to be part of an output stage, having relatively low gain, and being rather leaky, perhaps describing the typical emitter follower output stage of an audio amplifier. Although it is customary to have a Miller capacitor across the voltage-gain stage, if this capacitor also encompasses the output stage, and feeds back to the virtual ground input of the voltage-gain stage, it becomes the $s\tau_{2f}$ differentiating feedback shown in the diagram. The effect on amplifier distortion is very beneficial, and this point has been emphasised by Cherry [2].

There are other strong parallels between NDFLs and "distortion- shaping" topologies. The order of each can in principle be increased without limit by adding more stages, giving even

more feedback at lower frequencies. Another aspect is the internal stability of the loop with respect to the output stage. It is the output stage which generally is considered the dominant source of the distortion, and a high loop gain will act to reduce it. The output-stage loop gain A in the circuit of Figure 2(b) (and hence also of Figure 2(c)) can (setting the input X(s) = 0) be shown to be given by,

$$A = \frac{Q(s)}{Y(s)} - 1 = G\left[\frac{1}{s\,\tau_{3f}} + \frac{1}{s^2\,\tau_{2f}\,\tau_{2s}} + \frac{1}{s^3\,\tau_{3f}\,\tau_{2s}\,\tau_{3s}}\right] \qquad \ldots 3$$

and this is easily generalized to higher or lower order. At the highest frequencies, only the first term survives, and for good stability the phase shift must be considerably less than -180°, ideally being close to -90°, the lag of a single integrator. It is the τ_{3f} feedback path that ensures this stability at high frequencies, but the other loops allow increasing feedback at lower frequencies, reducing distortion in the process.

As an example, let us consider a 5th-order distortion shaper feedback circuit, for which G = 1, and the transfer function Y(s)/X(s) is by proper choice of τ parameters a 5th-order Butterworth unity-gain low-pass with cut off frequency of say 50 kHz. The loop gain can be written as,

$$A = \frac{s_n^5 + a\,s_n^4 + (a+2)\,s_n^3 + (a+2)\,s_n^2 + a\,s_n + 1}{s_n^5} - 1$$

$$= \frac{a}{s_n} + \frac{a+2}{s_n^2} + \frac{a+2}{s_n^3} + \frac{a}{s_n^4} + \frac{1}{s_n^5} \qquad \ldots 4$$

where, $a = 1 + 2\cos(\pi/5) + 2\cos(2\pi/5)$

and the normalized Laplace variable $s_n = s/\omega_0 = s/(2\pi f_0)$, and f_0 = 50 kHz. Figure 3 is a plot of equation (4), showing the basic 6 dB/octave rolloff of the loop gain at high frequencies, but with a rise of $1/f^5$ for lower frequencies. If the integrators in the circuit are leaky, the graph will look similar, but will limit near dc at some high value of gain, giving a loop gain for the output stage very similar to that indicated by Cherry [1]. In this example, unity gain occurs at 157 kHz and the attendant phase shift is - 120°, representing good stable behaviour. Note that at 20 kHz, there is already 40 dB of distortion reduction, rising in an ever-increasing way at lower frequencies.

For high orders such as discussed above, this system displays conditional stability, as pointed out by Cherry [1] and in a recent discussion of NDFL [5]. Hence it is important to consider large-signal clipping behaviour experimentally to see if the system can be provoked into self-oscillation or other bizarre behaviour. It is clear that there is no simple limit to the amount of distortion reduction at high orders, but increasing care must be taken in defining stable circuit parameters and behaviour for large signals.

5 An Overview

At first glance, it almost seems that the trivial reassignment of the feedback loops around the integrators in Figure 2(b) results in the NDFL structure of Figure 2(c). However, there are differences in realizability and circuit tradition. Cherry [1] also works out a great deal of the mathematical aspects of NDFL, with sensitivities and analysis of appropriate models.

Presumably most of this work applies also to the distortion-shaping topology as well, with appropriate measuring points or circuit associations in the two approaches. We do not in this didactic paper work out such details or attempt a mapping between the two topologies.

In some ways, the distortion-shaping approach is easier to grasp initially. But NDFL is perhaps better if one is faced with a traditional class AB audio power amplifier, and wishes to improve its performance. In fact the title of one of Cherry's papers [6] suggests this. Also, when NDFL was invented, it appeared as a new result since the traditional methods of stabilising amplifiers had limitations on the amount of simple feedback that could be applied, as Bode [7] had shown. There have been engineers who have employed selected aspects of a differentiating loop all along, but the general concept of NDFL puts a firm footing on new aspects of feedback. What this paper shows is that there is another way of looking at the application of large amounts of negative feedback, and that the two are closely related.

References

[1] **E.M. Cherry**, "A new result in negative-feedback theory and its application to audio power amplifiers", *IEEE J.Circuit Theory and Applns.* **6** 265-288 (1978).

[2] **E.M. Cherry**, "Nested differentiating feedback loops in simple power amplifiers", *JAES* **30** 295-305 (1982 May).

[3] **M.O.J. Hawksford**, "Chaos, oversampling and noise shaping in digital-to-analog conversion", *JAES* **37** 980-1001 (December 1989)

[4] **M.W. Hauser**, "Principles of oversampling A/D conversion", *JAES* **39** 3-26 (1991 Jan/Feb)

[5] **J.Scott and G. Spears**, " On the advantages of nested feedback loops', *JAES* **39** 140-147 (1991 March) and E. M. Cherry, discussion in this paper.

[6] **E.M. Cherry**, "A power amplifier improver" *JAES* **29** 140-147 (1981 March).

[7] **H.W. Bode**, "Network analysis and feedback amplifier design" van Nostrand Princeton N.J. (1945)

Figure captions

Figure 1 A progression of equivalent circuits for first-order and second-order digital noise shapers. The left column shows equivalent first-order shapers, the right column second-order. The inset in (c) shows the construction of the digital integrator block. The configurations in (c) show how to extend the order to any desired number.

Figure 2 Derivation of the nested-differentiating feedback loop (NDFL) concept. The digital third-order noise shaper in (a) is converted to analogue form in (b). By moving the feedback paths labelled τ_{2f} and τ_{3f} in (b) to the left, and adding compensating differentiators to maintain the same circuit transfer function, we achieve the NDFL representation of (c). The dotted lines in (c) encompass what we might consider a realistic output stage of a normal amplifier.

Figure 3 Loop gain as seen by the output stage, of a fifth-order noise shaper having a transfer function of a fifth-order Butterworth low-pass filter with cutoff frequency of 50 kHz. The rising loop gain at low frequencies vastly reduces distortion at these frequencies and effectively introduces "distortion shaping".

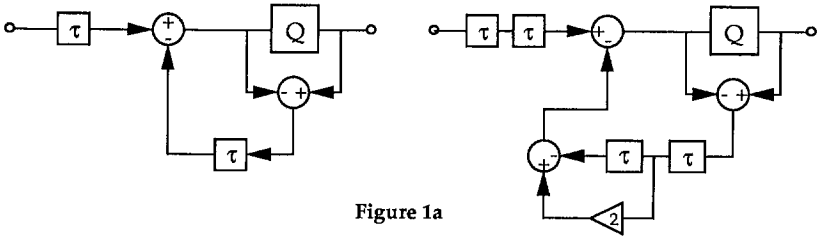


Figure 1a

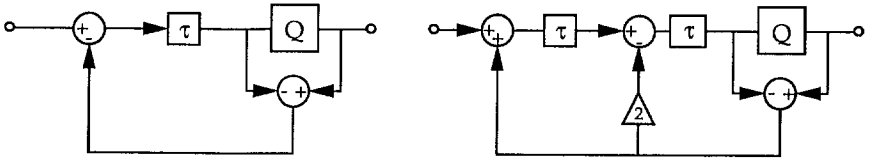


Figure 1b

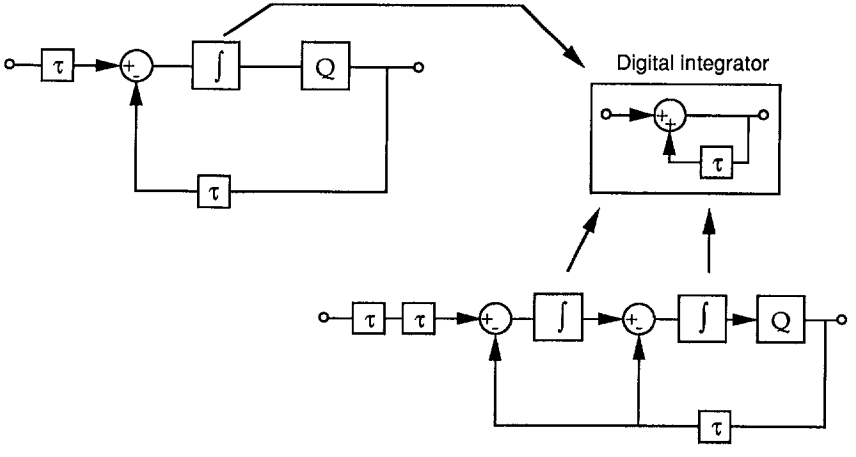


Figure 1c

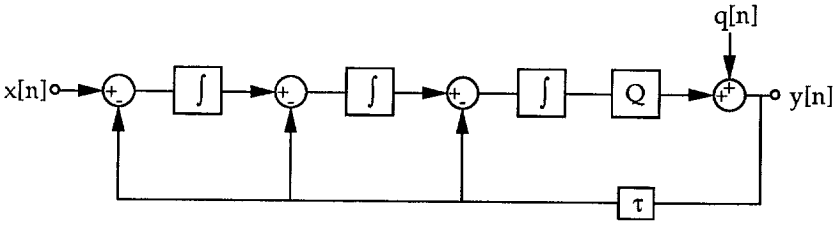


Figure 2a

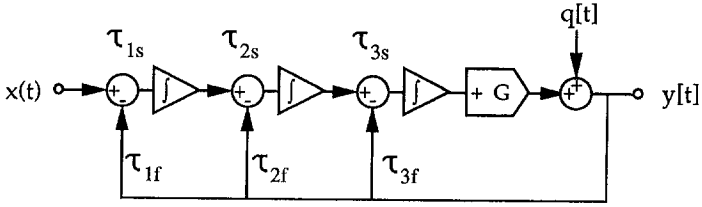
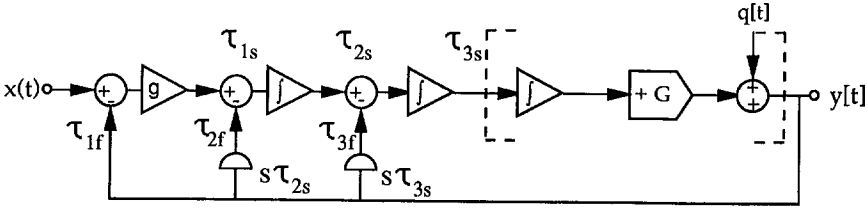


Figure 2b



(c)

Figure 2c

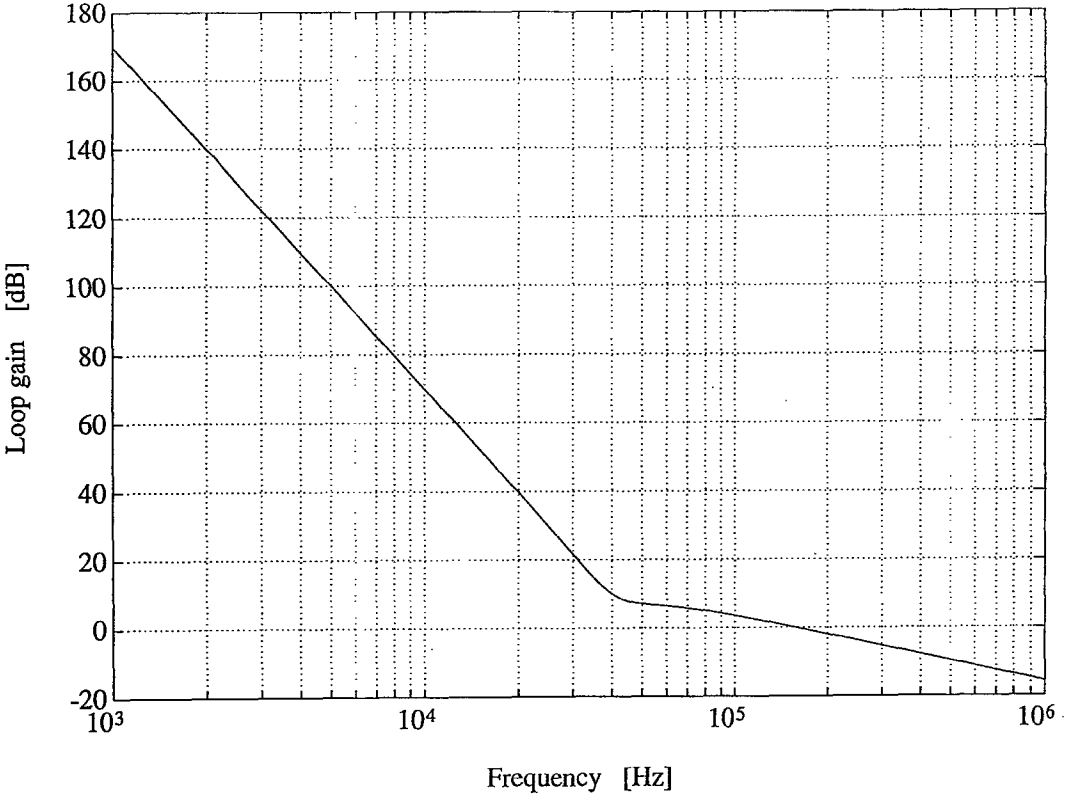

180
160
140
120
100
80
60
40
20
0
-20
Loop gain [dB]
10³
10⁴
10⁵
10⁶
Frequency [Hz]

Figure 3

"SOUND THE CHARGE" HFN/RR

Vol 33, No 5
pp 338 39 May 1988

LFD MC Preamplifier 1988

The philosophy of the LFD MC preamplifier evolved about six years ago, shortly followed by the first prototype circuit. Since then, considerable effort and many arduous hours of subjective appraisal have been directed towards component selection and constructional technique, a task for which my friends and associates, Richard Bews and Paul Mills, must take substantial credit. Interestingly, the initial circuit topology has deviated little during this time, witnessing only subtle revision.

It was appreciated from the project commencement that, if an optimum component selection was to be achieved, then the number of parts influencing sound quality must be minimised, within the bounds of engineering requirements. If a circuit is relatively complicated and includes contributions from power supply and decoupling components, the effect of individual circuit elements appear to become diluted and sonic optimisation may prove illusive.

The design aim for LFD was to achieve a low noise, wide band circuit with, for personal ideology, a relatively high signal power coupling to the input devices derived from a low source impedance moving coil (MC) cartridge, with no ac coupling anywhere within the signal path. Also, the circuit must be capable of achieving output signals circa 500mV from cartridges such as the Koetsu without recourse to a pre-preamplifier or line stage.

Recently there has been debate on the effect and quality of power supply components and interaction with mains circuitry. It was evident that an ideal solution to achieve total mains isolation was to use rechargeable cells to replace the more conventional power supply. Such a strategy is not, of course, new and has been employed in several preamplifiers. However, in most of these systems, the cells are just replacements for the mains supply and, as such, achieve only some of the advantages on offer using battery power.

The LFD preamplifier imbeds high capacity, lead acid cells configured into a distributed and partially floating topology, where individual power sources are localised within a symmetrical structure. This method immediately eliminates interstage interaction via a common supply and allows well-defined local signal current loops that do not flow in the ground rail. As such, a single ended circuit can achieve most of the advantage of a fully balanced topology but without the disadvantage of duplicating circuitry. The circuit is biased from a single floating 2V cell and requires no current sources, where the only active circuitry is that used for signal amplification.

Lead acid cells form almost perfect wide-band supplies. However, even if the cells exhibit non-ideality, in LFD, circuit symmetry and cell placement dramatically desensitises performance dependence on supply induced errors. Consequently, a circuit has evolved that uses no electrolytic capacitors or other decoupling and coupling elements and where the only capacitors are for RIAA EQ and for setting output dc conditions to a fraction of a mV.

LFD follows the route of passive RIAA EQ [ref. 1, 2, 3] where the input stage operates as a transconductance amplifier accepting signals directly from a MC cartridge with only local feedback applied; the output current is then converted to a voltage using just the RIAA EQ network. The design of the input cell must offer a wide-band low-distortion characteristic and yield a very high output impedance to circumvent both RIAA EQ inaccuracies and the effects of inherent transistor non-idealities. Also, the input transistors should operate with near constant collector-emitter voltage to minimise distortion through internal transistor slope impedance modulation.

The main contenders for low-feedback input stages are the single transistor common emitter amplifier, the long-tail pair and the complementary common emitter stages, as illustrated in Figure 1.

The common emitter stage was rejected because of both poor distortion characteristics and the difficulty in achieving dc coupling within a low feedback cell. Also, you may recall that this cell was previously used to illustrate a distortion mechanism [ref. 4, see article also for further background to early LFD philosophy] that has been occasionally employed to artificially enhance signal dynamics, where a low bias current and low feedback can be used to achieve a dynamically enhanced sound. Douglas Self [ref. 5] has, however, shown an elegant method for reducing distortion in a common emitter stage using a high overall feedback operational amplifier topology in association with a low impedance feedback network to minimise noise, though ac coupling is retained. The long-tail pair overcomes the dc biasing problem by allowing V_{BE} cancellation. However, the distortion performance, although substantially better than the single stage, still exhibits high-level saturation when one or other transistor is starved of current. Also, there is a noise penalty as the two transistors are in series. Again, a high feedback method of distortion reduction, similar in concept to the Self circuit, has been shown by Walter Jung [ref. 6].

The third circuit using the complementary configuration is the one chosen for LFD due to its enhanced overload performance and parallel transistor connection that lowers noise levels. Also, dc coupling is possible with almost complete base current cancellation as current flowing out of the base of the PNP device flows into the base of the NPN device. (In LFD, small beta imbalances are further nulled to reduce the possibility of cartridge magnetisation.)

The principal source of non-linearity in the common-emitter cell is I_E/V_{BE} curvature which follows approximately an exponential law. However, when complementary devices are paralleled [ref. 7], this non-linear law is modified to advantage and realises a very acceptable lower overall distortion characteristic. Indeed, the lower the inter-emitter resistance in this configuration, the lower is the distortion for a given transconductance defining resistance, within the bounds of class A working. The effective emitter resistance R_{eff} of the compound cell can be written [ref. 8],

$$R_{eff} = R_o + r(i)$$

where R_o is the effective resistance under quiescent conditions and r(i) is the change in resistance when the cell conveys a signal current i. A distortion factor D(i) can then be defined as D(i) = $r(i)/R_o$. In Figure 2, a set of computer generated curves is shown to illustrate the trend in behaviour of D(i) with signal current and effective inter-emitter resistance.

The basic LFD circuit topology is shown in Figure 3, where to increase the output impedance of the input stage, a modified form of cascode is employed where the transistor bases are effectively driven by the emitters of the input transistors. An important characteristic of this cell is that the base currents of the second cascode stage are not returned to ground, but form closed paths that bypass the RIAA EQ network. Consequently, the cell offers an extremely high output impedance that has only a second-order dependence upon the output slope characteristics of the transistors - even if these are non-linear and frequency dependent under large signal conditons. The cell which floats on the input signal is therefore an ideal match to the transconductance fed RIAA equalisation network and approaches the theoretical optimum within the bounds of RIAA network tolerance.

The circuit is completed by a high current gain, unity-gain output buffer similar in form to the input stage, which enables interconnect/power amplifier to be driven directly via a hard wired, precision switched passive attenuator, which incorporates no more than two resistors at any of the 29 positions. The output cell uses a local ±6V supply which is independent of the input stage. This supply also powers the loosely coupled servo-amplifier and establishes an output offset of typically <0.2mV. The servo is an integrator in association with a compensation network designed to yield minimum distortion of the RIAA EQ accuracy [ref. 4].

Each channel of LFD is totally independent and housed in individual enclosures, as is the comprehensive cell management system, where Figure 4 illustrates the complete system and Figure 5 shows a composite graph of typical measured performance, presenting both the spectral noise floor and 19/20kHz IM performance.

The functionality of the cell management system (CMS) endows the preamplifier with most of the advantages of a mains powered system but with the added advantage of complete circuit isolation that prevents mains borne interference being injected into the signal path. CMS monitors the state of the preamplifier ON/OFF controls (one per monoblock) and also the state of charge of each preamplifier (in fact two independent CMS circuits are provided that are each dedicated to a preamplifier channel). The system automatically charges the preamplifier at appropriate rates when in the OFF state so as to maintain optimum cell condition but is fully isolated in the ON state, where CMS is designed to remain permanently mains powered. Philosophically, it is interesting to observe that the complexity found in many preamplifier circuits of smoothing, regulation, decoupling and current source biasing which potentially can degrade sound are now completely removed from the active circuitry and located under modified guise in the CMS, but in total isolation. Of course, the advantages of an isolated, distributed and partially floating supply are not

easily replicated in a mains powered system.

The only "dangerous" state that a battery powered system can experience is when the system is inadvertently left in the ON state for an extended period. Although the preamplifier will run for about two days from full charge, an internal alarm sounds after about sixteen hours, warning that the unit should be switched OFF, whereby CMS automatically takes control of charging in readiness for the next listening session.

To conclude this summary of the engineering and design aspects of the LFD preamplifier, the following features are highlighted:

* input amplifier cell is optimised for wide bandwidth, low distortion operation with only local feedback

* input transistors operate with near-constant collector-emitter voltage under dynamic conditons

* second stage of input cell offers a very high output impedance that is not adversely affected by non-linear output slope impedance modulation of transistors under large signal conditions and allows RIAA EQ to be defined by a single passive network

* virtual elimination of the effect of noise sources in second stage of input cell

* any modulation of the floating, 2V input bias cell, whether linear or non-linear (though negligible in practice) results in a circulating current that bypasses the RIAA EQ network, thus causing virtually zero contribution to output signal

* modulation of the floating ±6V supply (though negligible in practice) has negligible effect on the output signal due to the very high output impedance of the transconductance stage

* the signal current circulating through input cell, floating supply and RIAA EQ forms a well-defined closed path that excludes the ground line, thus secondary errors are minimised which further augments the use of star grounding to a single mecca

* similarly, output currents demanded by interconnect capacitance or a load impedance form a closed path including a local supply that is independent of the input circuitry

* due to the low number of circuit elements that actively contribute to sound quality and the nature of these components, "warm up" time is very short, a mandatory requirement in a battery powered system (I'm sure Ken would not want to keep Ella waiting!).

* a frequency compensated servo, that is loosely coupled to the main preamplifier circuit, offers an accurate first-order, low frequency, high pass response with minimal RIAA EQ error, together with a low output offset voltage (typically) <0.2mV to prevent attenuator switching transients and enable dc coupling to an appropriate power amplifier with no low frequency roll off; due to attenuator location, dc offset falls progressively below 0.2mV with increased attenuation

* hard wired, single strand PTFE coated LC conductors for signal routing.

Although this article is a discussion of the engineering aspects and design of the LFD MC preamplifier, the considerable time and effort contributed by my associates in prototype construction and component selection is equally important with respect to total system development. The preamplifier system is relatively complicated to construct and requires much hand assembly. In particular, the hard wired switched attenuators which are assembled personally by Richard Bews are a true labour of love and extremely time consuming.

Although the circuit topology was designed to minimise most secondary dependencies and to match my own personal idiosyncracies on low-level amplifier stage design for MC applications, it has proved a sensitive tool for the subjective appraisal of RIAA passive components, the results of which are now incorporated into the latest version.

It is envisaged that several of the techniques developed for the disc amplifier will be applied to a complete amplifier system. At present, however, the disc stage can be used to interface appropriate cartridges and amplifiers without recourse to pre-preamplifiers or line stages, whose addition represents unnecessary complication.

Finally, for information, Richard Bews has now established LFD Audio, to manufacture limited numbers of the preamplifier system shown in Figure 4, together with a passive preamplifier that can be personalised to individual requirements and incorporates the same hand assembled, hard wired switched attenuator used in the MC preamplifier.

References

1 Y. Miloslavskij, "Audio preamplifier with no TID", *Wireless World*, **Vol.85**, No. 1524, pp.58-60, August 1979

2 I. Humphreys, "Getting the Pip", *HFN/RR*, **Vol.29**, No. 3, pp.19-21, March 1984

3 O. Jones, "About the genesis of the Pip", (Letter to Editor, HFN/RR, **Vol.30**, No.12, p.25, December 1985

4 M.J. Hawksford, "The Essex Echo: on errors, low feedback and fuzzy distortion", *HFN/RR*, **Vol.29**, No.9, pp.37-41, September 1984

5 D.R.G. Self, "Ultra-low-noise amplifiers and granularity distortion", *JAES*, **Vol.35**, No.11, pp.907-915, November 1987

6 W.G. Jung, "Audio IC Op-Amp Applications", Howard W. Sams and Co. Inc., ISBN 0-672-21558-6, p.128

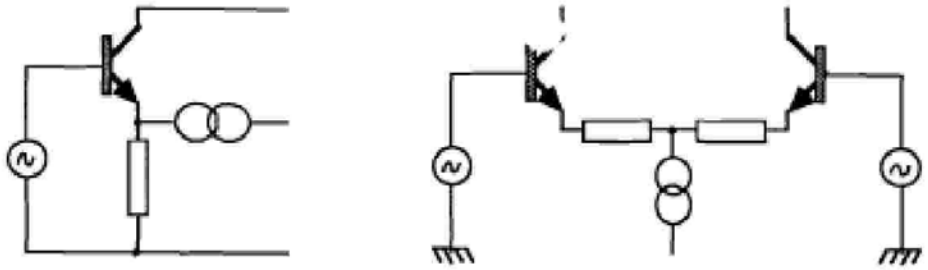
7 B.M. Oliver, "Distortion in complementary-pair class-B amplifiers", *Hewlett Packard Journal*, **Vol.22**, No.8, pp.11-16, February 1971

8 M.J. Hawksford, "Distortion correction circuits for audio amplifiers" *JAES*, **Vol.29**, pp.503-510, July/August 1981

Dr M.J. Hawksford
Department of Electronic Systems Engineering
University of Essex

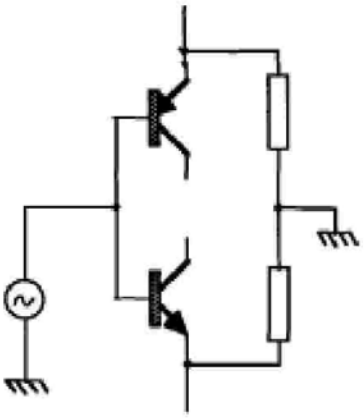
2nd February 1988

MJH/lf



(a) Common emitter cell

(b) Long tail pair



(c) Complementary cell

Fig.1 Basic input stage configurations

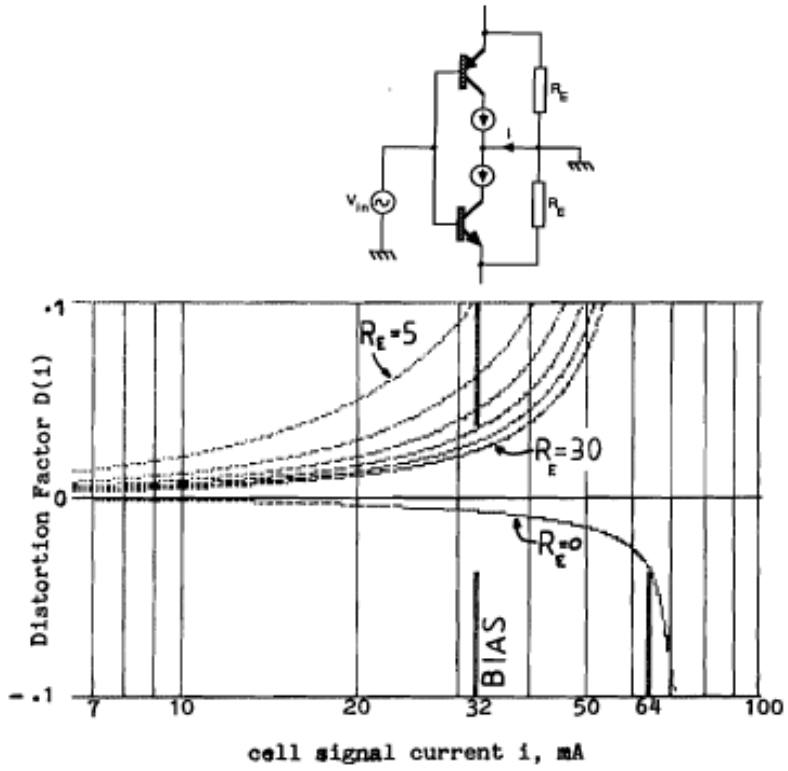


Figure 2 Plot of distortion factor D(i) as a function of signal current i for a quiescent bias current of 32mA and emitter resistance of 0Ω to 30Ω in steps of 5Ω for complementary emitter follower stage

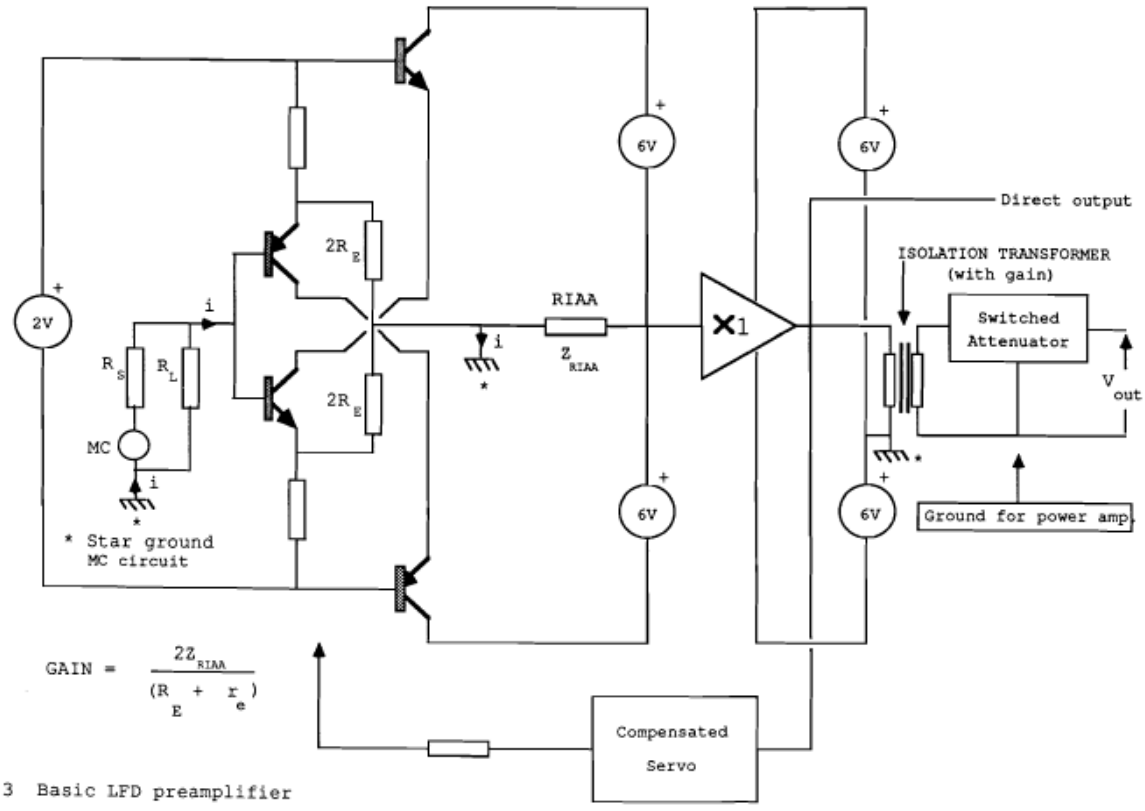


Fig 3 Basic LFD preamplifier

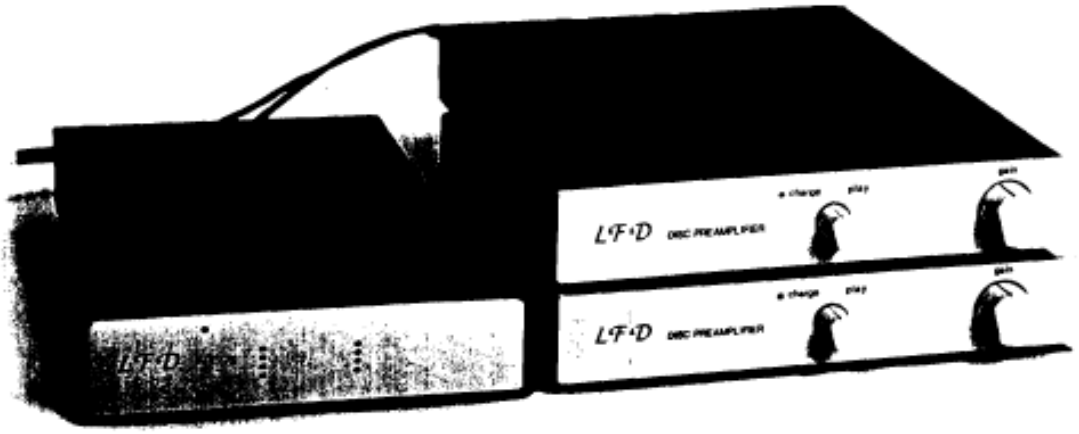


Figure 4 Complete LFD preamplifier showing CMS and the two monoblocks

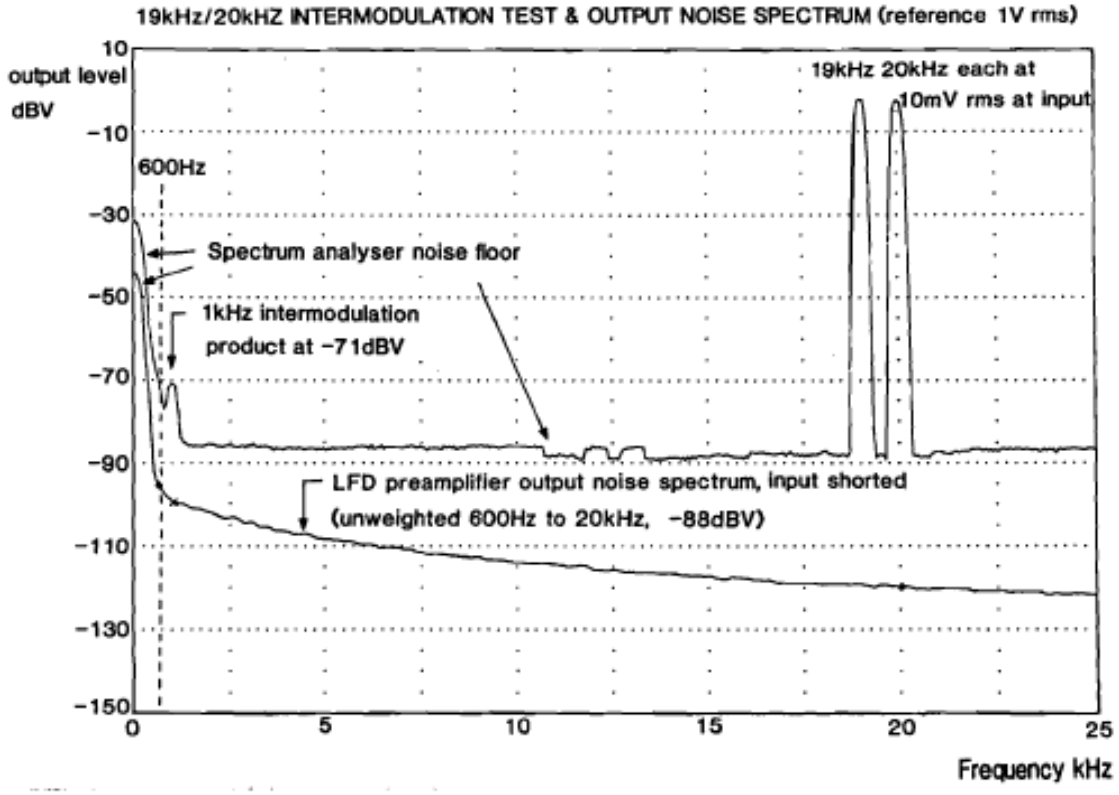


Figure 5 Typical spectral noise floor and two tone intermodulation performance of LFD

M. O. J. Hawksford
Centre for Audio Research and Engineering
University of Essex
Colchester, Essex CO4 3SQ, UK

Presented at
the 109th Convention
2000 September 22-25
Los Angeles, California, USA

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

Current-steering transimpedance amplifiers for high-resolution digital-to-analogue converters

M.O.J. Hawksford
Centre for Audio Research and Engineering
University of Essex
UK CO4 3SQ

mjh@essex.ac.uk
http://esewww.essex.ac.uk/research/audio

Abstract- A family of current-steering transimpedance amplifier circuits is presented for use in high-resolution, digital-to-analogue converters. The problems of achieving accurate current-to-voltage conversion are discussed with a specific emphasis on digital audio applications. Comparisons are made with conventional virtual-earth feedback amplifiers and the inherent distortion mechanisms relating to dynamic open-loop gain are discussed. Motivation for this work follows the introduction of DVD-audio carrying linear PCM with a resolution of 24 bit at a sampling rate of 192 kHz.

1 Introduction

This paper investigates the design and performance requirements of the transimpedance amplifier used in association with a current-output, digital-to-analogue converter (DAC) [1]. The principal motivation for this work stems from the extreme resolution requirements determined by the advanced audio specification available in digital versatile disc (DVD) applications [2]. Following a theoretical discussion, two principal circuit topologies are presented, the first based upon wide-band, current steering circuit techniques enhanced by input-stage error correction [3], while the second incorporates dual operational amplifiers with nested differential feedback and an embedded low-pass filter.

The DVD-video specification includes linear pulse-code modulation (LPCM) at 96 kHz sampling with a 24-bit resolution while DVD-audio extends this to a maximum of 192 kHz at 24 bit in its two-channel mode. Although DVD includes alternative audio formats such Dolby AC-3[1] and DTS[2] together with lower specification LPCM options, it is the most demanding parameters that dictate the performance requirements of the converters and associated analogue circuitry. Techniques incorporating oversampling and multi-bit noise shaping DACs have been proposed to achieve the required accuracy, which include methods to randomise DAC errors to decorrelate distortion into a noise residue [4,5]. The performance of R-2R ladder network DACs has also improved where accuracy exceeding 21 bit is now claimed for consumer grade products.

However, although the performance of the digital processing and digital converter circuitry can be exemplary, there are error mechanisms in the analogue circuitry immediately following the DAC which produce non-linear distortion. Most multi-bit DACs are current-output devices and should therefore drive low (ideally zero) input impedance transimpedance amplifiers to perform current-to-voltage conversion (I/V conversion). However, DACs operate at high sampling frequencies, typically ≈ 384 kHz, and produce rapid changes in output current with typically nano-second settling times. Consequently, a transimpedance amplifier requires a rapid yet linear response time with low dynamic modulation of its principal parameters.

Distortion mechanisms are discussed specific to a transimpedance amplifier driven by a rapidly changing input current. The errors are assessed both by linear and non-linear analysis including simulation, where it is shown that differential-phase distortion induced by non-linearity, can be represented approximately as an additional correlated jitter distortion [6]. Solutions to these problems are presented that employ fast acting, current steering circuitry augmented by novel input-stage error correction to both linearize and lower the input impedance, also a 2-stage amplifier is investigated.

---

[1] Dolby AC-3: proprietary multi-channel, lossy perceptual coding algorithm.
[2] Digital Theatre Systems (trade name): proprietary multi-channel, lossy perceptual coding algorithm.

2 Virtual-earth transimpedance amplifier and transient distortion mechanisms

Distortion mechanisms in transimpedance amplifiers can be attributed jointly both to linear and to non-linear aspects of circuit behaviour. In the following Section some global observations are made and critical circuit factors examined.

2-1 Linear distortion in I/V conversion

A common approach to transimpedance amplifier design is to use a single high-gain, wide-bandwidth operational amplifier as illustrated in Figure 2-1.

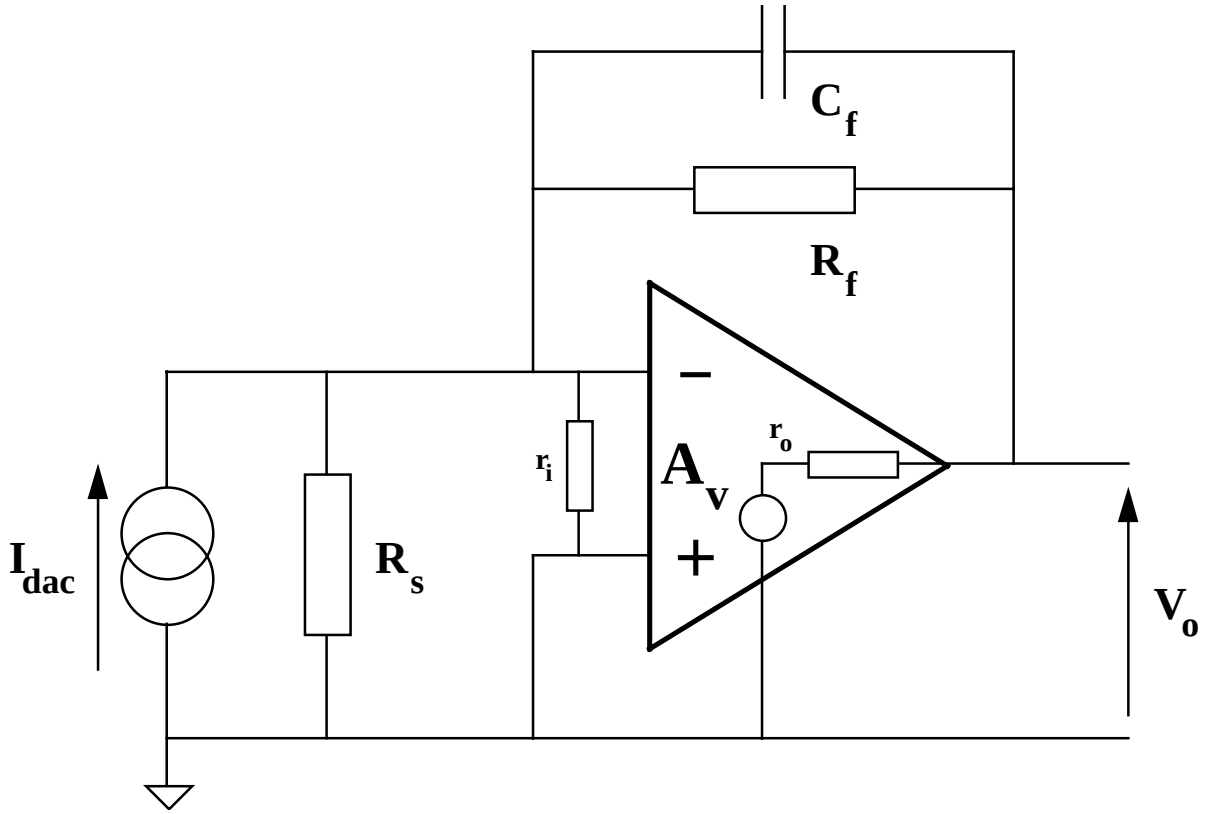


Figure 2-1 Transimpedance amplifier using operational amplifier with feedback.

The DAC output is represented as a Norton equivalent circuit with current generator I_{dac} and source resistance R_s. The operational amplifier is configured in shunt feedback mode with feedback impedance Z_f formed here by resistor R_f in parallel with capacitor C_f. This circuit yields a low value of input impedance z_{in} given by,

$$z_{in} = \left(\frac{R_f}{1 + j2\pi f R_f C_f} \right) \left(\frac{1}{1 + A_v} \right) \; // \; r_i = \frac{Z_f \; // \; r_i}{1 + A_v}$$

For the operational amplifier, A_v is the differential voltage gain and r_i is the differential input impedance that normally can be neglected, reducing z_{in} to,

$$z_{in} \approx \frac{Z_f}{1 + A_v} \qquad \ldots 2\text{-}1$$

For the case where $z_{in} << R_s$, then the current I_{dac} flows predominantly through the feedback impedance and for the ideal case of a vanishingly small differential input voltage, the target transimpedance $Z_T(f)$ is

$$Z_T(f) \quad = \quad \frac{V_o}{I_{dac}} \quad = \quad -Z_f \qquad \ldots 2\text{-}2$$

Equation 2-2 describes perfect I/V conversion and assumes $A_v = \infty$ for all frequency. However, in practical circuits even a good approximation to this criterion is difficult to achieve because of the wide bandwidth of the current signal I_{dac} that results from the rapid changes at sample boundaries. This aspect of performance will be examined in the paper and shown to be of particular significance.

Consider a transimpedance amplifier using an operational amplifier with an n-pole transfer function, where the differential voltage transfer function A_v is given by,

$$A_v \quad = \quad \frac{A_{v0}}{\sum_{r=0}^{r=n-1}\left(1 \; + \; j\frac{f}{f_r}\right)} \qquad \ldots 2\text{-}3$$

where f_0, to f_{n-1} are n respective break frequencies and A_{v0} is the zero-frequency differential voltage gain. If z_{in} is the effective input impedance of the transimpedance amplifier then the differential input voltage v_ε at the virtual earth is,

$$v_\varepsilon \quad = \quad I_{dac}\frac{z_{in}R_s}{z_{in} \; + \; R_s}$$

where by including $Z_T(f)$ from equation 2-2, the output voltage V_o is,

$$V_o \quad = \quad -Z_f\left(I_{dac} \quad - \quad \frac{v_\varepsilon}{r_i}\right) \quad - \quad v_\varepsilon \quad = \quad Z_T(f)I_{dac} \quad - \quad v_\varepsilon\left(1 \; + \; \frac{Z_T(f)}{r_i}\right)$$

In practice the input current to the operational amplifier is negligible as v_ε is small and r_i is large, so can be neglected. Consequently, eliminating v_ε the transimpedance $Z_A(f)$ is,

$$Z_A(f) \quad = \quad \frac{V_o}{I_{dac}} \quad = \quad Z_T(f) \quad + \quad \frac{z_{in}R_s}{z_{in} \; + \; R_s} \quad \approx \quad Z_T(f) \quad + \quad z_{in} \qquad \ldots 2\text{-}4$$

To quantify the error in the transimpedance response, an error function E(f) is defined as,

$$E(f) \quad = \quad 1 \; - \; \frac{Z_A(f)}{Z_T(f)} \qquad \ldots 2\text{-}5$$

Substituting for $Z_A(f)$ from equation 2-4, and incorporating equations 2-1, 2-2 and 2-3,

$$E(f) \quad = \quad -\frac{z_{in}}{Z_T(f)} \quad = \quad \frac{1}{1 \; + \; A_v} \quad = \quad \frac{\frac{1}{A_{v0}}\sum_{r=0}^{r=n-1}\left(1 \; + \; j\frac{f}{f_r}\right)}{1 \; + \; \frac{1}{A_{v0}}\sum_{r=0}^{r=n-1}\left(1 \; + \; j\frac{f}{f_r}\right)} \qquad \ldots 2\text{-}6$$

Equation 2-6 reveals an error function dependent only on the operational amplifier parameters.

Differential-phase distortion $\Delta\phi$

The differential-phase distortion $\Delta\phi$ is defined as the additional phase shift of the transimpedance amplifier introduced by the finite gain and frequency characteristics of the operational amplifier. By considering two I/V stages with respective transimpedances $Z_T(f)$ and $Z_A(f)$, the difference in phase response, hence $\Delta\phi$ is,

$$\Delta\phi \quad = \quad \arctan\left(\frac{Z_T(f)}{Z_A(f)}\right) \quad = \quad \arctan\left(1 - E(f)\right) \quad = \quad \arctan\left(\frac{A_v}{1 + A_v}\right) \qquad \ldots 2\text{-}7a$$

where substituting for A_v from equation 2-3,

$$\Delta\phi \quad = \quad -\arctan\left(1 + \frac{1}{A_{v0}}\sum_{r=0}^{r=n-1}\left(1 + j\frac{f}{f_r}\right)\right) \qquad \ldots 2\text{-}7b$$

The differential group delay T_{diff} is then calculated as,

$$T_{diff} \quad = \quad -\frac{1}{2\pi}\frac{\partial\Delta\phi}{\partial f} \qquad \ldots 2\text{-}8$$

2-2 Non-linear distortion in I/V conversion

The analysis presented in Section 2-1 demonstrates a benign linear distortion that is well controlled in the audio band by feedback providing there is adequate closed-loop gain. Also, the high output impedance of the DAC makes the feedback factor almost unity gain and nearly independent of the feedback path components R_f and C_f. However, for transient input currents that occur at sample boundaries, there can be modulation of the open-loop parameters of the operational amplifier. At these time instants the operational amplifier may appear almost "open-loop" and experience momentary dynamic changes in gain-frequency response. In extreme conditions the operational amplifier can exceed its slew-rate limit contributing further to transimpedance non-linearity, which in turn is reflected in the input impedance.

Non-linearity is modelled here for two cases: The first where no slew rate limiting occurs and there is only minor modulation of the operational amplifier gain as a function of its differential input signal and the second, where momentary slew-rate limiting also occurs.

Consider a change in DAC output current from I(n-1) to I(n) at the n^{th} sample where the sampling interval is τ. Assume the operational amplifier has positive and negative slew-rate limits of $\langle S_+, S_- \rangle$ volt/s and that the DAC output current is a step function. Figure 2-2 shows the output voltage waveform of the transimpedance amplifier, where the response may be divided into two regions. The first region is where the rate-of-change of the output voltage exceeds the slew-rate limit and has constant slope while in the second region, the waveform is controlled by the operational amplifier and its associated feedback network and the output approximates to an exponential waveform. The occurrence of slew-rate limiting causes an error in the area under the reconstructed sample compared to that of linear case.

2-2-1 Linear case

The area under reconstructed sample n for the linear case is defined as A_l. At sample n, the initial output voltage of the transimpedance amplifier is V(n-1), while at sample n+1, due to the finite response time of the amplifier, the output voltage attains a value V(n). Assuming an exponential linear response, the instantaneous waveform $v(n\tau+t)$, for $0 < t < \tau$, is given by,

$$v(n\tau + t) \quad = \quad V(n) \quad + \quad \{V(n-1) \quad - \quad V(n)\} \quad e^{-2\pi f_0 t}$$

where the maximum (initial) slope $= 2\pi f_0\{V(n) - V(n-1)\}$.

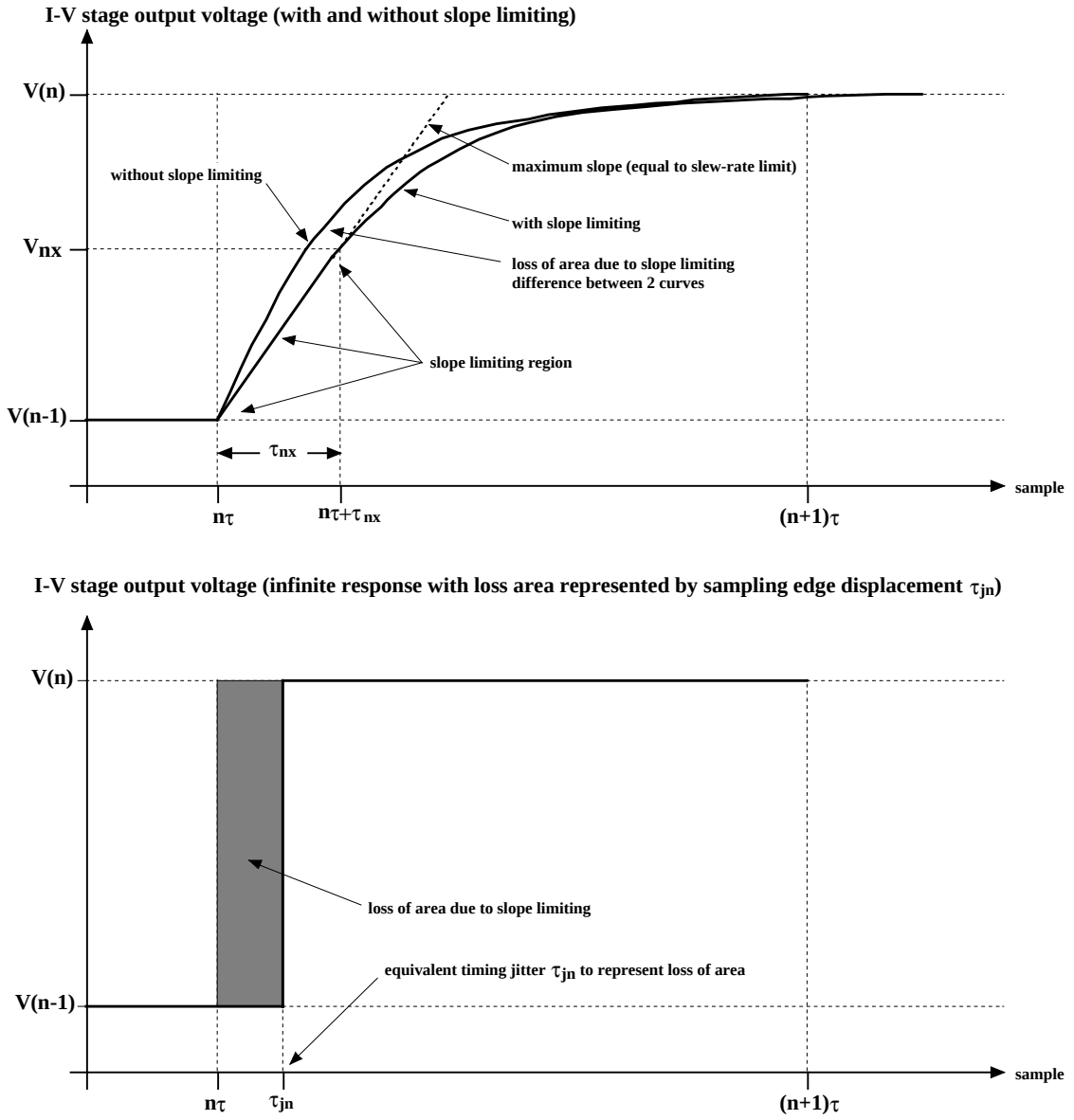


Figure 2-2 Jitter equivalence of slew-induced distortion at a sample boundary.

It is assumed here that the dominant pole in the closed-loop gain produces an exponential waveform between samples with a time constant $\tau_0 = (2\pi f_0)^{-1}$ that is sufficiently small for the exponential transient to have decayed within a sample period τ, that is $e^{-\tau/\tau_0} << 1$.

The area A_l under the n^{th} sample is then calculated,

$$A_l = \int_{t=n\tau}^{(n+1)\tau} v(t)dt$$

giving,

$$A_l = V(n)\tau - \{V(n) - V(n-1)\}\left(1 - e^{-\tau/\tau_0}\right)\tau_0 \qquad \ldots 2\text{-}9a$$

which approximates to,

$$A_l \approx V(n)\tau - \{V(n) - V(n-1)\}\tau_0 \qquad \ldots 2\text{-}9b$$

2-2-2 Mildly non-linear case without slew-rate limiting

Because a single low-frequency pole normally dominates the operational amplifier response, then within the linear operating region the differential input signal of the amplifier is proportional approximately to the time differential of the input signal. If the operational amplifier exhibits mild non-linearity, then there will be waveform distortion that relates to the inter-sample difference signal. It is therefore proposed to model the non-linearity by modulating the time constant τ_0 by a function of the inter-sample difference signal. That is,

$$\tau_0 \quad \Rightarrow \quad \tau_{0n} \quad = \quad (1 + \gamma_n)\tau_0$$

where,

$$\gamma_n \quad = \quad \sum_{r=1}^{k} \lambda_r \left(\frac{(V(n) \quad - \quad V(n-1))}{V_{\max} \quad - \quad V_{\min}} \right)^r$$

$\{\lambda_r\}$ are coefficients defining the non-linearity and equation 2-9b is re-written,

$$A_l \quad \approx \quad V(n)\tau \quad - \quad (V(n) \quad - \quad V(n-1))\tau_{0n} \qquad \ldots 2\text{-}10$$

2-2-3 Non-linear case with slew-rate limiting

In the non-linear case, assume the period τ is divided into two segments, τ_{nx} a period dominated by slew-rate limiting and $\tau - \tau_{nx}$ a linear period with an exponential response with the same time constant as the linear case (although mild non-linearity is introduced later, as in Section 2-2-2). The boundary between the two segments at a level V_{nx} is defined where the initial slope of the exponential at $t = \tau_{nx}$ equals either the positive or negative slew-rate limits of S_+ volt/s or S_- volt/s respectively. If the initial slope, as determined in the linear analysis, of the reconstructed signal case breaches either of these limits then slew-rate limiting occurs, i.e.

$$V(n) \; - \; V(n-1) \quad > \quad S_+\tau_0$$

or

$$V(n) \; - \; V(n-1) \quad < \quad S_-\tau_0$$

In the following analysis, the notation $\langle S_+, S_- \rangle$ implies the slew-rate limit is selected to match the appropriate positive or a negative signal encounter.

In the linear region $\tau > t > \tau_{nx}$,

$$v(n\tau + t) \quad = \quad V(n) \quad + \quad \{V_{nx} \quad - \quad V(n)\} \quad e^{-(t-\tau_{nx})/\tau_0}$$

At the non-linear/linear transition where $v(n\tau + \tau_{nx}) = V_{nx}$ set $\langle S_+, S_- \rangle = \partial v(n\tau + t)/\partial t$, whereby

$$V_{nx} \quad = \quad V(n) \quad - \quad \langle S_+, S_- \rangle \tau_0$$

At the termination of the slew-rate-limited region, V_{nx} is,

$$V_{nx} \quad = \quad V(n-1) \quad + \quad \langle S_+, S_- \rangle \tau_{nx} \qquad \ldots 2\text{-}11$$

Eliminating V_{nx}, then if the slew-rate limit is exceeded,

$$\tau_{nx} \quad = \quad \frac{V(n) \quad - \quad V(n-1)}{\langle S_+, S_- \rangle} \quad - \quad \tau_0 \qquad \ldots 2\text{-}12$$

otherwise,

$$\tau_{nx} = 0 \text{ and } V_{nx} = V(n-1) \qquad \ldots 2\text{-}13$$

By integration, the area A_{ln} under the reconstructed sample for the non-linear case is,

$$A_{ln} = V(n)\tau - 0.5\{2V(n) - V(n-1) - V_{nx}\}\tau_{nx} - \{V(n) - V_{nx}\}\tau_0 \qquad \ldots 2\text{-}14$$

Following the earlier analysis, modify the time constant $\tau_0 \Rightarrow \tau_{0n} = (1 + \gamma_n)\tau_0$ to account for mild non-linearity, where in this case

$$\gamma_n = \sum_{r=1}^{k} \lambda_r \left(\frac{(V(n) - V_{nx})}{V_{\max} - V_{\min}} \right)^r$$

Hence, the pulse-area error ΔA_{ln} is calculated by taking the difference between the linear and non-linear reconstructed samples and follows from equations 2-9b and 2-14 as,

$$\Delta A_{ln} = A_l - A_{ln}$$

$$\Delta A_{ln} = 0.5\{2V(n) - V_{nx} - V(n-1)\}\tau_{nx} + \{V(n-1) - V_{nx}\}\tau_0 + \gamma_n\{V(n) - V_{nx}\} \qquad \ldots 2\text{-}15$$

Since the DAC output impedance is large, the feedback network $R_f//C_f$ does not effect the degree of negative feedback although it does influence the rate-of-change of output voltage, hence onset of slew induced distortion. In a practical amplifier, C_f can be used to marginally band-limit the input signal and lower the output voltage slope, although it does not reduce significantly the differential input voltage of the operational amplifier, hence internal distortion associated with the early stages of amplification.

2-2-4 Equivalent jitter distortion

The above analysis demonstrates pulse-area modulation located close to sample boundaries that results from non-linearity in the transimpedance amplifier during rapid changes of signal. This non-linear change of area can be mapped approximately to an equivalent sample jitter [6] (absolute jitter τ_{jn} being linked with the n^{th} sample) where the reconstructed samples are otherwise linear. If the equivalent jitter τ_{jn} is only a small fraction of a sample period τ and the sampling frequency is high (e.g. 8 times Nyquist sampling rate), then the approximate distortion is an error impulse of area ΔA_{jn} located at the n^{th} sample given by,

$$\Delta A_{jn} = \{V(n) - V(n-1)\}\tau_{jn} \qquad \ldots 2\text{-}16$$

Consequently, the distortion resulting from transimpedance amplifier non-linearity can be represented as a uniformly sampled sequence $\{\Delta A_{jn}\}$ by equating $\Delta A_{jn} = \Delta A_{ln}$ or alternatively as an equivalent sample timing jitter sequence τ_{jn}.

3 Example results of operational amplifier based transimpedance stages

This Section presents some example results to demonstrate the typical levels of linear and non-linear distortion inherent in transimpedance amplifiers used with fast switching, current-output DACs.

3-1 Linear distortion

By way of example, consider a transimpedance amplifier based on the topology in Figure 2-1, using a 3-pole operational amplifier, where the principal parameters are,

R_s	=	4 kΩ	R_f	=	2 kΩ	C_f	=	1 nF
f_0	=	100 Hz	f_1	=	20 MHz	f_2	=	50 MHz

with dc gain A_{v0} swept from 50000 to 200000 in steps of 10000. The results shown in Figure 3-1 correspond to the transimpedance and error function responses defined in Section 2-1, while those shown in Figure 3-2 and 3-3 correspond to $\Delta\phi$ and T_{diff} respectively as defined by Equations 2-7, 2-8.

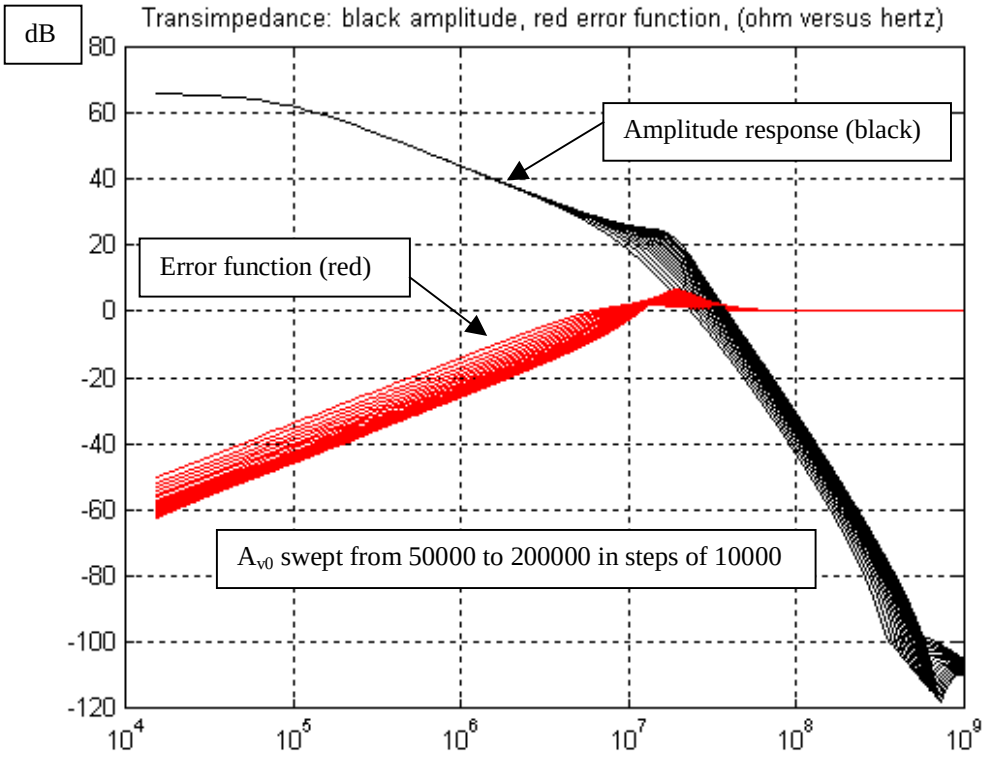


Figure 3-1 Transimpedance and error function as a function of frequency.

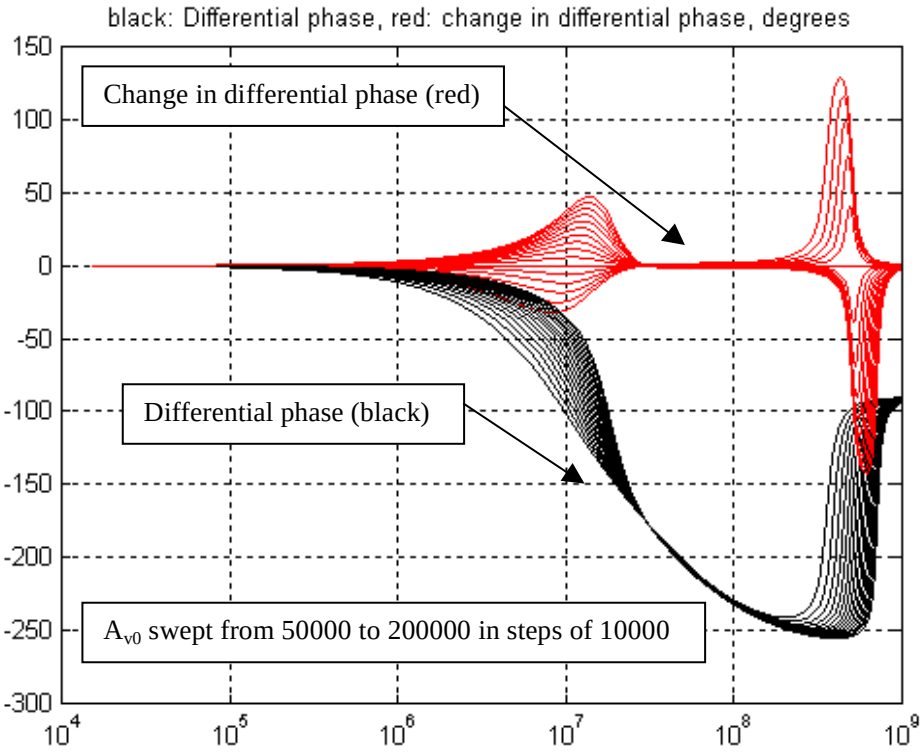


Figure 3-2 Differential phase error $\Delta\phi$ as a function of frequency for varying A_{v0}.

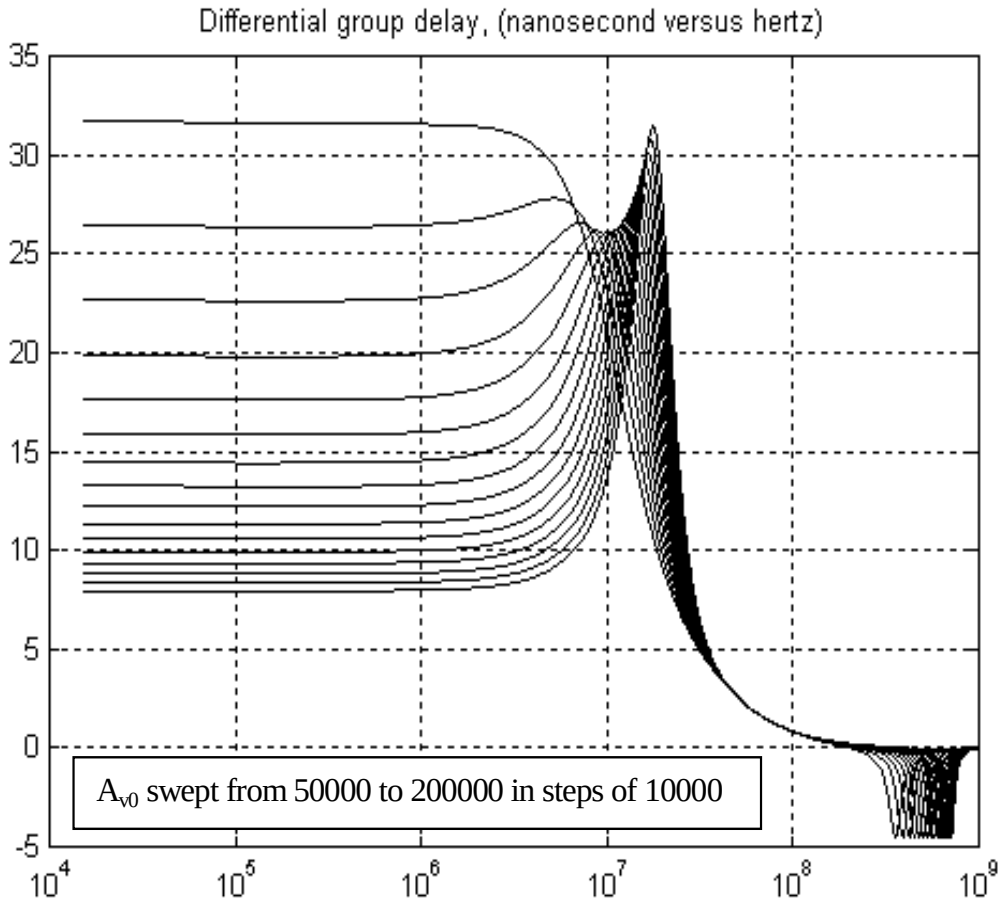


Figure 3-3 Differential group delay T_{diff} as a function of frequency for varying A_{v0}.

Although the analysis presented in Section 2-2-1 is linear, the results illustrate the dependence on A_{v0}, a parameter that may change dynamically both with signal and power supply voltage. Observing $\Delta\phi$ the gross changes appear concentrated at high frequency, yet on closer inspection, T_{diff} reveals that at lower frequency there is an almost constant differential time delay that is strongly dependent on A_{v0}.

To explore this dependency on A_{v0}, a plot of T_{diff} against $(A_{vo})^{-1}$ is presented in Figure 3-4 where the characteristic is almost linear with a slope of $1.5882*10^6$ ns-gain, such that

$$T_{diff} = \frac{1.5882*10^6}{A_{v0}} \qquad \ldots 3\text{-}1$$

Hence, for a change in dc gain ΔA_{v0}, the corresponding change in group delay ΔT_{diff} is,

$$\Delta T_{diff} = -\frac{1.5882*10^6}{A_{v0}}\left(\frac{\Delta A_{v0}}{A_{vo}}\right) \quad \text{nano-second} \qquad \ldots 3\text{-}2$$

This implies that for a nominal gain of $A_{v0} = 10^5$, there is a group delay change of 158.82 ps per 1% gain change. It should be noted that because the output resistance of the DAC is significantly greater than the impedance of the feedback network, the values of R_f and C_f are uncritical with respect to the dependence of T_{diff} on A_{v0}.

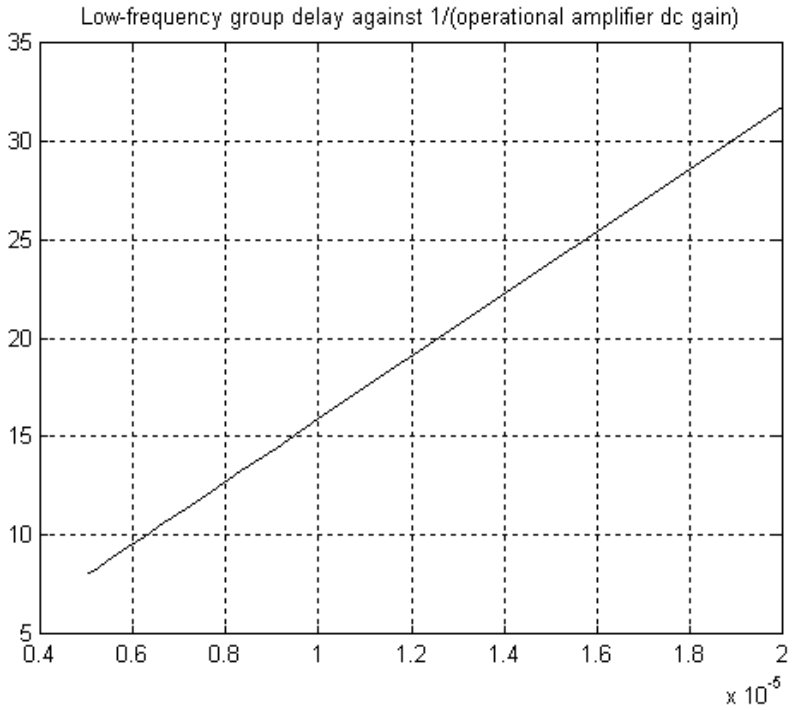


Figure 3-4 Low-frequency differential group delay T_{diff} ns as a function $1/A_{v0}$.

Jitter equivalence

Although these results do not describe non-linear performance in a way that enables exact prediction of distortion, they do give insight into basic mechanisms. For example, the dependency of T_{diff} on A_{v0} can be observed as correlated jitter [6]. Any signal dependent modulation of A_{v0} will cause timing displacement of the signal. This is exacerbated by the presence of high-frequency signal components arising from the structure of sampled audio. In practice there will be modulation of the sampled signal with the dynamic phase-dependent amplifier parameters, allowing high-frequency signal components to alias into the audio band, where examples are presented in Sections 3-2 and 3-3. In making this observation, the role of the feedback capacitor should be observed, which acts to partially bandlimit the input signal as well as lower slew-rate dependent distortion, even though the feedback factor remains close to unity of a broad frequency range.

3-2 Mild amplifier non-linearity

In Section 2-2-2 a transimpedance stage with mild non-linearity was analysed while operating with a sampled data time-domain waveform. Simulation results presented here are performed with the following characteristics selected to prevent the onset of slew-rate limiting even at the lowest sampling rate of 48 kHz:

Positive and negative slew rates:	S_+ = 500 V/μs S_- = -500 V/μs
Transimpedance amplifier first break frequency:	f_0 = 100 Hz
Signal resolution:	24 bit
Non-linearity parameters of operational amplifier:	λ_1 = 0.01 λ_2 = 0.001 λ_3 = 0.0001

Input consists of two sinusoidal currents each of amplitude $\sqrt{2}$ mA and respective frequencies 19 kHz and 20 kHz, where the low-frequency transimpedance is 1 kΩ, where the assumed peak-to-peak current output range of the DAC is $-2\sqrt{2}$ to $2\sqrt{2}$ mA. Output spectra are shown in Figures 3-5a, 3-6a and 3-7a respectively for sampling rates of 48 kHz, 192 kHz and 384 kHz together with corresponding equivalent time-domain jitter waveforms shown in Figures 3-5b, 3-6b and 3-7b. Two sine wave reference signals are also superimposed on the spectra to benchmark the 24-bit dynamic range, one at set at the full amplitude of $2\sqrt{2}$ mA peak, while the other is reduced in level by 2^{24} (i.e 144 dB).

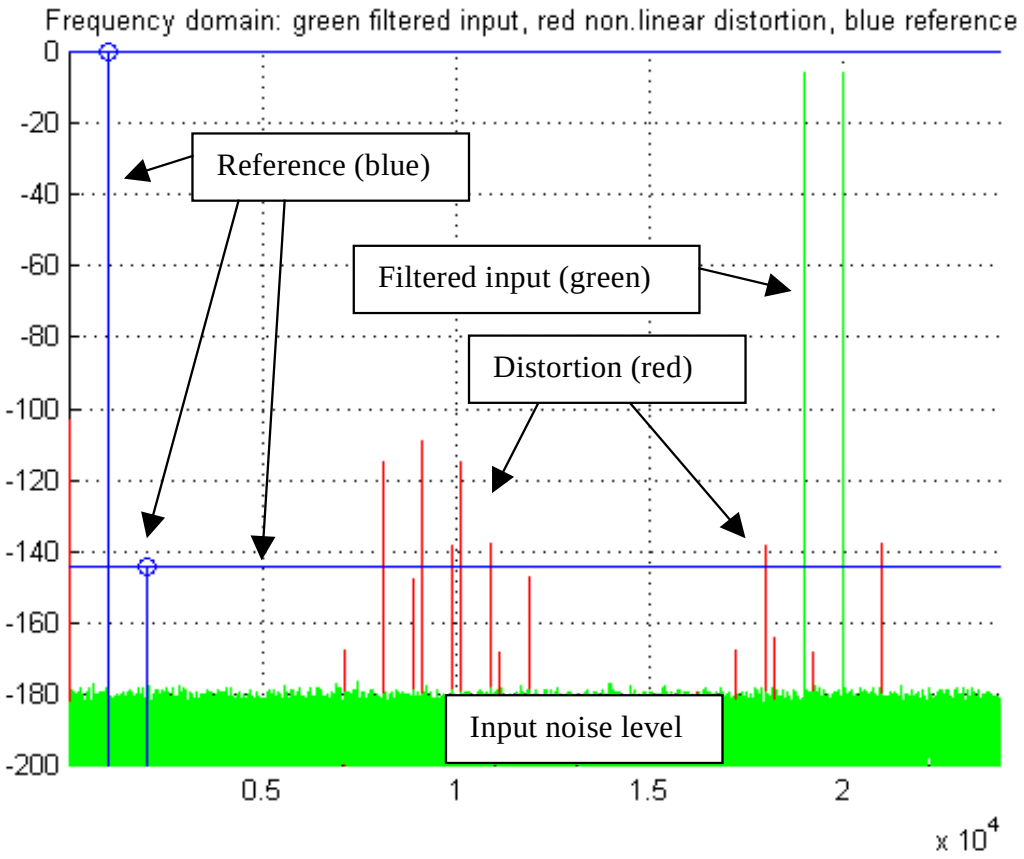


Figure 3-5a Output spectrum, sampling rate 48 kHz

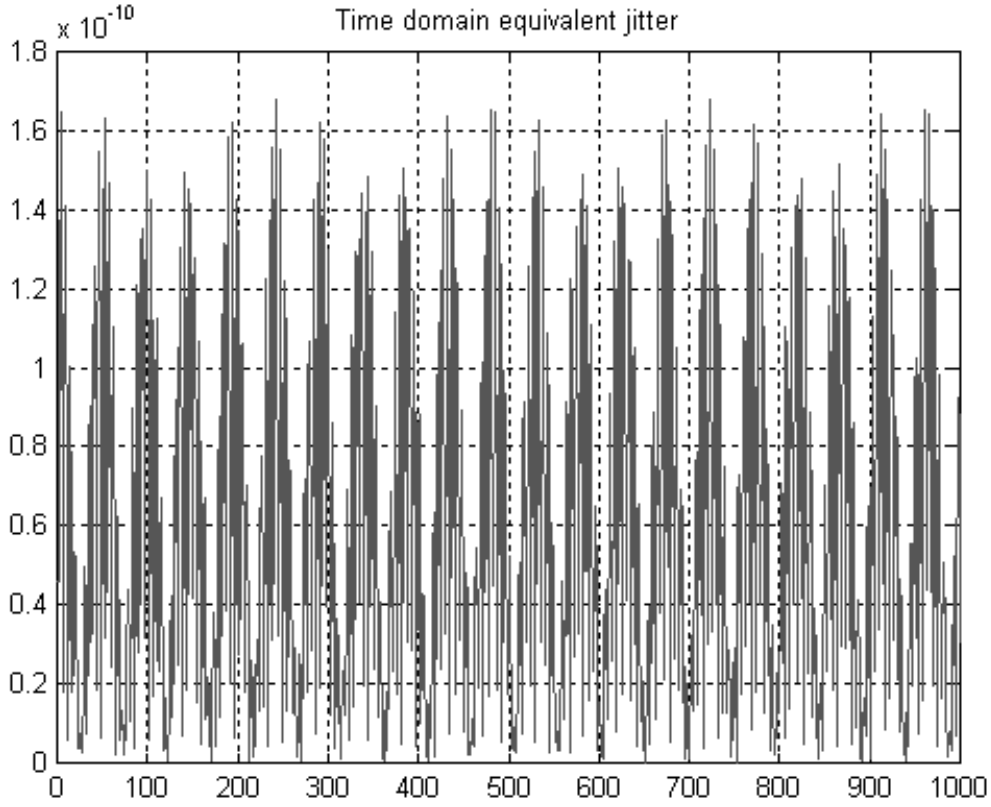


Figure 3-5b Equivalent sampling jitter, sampling rate 48 kHz.

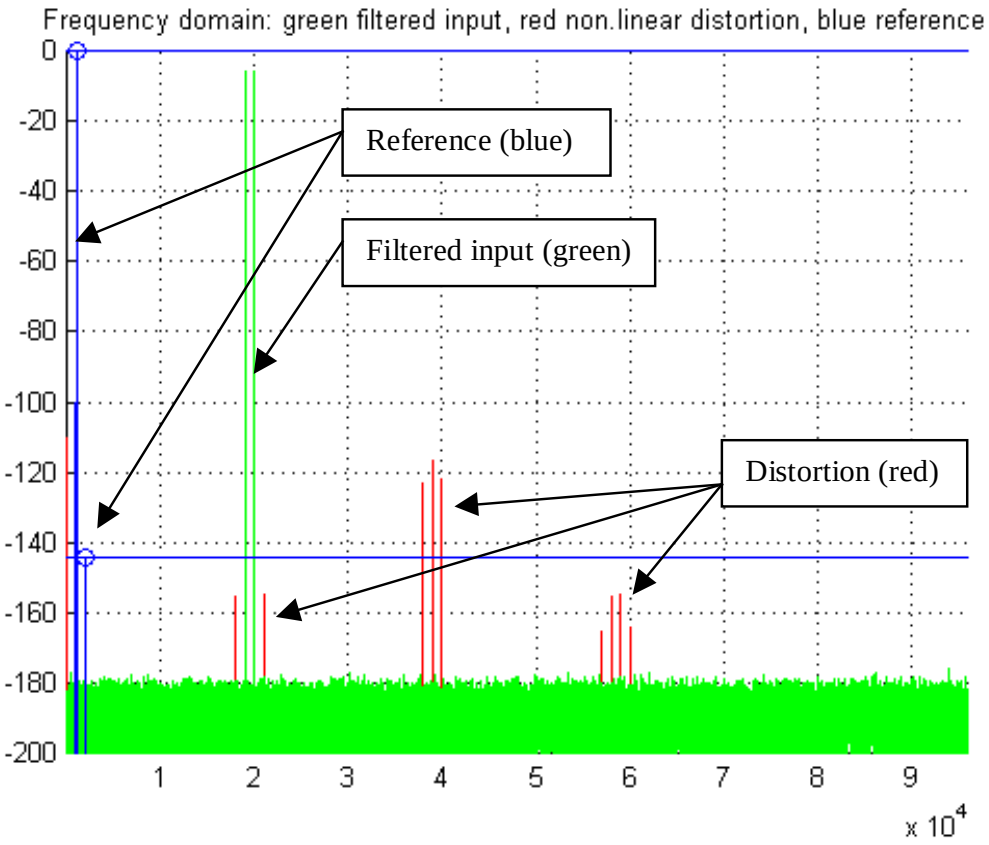


Figure 3-6a Output spectrum, sampling rate 192 kHz.

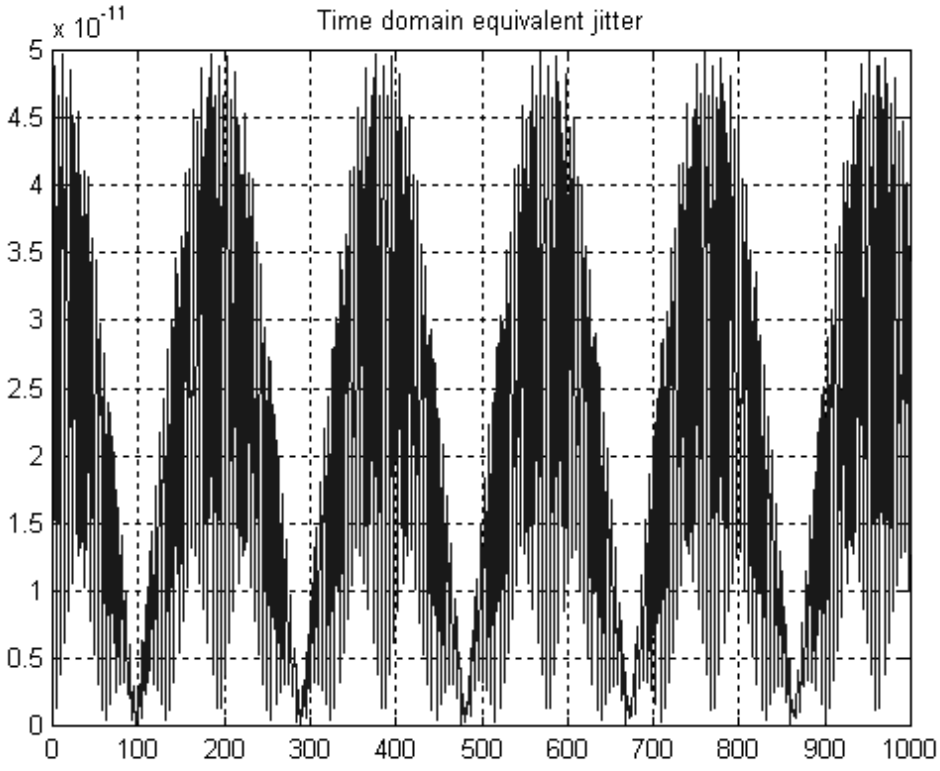


Figure 3-6b Equivalent sampling jitter, sampling rate 192 kHz.

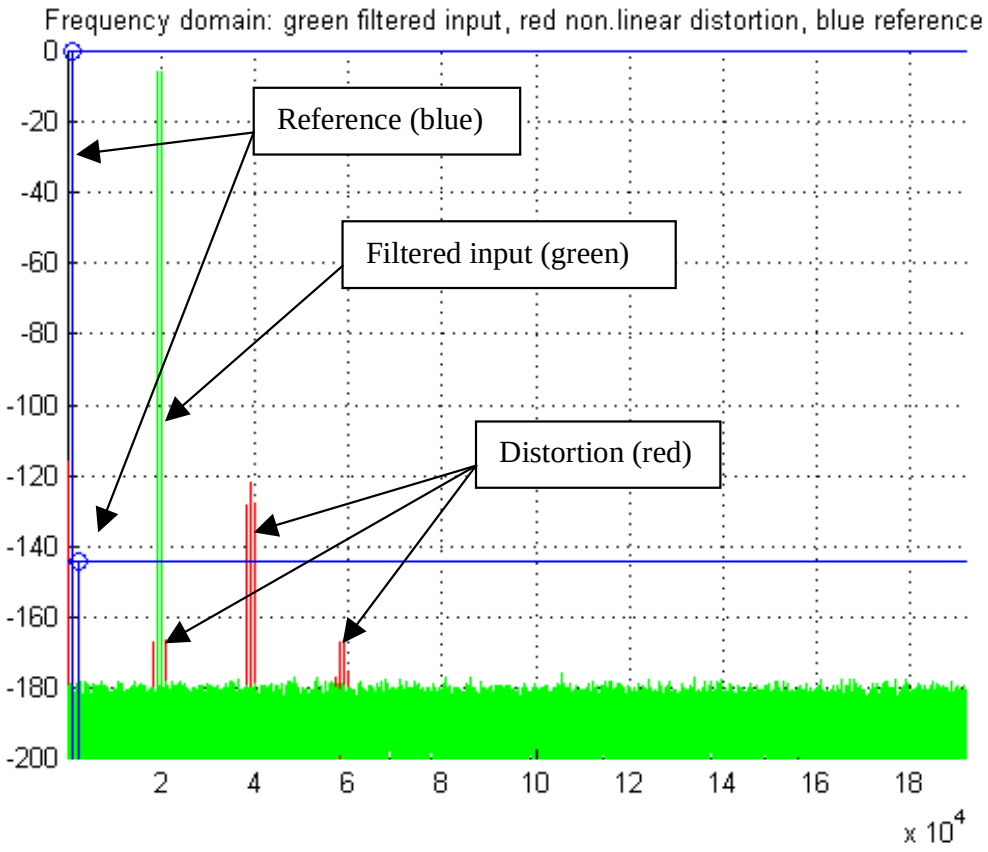


Figure 3-7a Output spectrum, sampling rate 384 kHz

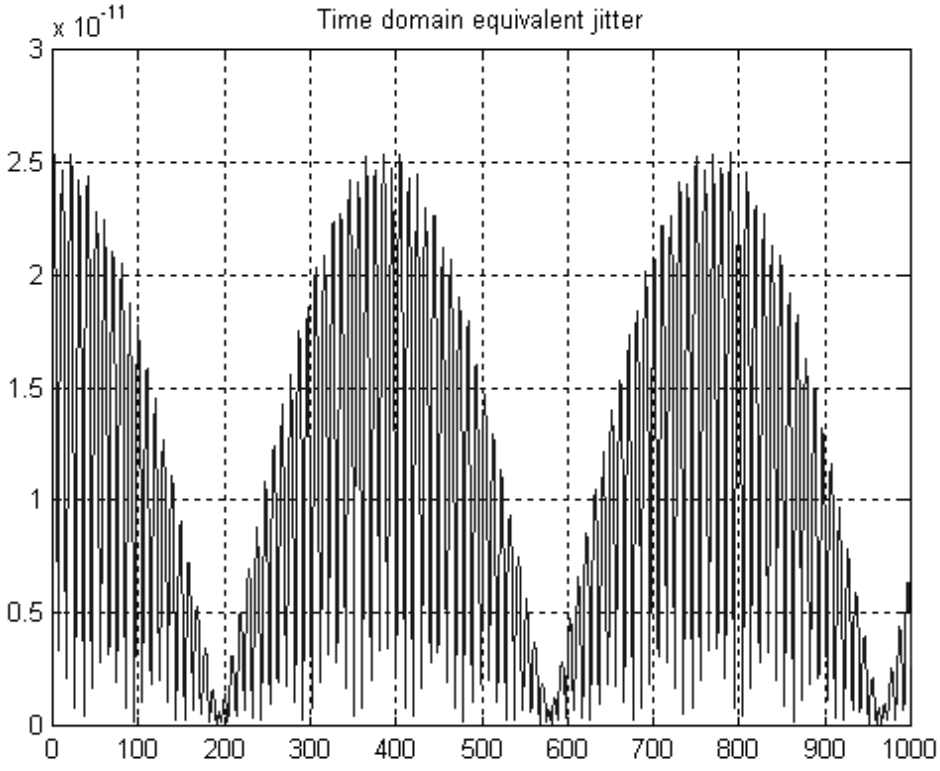


Figure 3-7b Equivalent sampling jitter, sampling rate 384 kHz.

3-3 Mild amplifier non-linearity with slew-rate limiting

In Section 2-2-3, the analysis was extended to include slew-rate limiting. As to whether slew-rate limiting occurs depends upon the inter-sample difference and the closed-loop bandwidth f_0 of the transimpedance stage. By way of example, the simulations presented in 3-2 are repeated but with the slew-rate limits of the operational amplifier modified to,

Positive and negative slew rates: $S_+ = 50$ V/μs $S_- = -50$ V/μs

Output spectra are shown in Figures 3-8a, 3-9a and 3-10a respectively for sampling rates of 48 kHz, 192 kHz and 384 kHz together with corresponding equivalent time-domain jitter waveforms shown in Figures 3-8b, 3-9b and 3-10b.

3-4 Observations

In the following discussion the objective is to compare distortion levels against a system aspiring to 24-bit resolution. As the sampling rate is lowered, inter-sample differences increase so increasing the differential drive to the transimpedance stage, hence higher distortion is anticipated. However, it is evident that the greatest distortion arises because of slew-rate limiting, even if this is only a momentary event at the commencement of each sample, so it is imperative to design a system such that amplifiers operates well clear of slope overload. Although the slew rate in the analysis was referred to the output voltage, it is conceivable that other slope related distortions could occur in the operational amplifier. This was partially accounted by the inclusion of a mild non-linearity operating on the inter-sample difference signal.

The use of equivalent jitter was used to demonstrate how distortion calculated on a sample-by-sample basis compares with conventional timing jitter, as notional benchmarks have been suggested as to permissible levels of jitter. For example, critical listening tests have been used to evaluate the effect of sampling rate on audible performance. Interestingly, results suggest that the level of jitter must be held to an extremely low level for valid results to be obtained. Of course this is an oversimplification, as the spectral content of jitter and its correlation with the signal are critical and the interactions can be extremely complicated. It is evident that quite mild levels of non-linearity in the open-loop behaviour of an operational amplifier can map through to equivalent jitter figures that are significant. The linear analysis presented in Section 3-1 is illuminating with regard to this phenomena, where dynamic modulation of the parameters such as dc gain and/or the dominant-pole frequency, will map effectively into timing errors. It is important to note that parametric modulation is exacerbated by the presence of rapid signal changes at the sample boundaries, where one can envisage a transient modulation of pulse timing, making the concept of jitter equivalence more tractable. However, if the signal is filtered to remove the sample structure this should reduce the level of modulation, where this appears to be born out by the process of pre-filtering of the DAC output current prior to I/V conversion.

The inclusion of capacitor C_f in the feedback path reduces the output slew-rate. Consequently, in this sense is helpful but because the DAC output impedance is high, the capacitor has little effect on the level of feedback which is already close to maximum, so will not influence the output distortion other than by introducing high-frequency attenuation of the input. Inter-modulation and timing modulation remain. Observe how in the analysis in Section 2-2-1, group delay changes with operational amplifier dc gain A_{v0} occurred even when capacitor C_f was included in the feedback loop. Although C_f has no direct effect on timing performance, it does affect distortion resulting from rapid output voltage changes, which in turn then modulates the amplifier parameters, hence dynamically altering in-band group delay.

To mitigate the problem of timing modulation three strategies are proposed:

- Open-loop/current-feedback, wide-band I/V conversion where the in-audio band, signal delay is low and the amplifier parameters are established with minimal parametric modulation.
- Pre-filter the DAC output current with a passive low-pass filter.
- Multiple amplifier stages with nested feedback to achieve extremely high loop gains with low in-band phase group delay distortion.

Section 4 discusses low-feedback current-steering circuits, while pre-filtering is investigated in Section 5 and a dual-loop I/V stage is presented in Section 6.

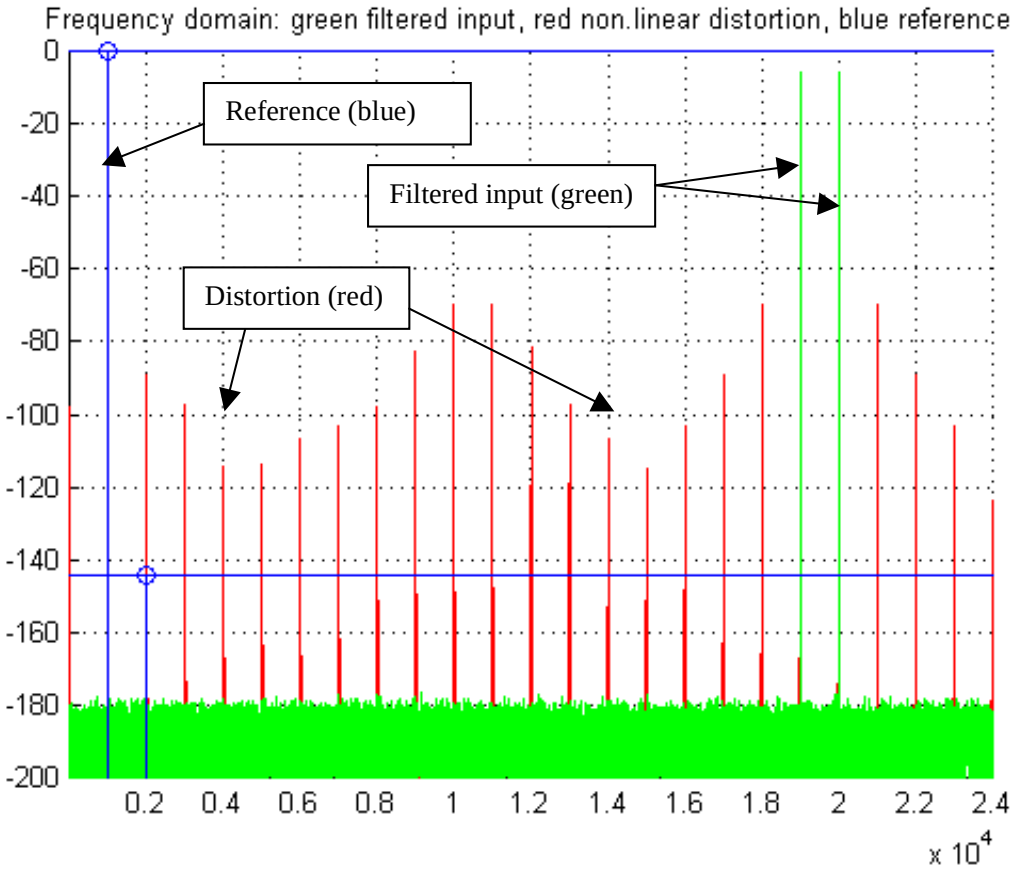


Figure 3-8a Output spectrum, sampling rate 48 kHz

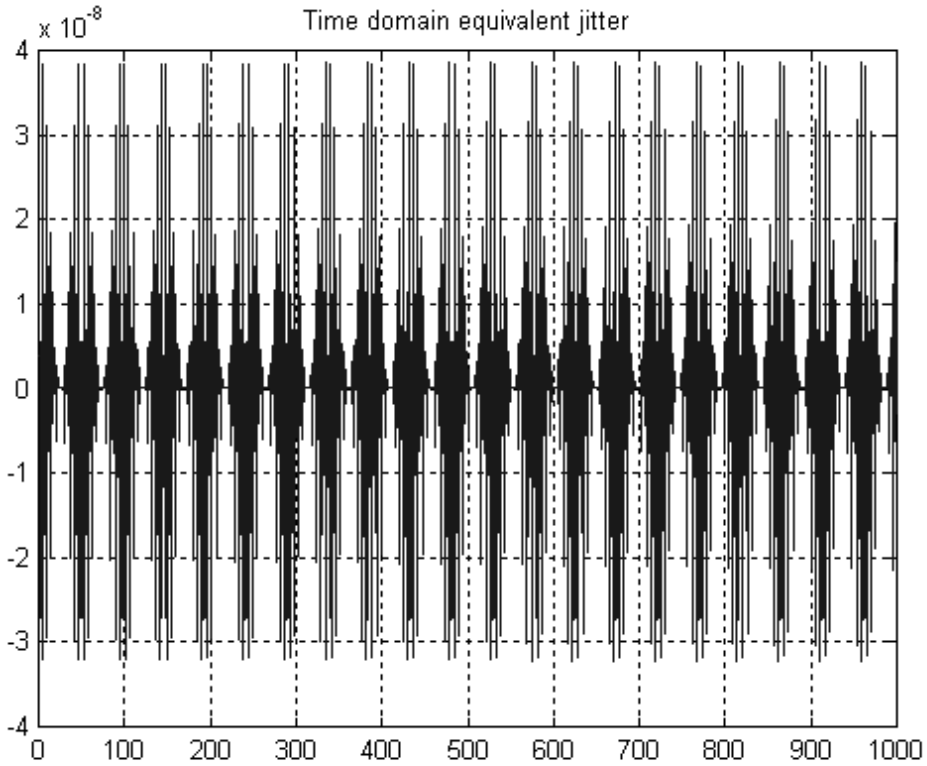


Figure 3-8b Equivalent sampling jitter, sampling rate 48 kHz.

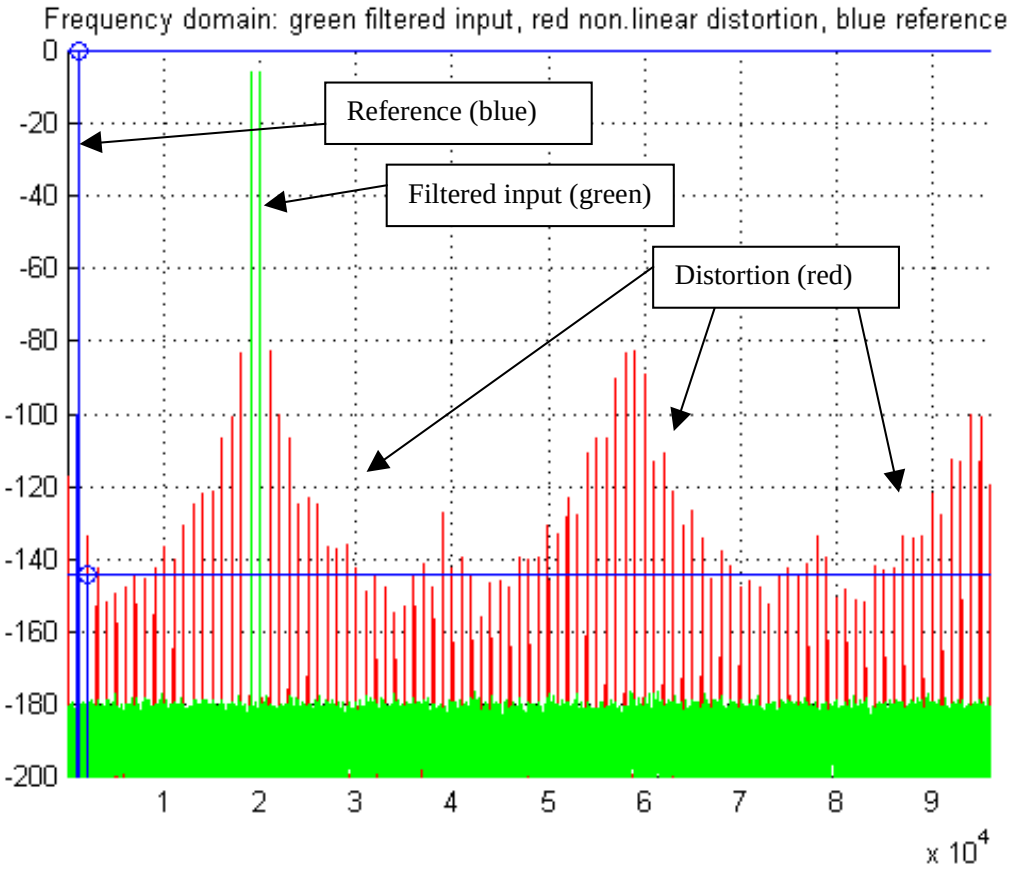


Figure 3-9a Output spectrum, sampling rate 192 kHz

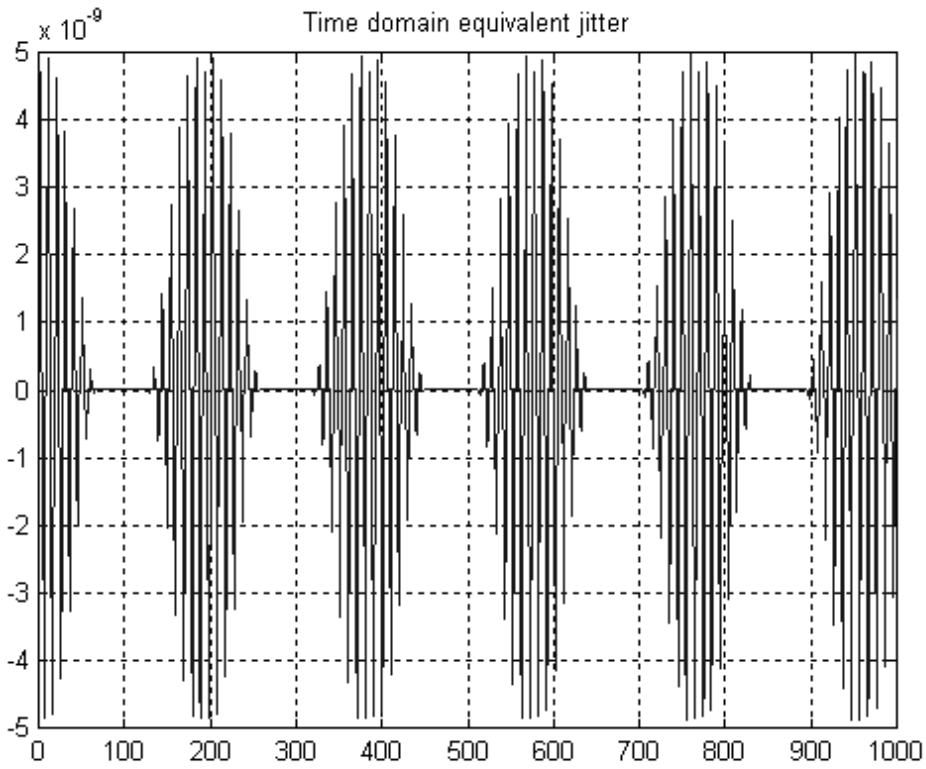


Figure 3-9b Equivalent sampling jitter, sampling rate 192 kHz.

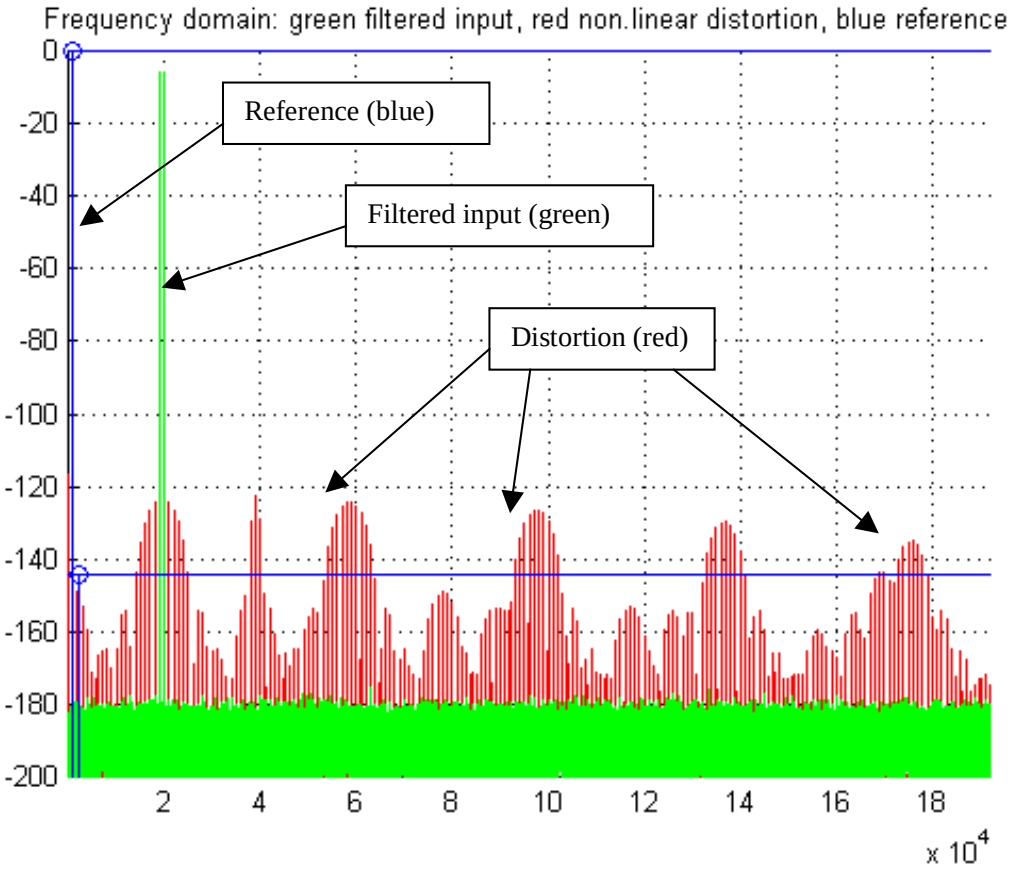


Figure 3-10a Output spectrum, sampling rate 384 kHz

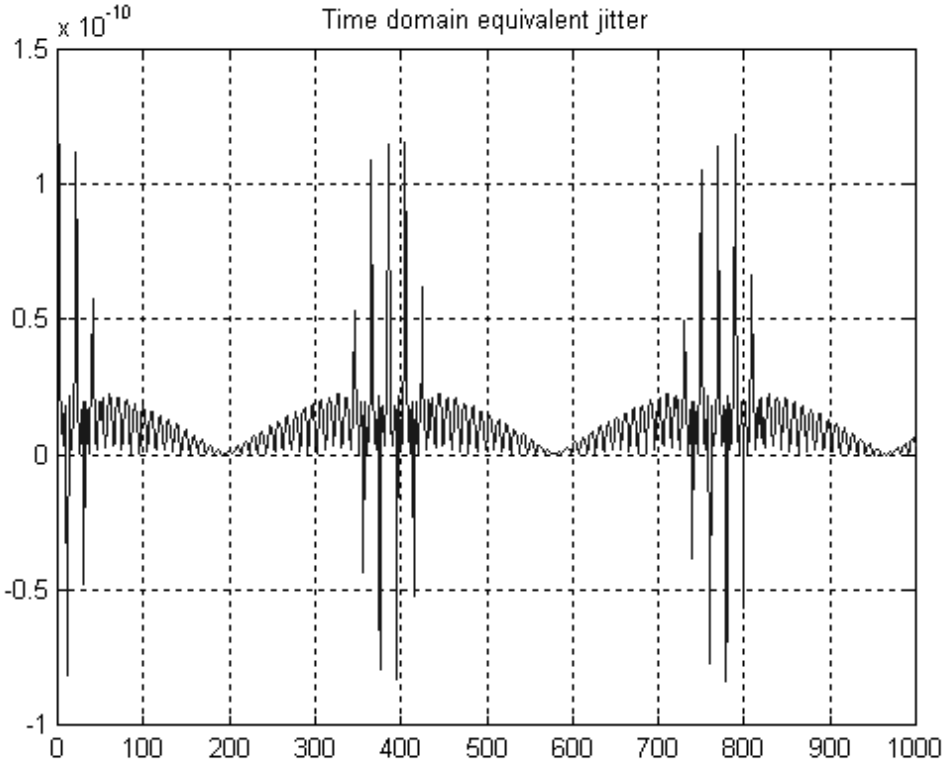


Figure 3-10b Equivalent sampling jitter, sampling rate 384 kHz.

4 Current-steering amplifiers

An alternative approach to the feedback amplifier is to use a current-steering, open loop configuration as shown conceptually and without dc biasing in Figure 4-1. This transimpedance stage uses a grounded-base amplifier that steers the output current of the DAC into the collector load impedance. The load impedance includes a shunt capacitance to partially bandlimit the signal and to reduce the rate of change of output voltage. For the special case of a DAC with infinite output impedance, the only distortion mechanism is the minor modulation in the slope parameters of the transistors. However, a finite DAC output impedance will result in minor distortion due to the modulation of input resistance with signal current. The amplifier is completed by using a unity-gain buffer amplifier and low-pass filter.

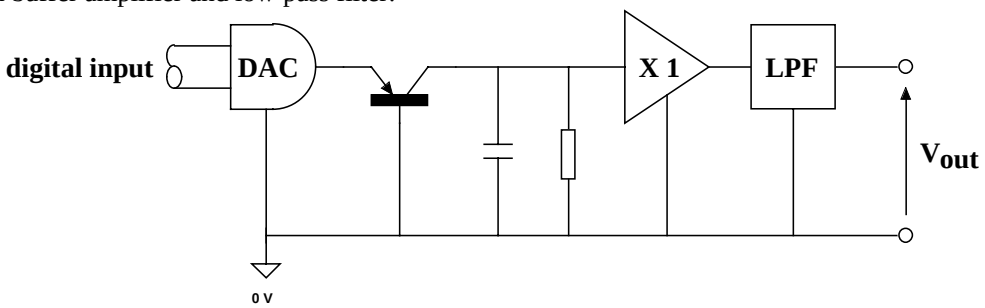


Figure 4-1 Grounded-base open-loop current steering transimpedance stage.

4-1 Input stage error correction

The use of optimally balanced, error feedback can virtually eliminate the non-linear modulation of transistor base-emitter slope resistance. In Figure 4-2 a modified amplifier is shown where the input stage consists of two matched complementary transistors T_1 and T_2 together with a grounded base stageT_3. T_1 and T_3 act as a cascode stage to steer the DAC output current i to the current mirror formed by T_4, T_5 and the three equal valued resistors R_0 such that the collector currents of T_4 and T_5 each carry a mirror the current i. As a result, changes in emitter currents of T_1 and T_2 are identical, where providing parametric and thermal matching, then $V_{BE1} = -V_{BE2}$. Consequently, the emitter potential of T_1 remains theoretically zero even though the base-emitter voltages may change non-linearly with signal current. This implies zero input impedance even under large signal conditions. A constant current generator I_B sinks the collector current bias component of T_5 while a parallel resistor-capacitor network R_1, C_1 converts the DAC signal current to a voltage. A unity-gain buffer with high input impedance completes the conversion yielding an overall transimpedance $Z_{I/V}$ of,

$$Z_{I/V} = \frac{-R_1}{1 + j\omega R_1 C_1} \qquad \ldots 4\text{-}1$$

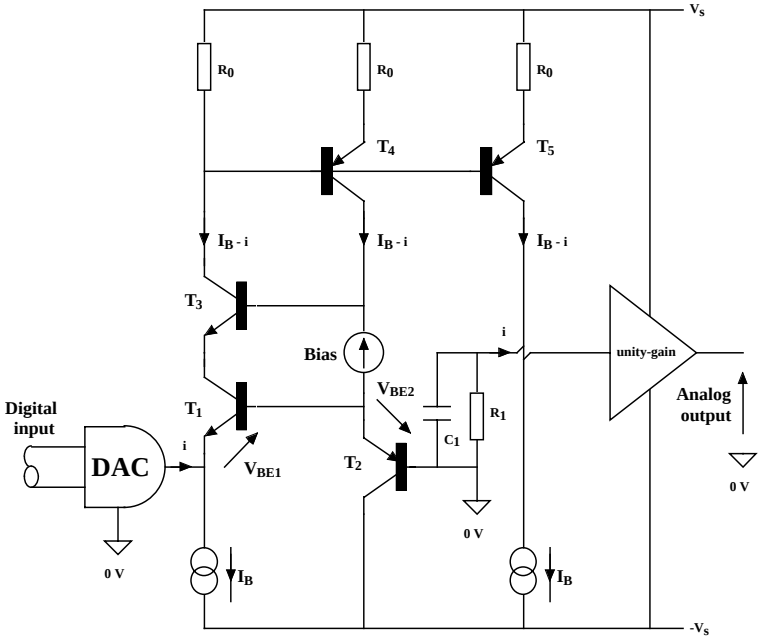


Figure 4-2 Open-loop transimpedance amplifier with input stage error correction.

4-2 Current-feedback transimpedance amplifier, embedded low-pass filter

The performance of the transimpedance amplifier can be improved by current feedback as shown in Figure 4-3.

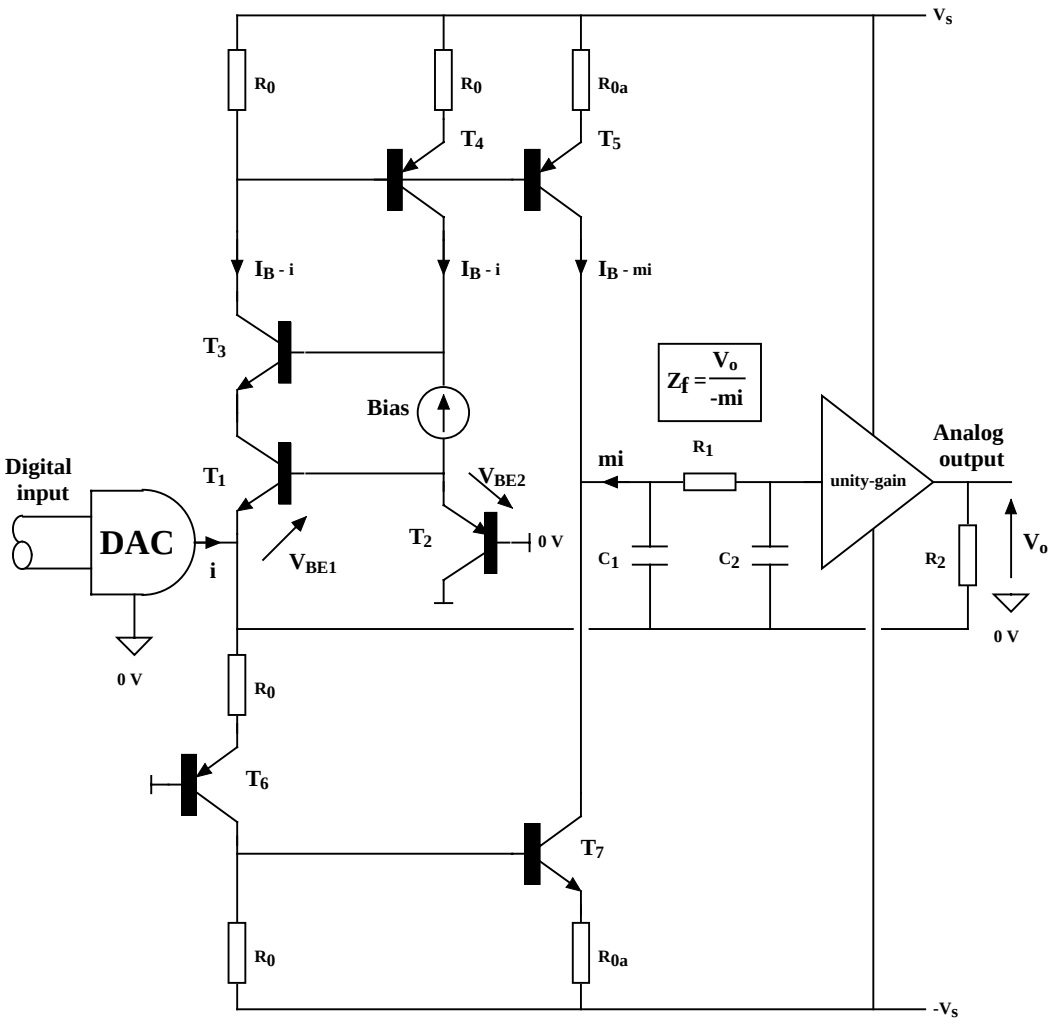


Figure 4-3 Transimpedance amplifier with error correction and embedded low-pass filter.

A principal feature of this circuit is the inclusion of a second-order, low-pass filter embedded in the output stage formed by the π-network R_1, C_1 and C_2, a unity-gain buffer and resistor R_2. However, it will be observed that the filter ground line is returned to the input node rather than the true ground and this constitutes the current-feedback path to the emitter of T_1. The operation is such that at low frequency, feedback is derived via R_2 and thus includes the output buffer, while at higher frequency, in the filter attenuation region, the current path is derived primarily from the collector current of T_6.

In this respect the high-frequency feedback path is similar to a simple dc-coupled feedback pair of transistors. Benefits derived from this configuration include reduced output impedance and enhanced linearity together with an embedded low-order reconstruction filter, where the filter behaves as an integral part of the feedback path while returning no currents to ground to aid ground-rail purity. Although a second-order low-pass filter is shown in Figure 4-3, higher-order filters can be accommodated without incurring a stability penalty. In effect, the low-pass filter acts as a nodal transition filter, allowing the feedback path to be gradually redirected as a function of frequency from the output node to the buffer input. Also, the DAC signal current i is returned to the power supply and does not require transient currents to flow in the ground bus.

Assuming the current gain of the mirror formed by transistors T_4 and T_5 is m and the filter formed by C_1, R_1 and C_2 has a transimpedance Z_f (see Figure 4-3 for component definitions), then the closed-loop transimpedance $Z_{I/V}$ of the overall amplifier is,

$$Z_{I/V} = \frac{-R_2}{1+\left(\frac{1+m}{m}\right)\frac{R_2}{Z_f}} = \frac{-R_2}{1+\alpha\frac{R_2}{Z_f}} \quad \ldots 4\text{-}2$$

where $\alpha = m/(m+1)$. Taking the filter example shown in Figure 4-3, then the transimpedance of this filter stage Z_f is,

$$Z_f = \frac{1}{j\omega(C_1 + C_2) + (j\omega)^2 R_1 C_1 C_2} \quad \ldots 4\text{-}3$$

resulting in,

$$Z_{I/V} = \frac{-R_2}{1 + j\omega\alpha R_2(C_1 + C_2) + (j\omega)^2 \alpha R_1 C_1 R_2 C_2} \quad \ldots 4\text{-}4$$

4-3 Discrete transimpedance amplifier with input-stage error correction

Figure 4-4 illustrates a complete transimpedance amplifier based upon the skeleton topology presented in Section 4-2. This circuit incorporates a second-order, low-pass output filter together with an optional RC network to implement the standard de-emphasis characteristic required for CD replay. A servo amplifier is also included to control the output voltage offset voltage to facilitate dc coupling. The overall transimpedance of this amplifier without de-emphasis but including a first-order servo follows as,

$$Z_{I/V} = \left[\frac{-1100}{1 + j\omega 11*10^{-6} + (j\omega)^2 2.42*10^{-12}}\right]\left[\frac{j\omega*10}{1 + j\omega*10}\right] \quad \ldots 4\text{-}5$$

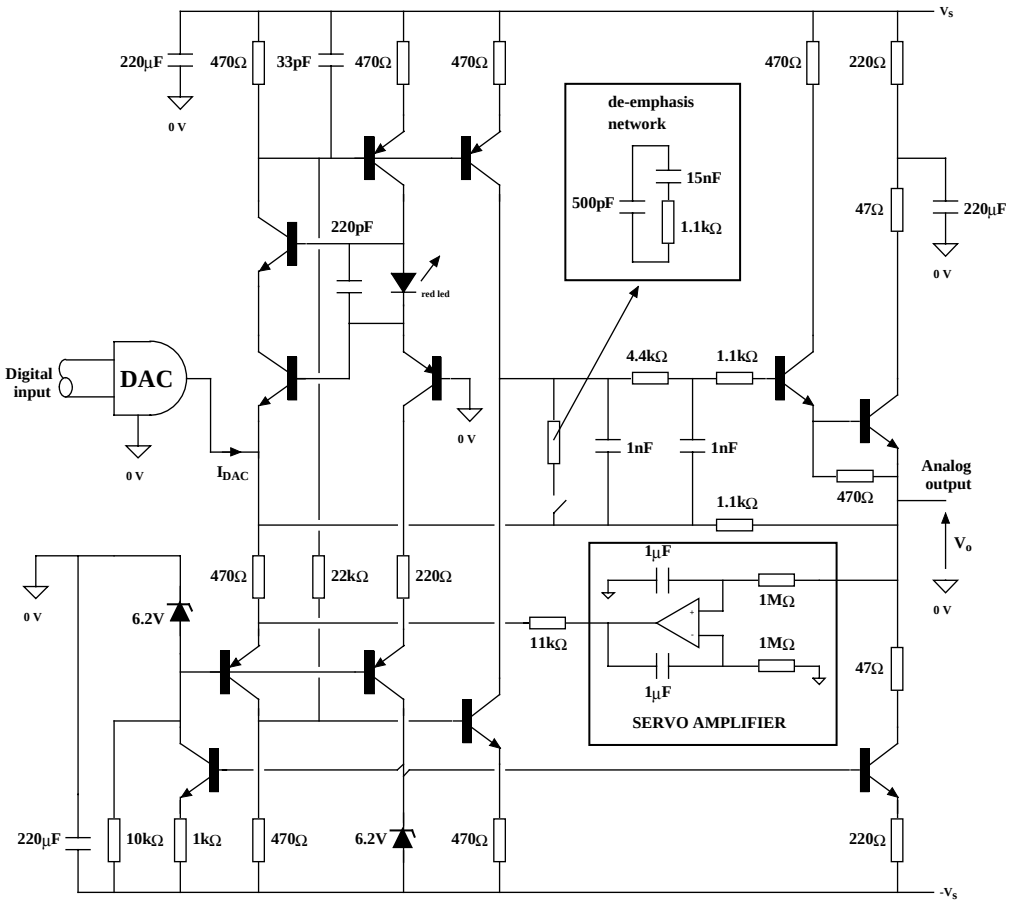


Figure 4-4 I/V stage with error correction, embedded low-pass filter and servo amplifier.

5 DAC output current pre-filter prior to I/V conversion

To reduce the rate-of-change of signal current applied to the transimpedance stage, pre-filtering can be used where an example circuit is shown in Figure 5-1.

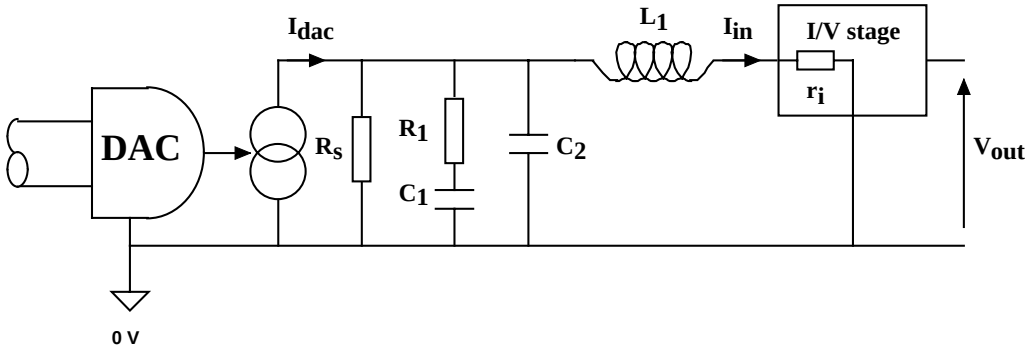


Figure 5-1 Pre-filter for DAC output current using lumped RLC network.

The filter band-limits the input current prior to conversion and militates against distortion resulting from rapid signal changes occurring at sample boundaries, thus lowering distortion correlation with the sampling rate. However, a negative feature of this technique is that it reduces the source impedance seen by the transimpedance stage, which has a detrimental effect on distortion and noise performance. Also, the input impedance of the filter changes with frequency implying that the DAC no longer feeds a near zero impedance. Nevertheless, such additional processing is often effective in association with low-cost operational amplifier transimpedance stages.

6 Dual-loop transimpedance stages

An enhancement to the single operational amplifier I/V stage is shown in Figure 6-1, where three operational amplifiers are used together with nested-differentiation feedback to achieve stability [7,8,9].

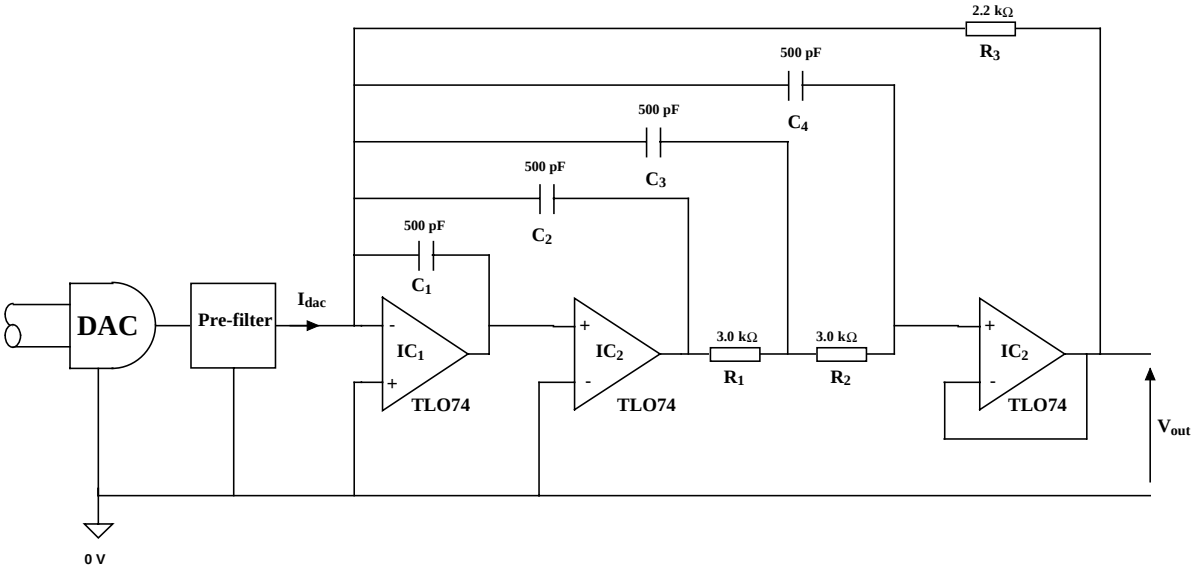


Figure 6-1 Dual-loop operational amplifier transimpedance stage with nested-differentiating feedback configured as a 3rd-order low-pass filter.

Two operational amplifiers are used for amplification, while the third is used as unity-gain output buffer. Nested feedback has a dual function and also forms a third-order, low-pass filter for signal recovery. The principal advantage of dual stages is a large reduction in sensitivity to the operational amplifier characteristics that achieves a corresponding reduction in dynamic phase modulation. Amplitude and phase responses are shown in Figure 6-2 for a design configured for a 50 kHz signal bandwidth, where the input reference current level is 1 mA. Figure 6-2 also shows the output response of the first operational amplifier IC_1, where the signal level is typically better than 70 dB below that of the final output, this helps to reduce transient modulation in IC_1 by constraining its output signal amplitude.

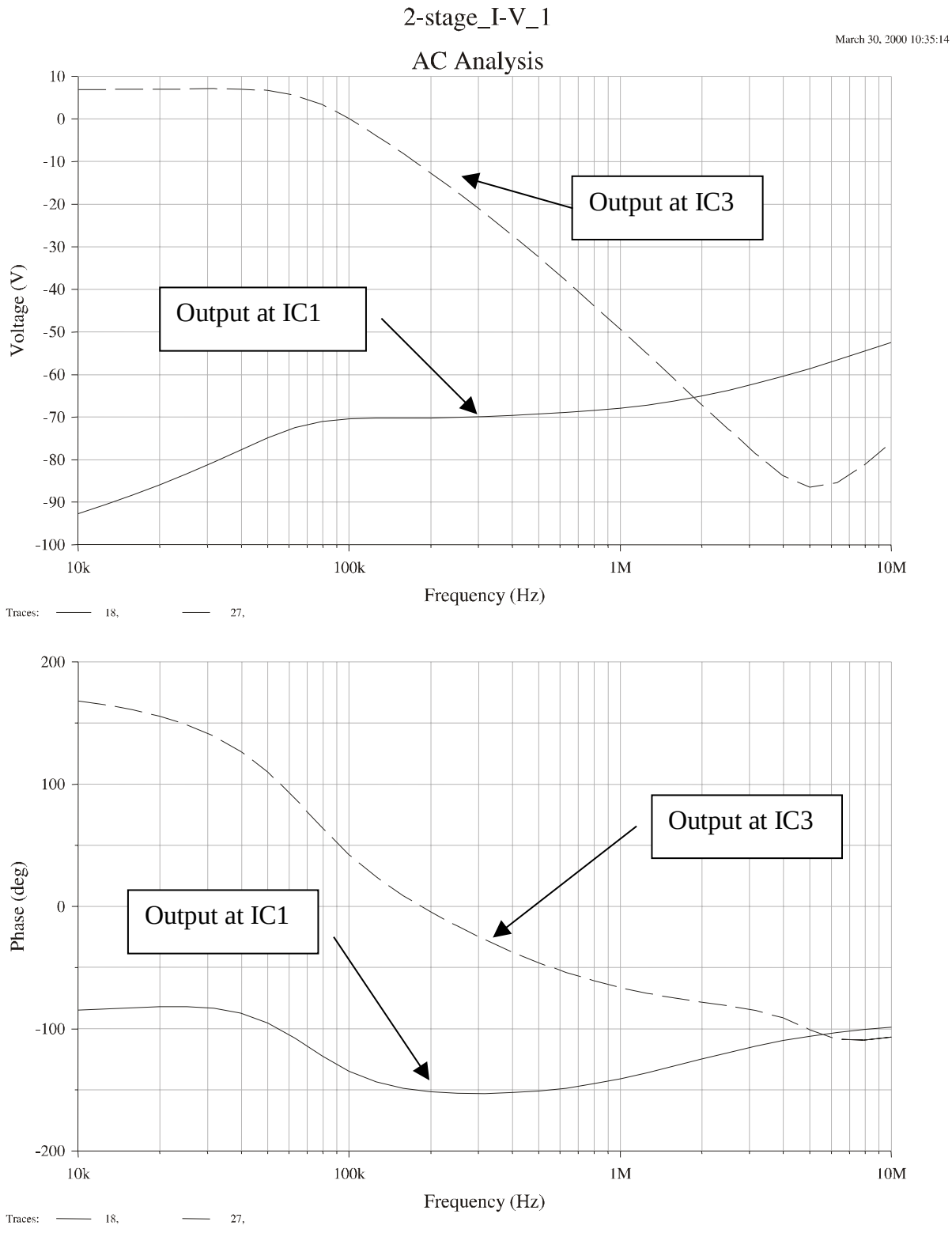


Figure 6-2 Transimpedance amplitude and phase response plots (reference Figure 6-1).

Next, some aspects of overall error of a 2-stage I/V stage are investigated. Figure 6-3 shows a pair of 2-stage transimpedance amplifiers with an error signal derived from the difference of their respective outputs, where this configuration is used as a simulation model. By way of example and to demonstrate the sensitivity of the overall error to operational amplifier gain, the input current to output error voltage response is shown in Figure 6-4 for a first-stage gain error g = 0.9. Because two stages of amplification are used, it follows that the low-frequency error has only a small phase shift of less than 20 degree. This coupled with a low level of error < -175 dB, means that the overall transimpedance has extremely low sensitivity to gain changes. By way of comparison, if the second stage is removed, Figure 6-5 shows the corresponding error level which is now only a little below –97 dB with a corresponding error signal phase shift of about -90 degree.

The observation that for the two-stage amplifier the error signal is almost in-phase is advantageous when considering equivalent jitter-distortion as a modulation in loop gain does not produce the same degree of differential-phase distortion as when the error is approximately in phase quadrature. Of course, the sensitivity being so much lower than for a single stage amplifier should virtually eliminate this problem and make the stage perform much closer to the specification dictated by high-resolution audio.

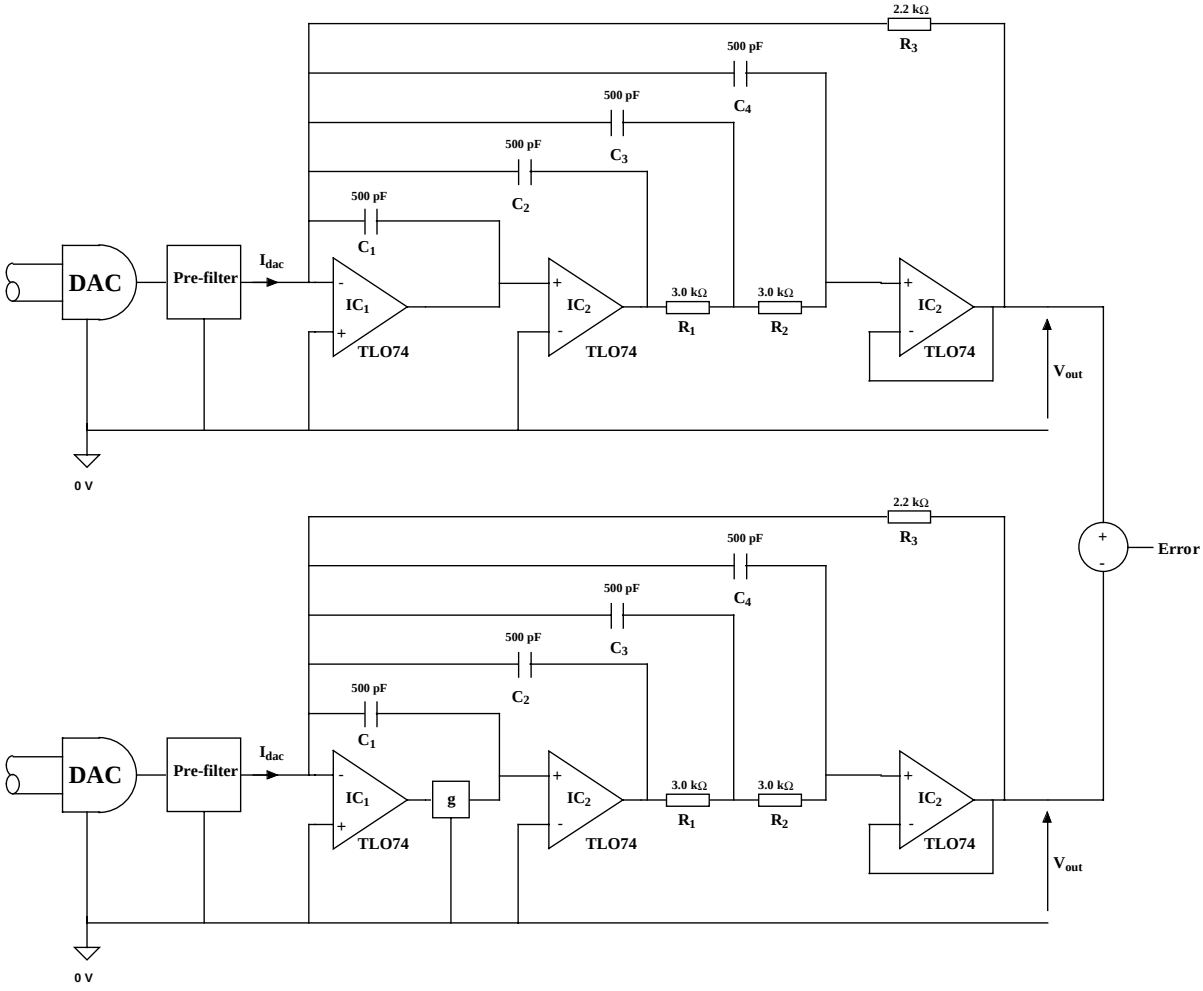
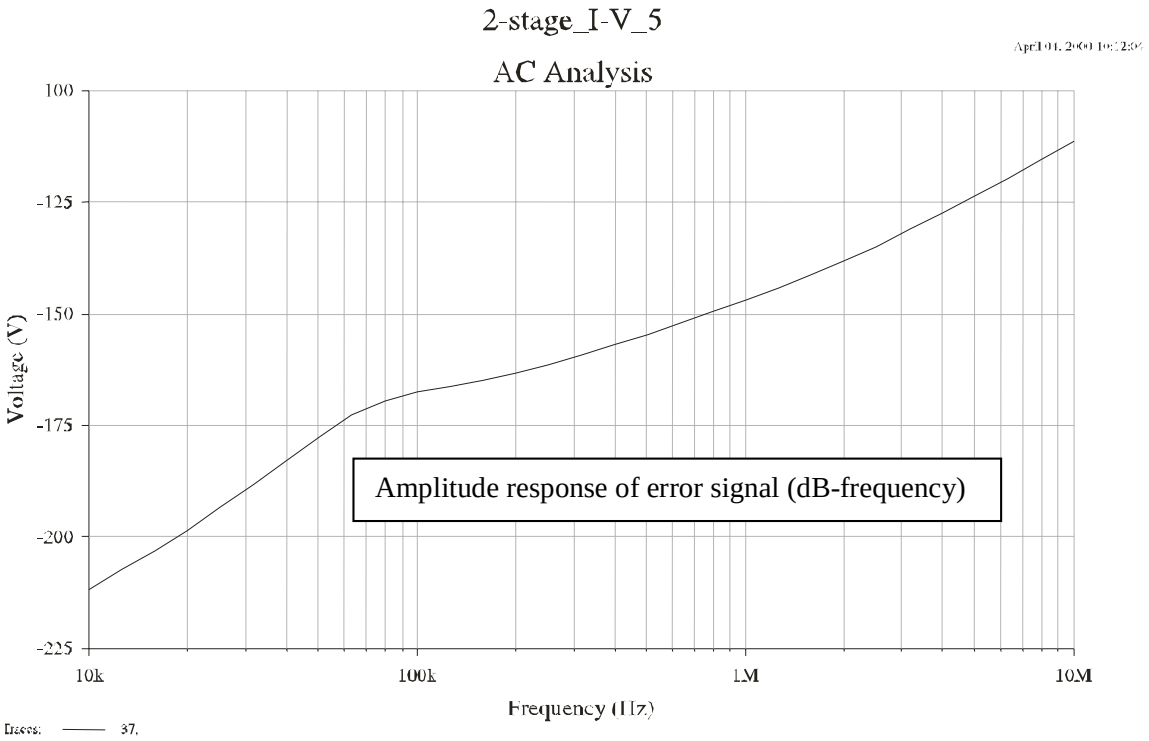


Figure 6-3 Parallel amplifier structure for determining error signal due to loop-gain error g.



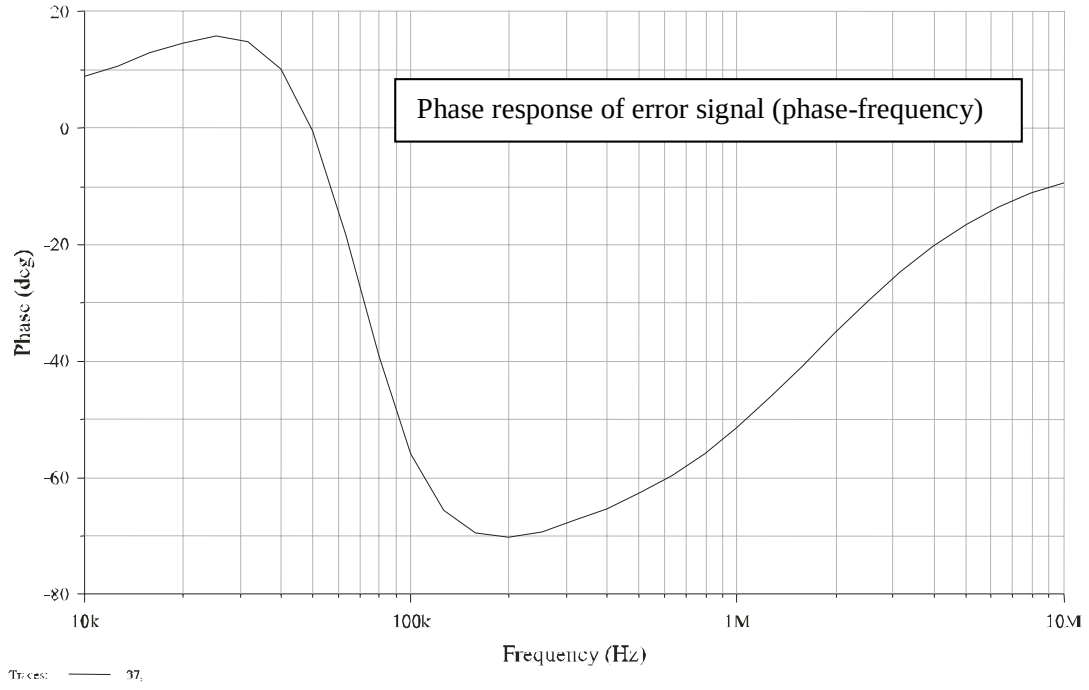
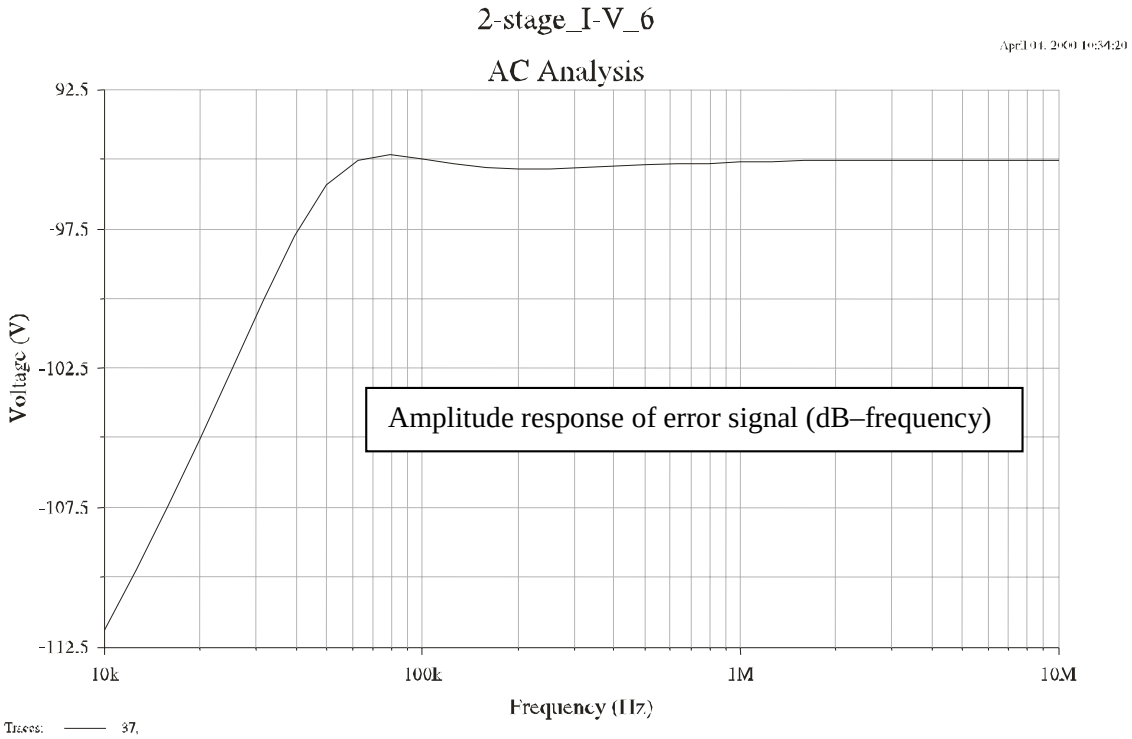


Figure 6-4 Error signal for 2-stage amplifier with 0.9 gain error in first stage.



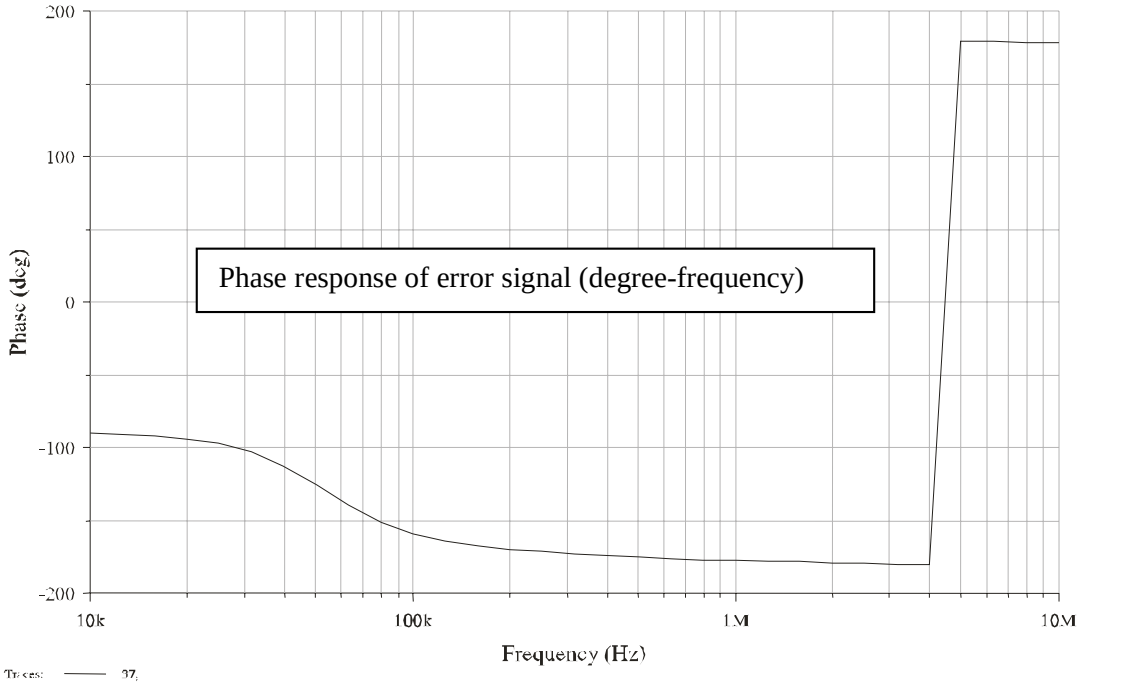


Figure 6-5 Error signal for 1-stage amplifier with a gain error of 0.9.

7 Conclusion

The design of current-to-voltage converters for high-resolution audio systems has been shown to be critical especially as transresistance amplifier non-linearity can be induced by the high-frequency signal components present in sampled audio signals at the output of a DAC. A number of distortion mechanisms have been discussed and the concept of jitter equivalence emphasised. Jitter is a performance indicator widely understood in digital audio, consequently transforming the distortion generated in an I/V conversion stage into an equivalent jitter sequence that is superimposed onto the sampled audio signal, forms a useful benchmark as well as unifying some aspects of the distortion.

As well as a detailed investigation of distortion mechanisms, the paper presented a number of circuit-level solutions using both discrete and operational amplifier topologies. Also, the use of input-stage error correction was illustrated. A method was described that enabled the low-pass recovery filter to be embedded into the topology where it became an integral structure within a nested-differential feedback topology. Such a technique allows the loop gain to be increased substantially, while band-limiting the signal within the transimpedance stage lowers slew-rate dependent distortion.

A dual-stage operational transimpedance amplifier was investigated and was shown to have a significantly lower sensitivity to operational amplifier loop-gain. It was also observed that the overall error signal in such stages has a significantly reduced phase shift at low frequency that is closer to 0 degree than to 90 degree, as occurs with a single stage amplifier. Another desirable attribute of the two-stage amplifier is that the output voltage of the first stage is extremely small, so eliminating slew-rate problems, while the embedded low-pass filter band-limits the signal applied to the second stage.

References

1 Burr Brown, PCM1702 20-bit digital-to-analogue converter, Application note PDS-1175, 1993

2 DVD-AUDIO FORMAT, Funasaka, E. and Suzuki, H., 103rd AES Convention, New York, NY, 1997 Sept. 26-29, preprint 4566

3 DISTORTION CORRECTION CIRCUITS FOR AUDIO AMPLIFIERS, M.J. Hawksford, *JAES*, vol.29, no.7, 8, July/August 1981

4 A 116 dB SNR MULTI-BIT NOISE SHAPING DAC WITH 192 kHz SAMPLING RATE, Adams, R., Nguyen, K. and Sweetland, K., 106th AES Convention, Munich, 1999, preprint 4963 S5

5 OVERSAMPLED ANALOG-TO-DIGITAL CONVERSION FOR DIGITAL AUDIO SYSTEMS, Hawksford, M.O.J. and Darling, T.E., *JAES*, vol. 38, no. 12, pp 924-943, December 1990

6 DIGITAL-TO-ANALOG CONVERTER WITH LOW INTERSAMPLE TRANSITION DISTORTION AND LOW SENSITIVITY TO SAMPLE JITTER AND TRANSRESISTANCE AMPLIFIER SLEW RATE, Hawksford, M.O.J., *JAES*, vol. 42, no. 11, pp 901-917, November 1994

7 RELATIONSHIP BETWEEN NOISE SHAPING AND NESTED DIFFERENTIATING FEEDBACK LOOPS, Hawksford, M.O.J. and Vanderkooy, J., *JAES*, vol. 47, no. 12, pp 1054-1060, December 1999

8 A NEW RESULT IN NEGATIVE-FEEDBACK THEORY AND ITS APPLICATION TO AUDIO POWER AMPLIFIERS, Cherry, E.M., *IEEE J.Circuit Theory and Applns*. 6, pp. 265-288, July 1978

9 DIFFERENTIAL-CURRENT DERIVED FEEDBACK (DCDF) IN ERROR CORRECTING AUDIO AMPLIFIERS, Hawksford, M.O.J., IEE Proceedings on Circuits, Devices and Systems, vol. 141, no 3, June 1994, pp 227-236

Low-Distortion Programmable Gain Cell Using Current-Steering Cascode Topology*

MALCOLM JOHN HAWKSFORD

University of Essex, Department of Electrical Engineering Science, Colchester, Essex, United Kingdom

A programmable gain cell is described which operates in a current-steering mode. The technique is shown to offer good linearity and a precisely defined maximum gain which exhibits near zero distortion. The cellular structure is presented together with an application circuit using error feedforward distortion correction within the input stage, while similar techniques realize a precision current mirror but configured using error feedback.

0 INTRODUCTION

The traditional circuitry that can be identified as the kernel of most analog programmable gain amplifiers (PGA) uses the translinear gain cell first described by Gilbert [1], with further derivatives reported in [2]–[6]. This basic cell is illustrated in Fig. 1.

In elementary form the cell is a differential current in (i_1), differential current out (i_2) structure that requires all transistors to have the same parameters (matched physically and thermally). If we assume

$$I_e = I_0 \, e^{qV_{be}/KT} \tag{1}$$

and since by Kirchhoff's law

$$V_{be1} - V_{be2} = V_{be3} - V_{be4}$$

then

$$\ln\left[\frac{I_1 + i_1}{I_1 - i_1}\right] = \ln\left[\frac{I_2 + i_2}{I_2 - i_2}\right]$$

which yields

$$\frac{i_1}{i_2} = \frac{I_1}{I_2} \, . \tag{2}$$

Eq. (2) suggests a seemingly ideal solution where a linear relationship exists between i_1 and i_2 and the gain is determined by the ratio of the bias currents I_1 and I_2.

The maximum gain setting is not specified by Eq. (2), and in fact will ultimately depend upon the device adherence to the exponential relationship specified by Eq. (1). That is, once there is deviation from Eq. (1), nonlinearity is introduced. Also unity gain ($I_1 = I_2$) is dependent upon device matching and as such is specified only as accurately as the transistors are matched.

The cell is also indirect. The signal current i_2 is derived by using T_3 and T_4 as a differential amplifier with T_1 and T_2 presented as nonlinear load resistors to the respective input currents $I_1 + i_1$ and $I_2 - i_2$. The nonlinear distortion is then minimized by matching the nonlinear device transfer characteristics.

The gain cell to be described in this paper uses a current-steering topology. It effectively eliminates a stage of amplification compared with the cell in Fig. 1 and most important, it has a well-defined upper unity gain (current in–current out) which is essentially linear and independent of device characteristics and matching—a most useful attribute for an audio channel.

This paper describes the theoretical basis of the current-steering gain cell and suggests an outline system topology that should enable high-performance PGAs to be designed. In addition to the main cell, a precision linear current mirror is described posessing wide dynamic range. The current mirror is used here within the PGA. However, it should also find application in mainstream constant-gain preamplifiers and power amplifier designs. Finally, the input circuitry proposed for the PGA includes error-correction feedforward to enhance transconductance, improve linearity, and minimize Johnson noise.

1 CURRENT-STEERING GAIN CELL

The elementary current-steering gain cell is illustrated in Fig. 2. The cell consists of two pairs of matched transistors that should adhere to the exponential I_e/V_{be} relationship defined by Eq. (1). Justification of the quiescent current distribution is given in the Appendix.

In Fig. 2, i_1 is the input signal current, $|i_1| < |I|$, i_2 and i_3 are output signal currents, I is bias current, m is a

* Presented at the 69th Convention of the Audio Engineering Society, Los Angeles, 1981 May 12–15.

 0004-7554/82/110795-05$00.75

gain-control parameter, $-1 < m < 1$, and V_k is a constant bias voltage.

1.1 Analysis

Applying Kirchhoff's law we obtain

$$V_k - V_{be1} + V_{be2} - V_{be3} + V_{be4} = V_k .$$

Therefore

$$(V_{be1} - V_{be2}) + (V_{be3} - V_{be4}) = 0 .$$

Using Eq. (1), we observe

$$V_{be1} - V_{be2} = \frac{KT}{q}\ln\left[\frac{I_{c1}}{I_0}\right] - \frac{KT}{q}\ln\left[\frac{I_{c2}}{I_0}\right] .$$

Therefore

$$V_{be1} - V_{be2} = \frac{KT}{q}\ln\left[\frac{I(1+m)/2 - i_2}{I(1+m)/2 + i_2}\right] .$$

Similarly for V_{be3} and V_{be4}:

$$V_{be3} - V_{be4} = \frac{KT}{q}\ln\left[\frac{I(1-m)/2 + i_3}{I(1-m)/2 - i_3}\right] .$$

Thus

$$\frac{I(1+m)/2 - i_2}{I(1+m)/2 + i_2} = \frac{I(1-m)/2 - i_3}{I(1-m)/2 + i_3}$$

where, after simplification,

$$\frac{i_2}{i_3} = \frac{1+m}{1-m} . \tag{3}$$

We also observe from Fig. 2 (where i_1 is the input current and i_2 and i_3 are output currents)

$$I + i_1 = \left(I\frac{1+m}{2} + i_2\right) + \left(I\frac{1-m}{2} + i_3\right) .$$

Therefore

$$i_1 = i_2 + i_3 . \tag{4}$$

Eqs. (3) and (4) reveal the operation of the current-steering PGA. The sum of the output currents is exactly equal to the input signal current (neglecting small base currents) and is independent of the I_c/V_{be} characteristics. Also the current division is linearly related to the gain parameter m. From Eqs. (3) and (4), i_2 and i_3 are derived:

$$i_2 = \frac{1+m}{2} i_1 \tag{5}$$

$$i_3 = \frac{1-m}{2} i_1 . \tag{6}$$

Thus the current gain of each cell is additionally complementary (that is, their sum is unity).

Eqs. (5) and (6) show that when $m = 1$ (or -1), then $i_2 = i_1$ (or $i_3 = i_1$), and on investigation Fig. 2 reveals that the input current is then steered through T_1 and T_2 (or T_3 and T_4) such that device nonlinearity has no effect, as the transistor pair acts as a cascode, assuming V_k is constant.

It is this latter mode of operation that is of greatest significance, since unity current gain results with excellent linearity due to the grounded base operation of transistors T_1 and T_2 or T_3 and T_4.

However, one problem still remains: the rejection of gain-control current under dynamic operation. The method by which control signal breakthrough is minimized is through the use of two techniques, a differential current mirror and two additionally complementary gain cells driven from a differential input stage. These are the subjects of Sections 2 and 4.

2 GENERALIZED PGA TOPOLOGY

This section describes the design of a basic PGA that uses two current-steering gain cells and two precision current mirrors. The circuitry is illustrated in the schematic of Fig. 3.

The input stage consists of a long-tail pair circuit with local emitter degeneration formed by R_1. Since a nonlinear fraction of the input signal $(V_1 - V_2)$ is developed across the two base–emitter junctions of the long-tail pair, two further long-tail pair stages are introduced to measure these error voltages and add corrective currents to each of the two output currents. Using feedforward error correction, greatly enhanced input stage performance with a significant reduction of nonlinearity in the transconductance transfer characteristic is achieved.

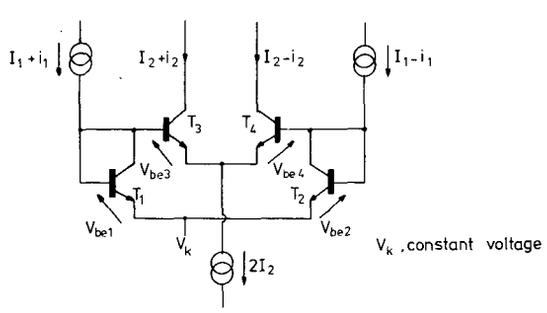


Fig. 1. Basic translinear gain cell.

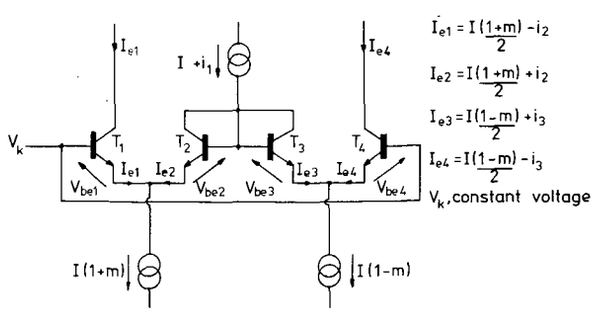


Fig. 2. Elementary current-steering gain cell.

2.1 Analysis

Referring to the input stage shown in Fig. 3,

$$V_1 - V_2 = i_1 R_1 + (V_{be1} - V_{be2})$$

where V_1 and V_2 are input signals.

Applying Eq. (1) to each error amplifier in Fig. 3,

$$V_{be1} = i_2 R_2 + \frac{KT}{q} \ln\left[\frac{I_2 + i_2}{I_2 - i_2}\right] + I_2 R_2$$

$$V_{be2} = -i_3 R_2 + \frac{KT}{q} \ln\left[\frac{I_2 - i_3}{I_2 + i_3}\right] + I_2 R_2$$

therefore,

$$V_1 - V_2 = i_1 R_1 + (i_2 + i_3) R_2 + \frac{KT}{q} \ln\left[\left(\frac{I_2 + i_2}{I_2 - i_2}\right)\left(\frac{I_2 + i_3}{I_2 - i_3}\right)\right] .$$

Assuming $i_2, i_3 < i_1$ such that $i_2, i_3 \ll I_2$,

$$V_1 - V_2 \simeq \left[i_1 + \left(\frac{i_2 + i_3}{R_1}\right)\left(R_2 + \frac{2KT}{qI_2}\right)\right] R_1 .$$

However, the output signal current i_0 is formed (see Fig. 3) by

$$i_0 = i_1 + i_2 + i_3 .$$

Hence if

$$R_1 = R_2 + 2\frac{KT}{qI_2} \tag{7}$$

then

$$i_0 = \frac{V_1 - V_2}{R_1} . \tag{8}$$

Since ΔV_{be1} and $\Delta V_{be2} < V_{in}/2$, then the error amplifiers operate well within their linear region. Thus provided that R_1 is selected according to Eq. (7), compensation for V_{be1} and V_{be2} is achieved, which results in the transconductance exhibiting excellent linearity with signal current.

The differential currents derived from the input stage provide signal currents for two current-steering gain cells, which in turn are both biased by identical gain-control currents. The outputs of the two cells are suitably crosscoupled, and when combined with two precision unity-gain current mirrors, produce additionally complementary output currents I_{01} and I_{02}. To aid understanding of this process, the circuit diagram shown in Fig. 3 should be observed with respect to the labeled currents. Under quiescent conditions $I_{x1} = I_{x2}$ and $I_{y1} = I_{y2}$, hence output currents I_{01} and I_{02} result, which are essentially independent of the gain-control currents.

The actual output currents generated when an input signal is applied are given by

$$I_{01} = I_k + \frac{V_1 - V_2}{R_1}(1 + m) \tag{9}$$

$$I_{02} = I_k - \frac{V_1 - V_2}{R_1}(1 - m) \tag{10}$$

where these currents can be converted to voltages by using suitably chosen load resistors R_L, as shown in Fig. 3.

3 LINEARIZATION USING NEGATIVE FEEDBACK

Although adequate performance can be achieved with the system shown in Fig. 3, it is possible to use the result that $I_{01} + I_{02}$ is independent of m. Thus an output voltage can be derived which is independent of the gain selected and used as a feedback signal that is returned to the input stage. Consequently the current-steering gain cells need operate only with small signal currents, which enhances linearity. The open-loop gain

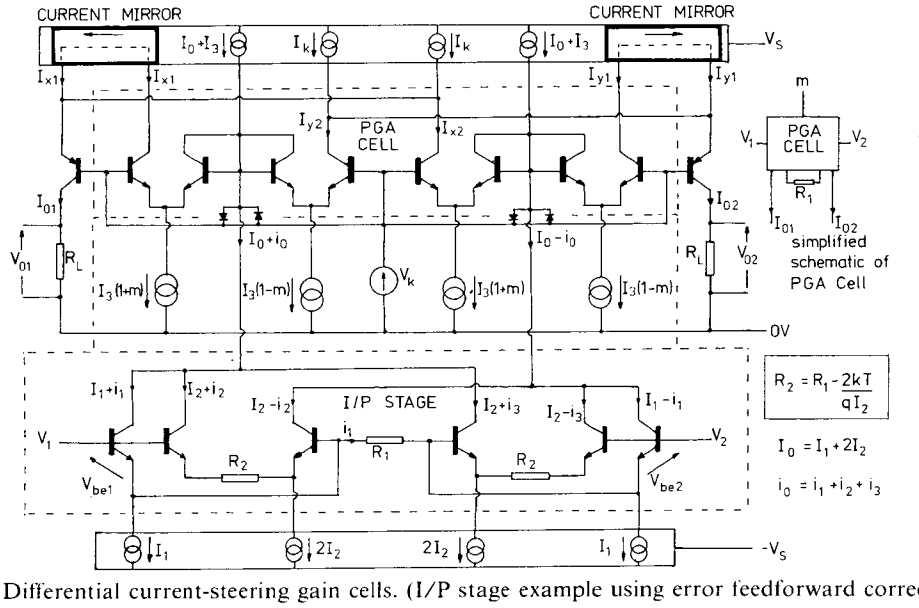


Fig. 3. Differential current-steering gain cells. (I/P stage example using error feedforward correction.)

is determined by selecting R_1 and R_5, while R_3 and R_4 determine the feedback factor. A basic feedback schematic is illustrated in Fig. 4, where the PGA cell is the circuit of Fig. 3. The output voltages V_{01} and V_{02} are then given by

$$V_{01} = V_1(1 + m)\frac{R_5(R_3 + R_4)}{R_1R_3 + R_1R_4 + R_4R_5} \quad (11)$$

$$V_{02} = V_1(1 - m)\frac{R_5(R_3 + R_4)}{R_1R_3 + R_1R_4 + R_4R_5} \quad (12)$$

4 CURRENT MIRROR

The two current mirrors in Fig. 3 can use an error correction feedback scheme to enhance linearity and minimize dependence upon the I_e/V_{be} characteristics. Fig. 5 illustrates a basic current mirror often used in discrete amplifier design, while Fig. 6 shows the enhanced design. Integrated current mirrors could be used, but they generally exhibit poor current gain linearity at currents in excess of a few milliamperes, especially at the extremes of their transfer characteristics.

4.1 Analysis

Since the bases of T_1 and T_2 are at the same potential, by Kirchhoff's law,

$$V_{be1} + I_1R + (I_1 + I_0 - i)R = V_{be2} + I_2R + (I_2 + I_0 + i)R .$$

Therefore

$$(V_{be1} - V_{be2}) + 2I_1R = 2iR + 2I_2R .$$

But within the error amplifier T_3, T_4,

$$V_{be1} - V_{be2} = (V_{be3} - V_{be4}) + iR_x$$

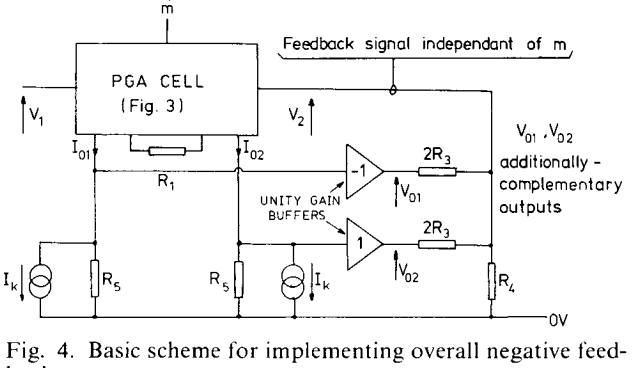


Fig. 4. Basic scheme for implementing overall negative feedback.

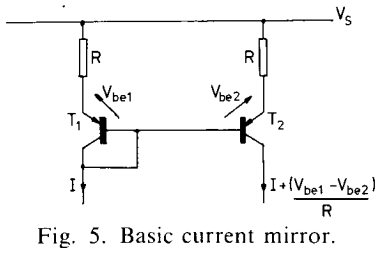


Fig. 5. Basic current mirror.

where

$$V_{be3} - V_{be4} = \frac{KT}{q}\ln\left[\frac{I_0 + i}{I_0 - i}\right] \simeq \left(\frac{2KT}{qI_0}\right)i, \quad \text{for modest } i .$$

Therefore

$$\left(\frac{2KT}{qI_0} + R_x\right)i + 2I_1R = 2iR + 2I_2R.$$

If

$$R_x = 2R - \frac{2KT}{qI_0} \quad (13)$$

then

$$I_2 = I_1 . \quad (14)$$

Since the error voltage $V_{be1} - V_{be2} \simeq 0$ due to T_1 and T_2 operating virtually with the same emitter current, the error amplifier is rendered linear. Consequently the correction system results in exceptional linearity over a wide range of the current transfer characteristic of the current mirror.

5 CONCLUSIONS

A topology for a programmable gain amplifier has been presented which can exhibit excellent gain stability and linearity. The cell offers the advantage of a precisely defined upper gain limit with the advantage that distortion is almost zero over a wide dynamic range.

The circuit requires transistor matching, but pair matching should be adequate, provided that transistors adhere closely to the logarithmic relationship between I_e and V_{be} (such as the LM394 at $I_c < 1$ mA).

An application circuit is presented which provides additionally-complementary gains, a topology which allows overall negative feedback to be used to further enhance linearity. However, the open-loop characteristics should yield adequate performance and may be preferred by designers of the low feedback school.

Additional to the basic cell, an input transconduc-

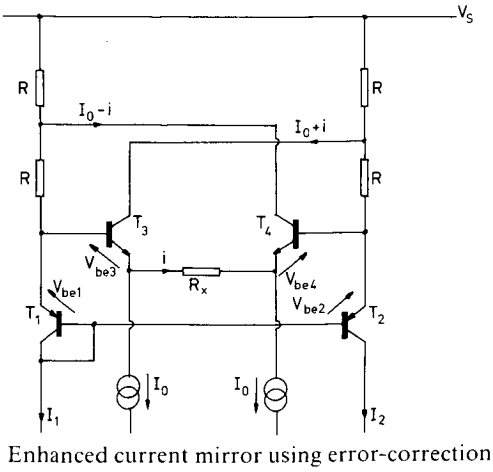


Fig. 6. Enhanced current mirror using error-correction feedback.

tance stage is described which uses error-correction feedforward to enhance linearity. Also a precision current mirror is presented which uses similar error correction, but configured within a feedback loop. These circuits offer excellent linearity, which is virtually independent of device characteristics.

6 REFERENCES

[1] B. Gilbert, "Translinear Circuits: A Proposed Classification," *IEE Electron. Lett.*, vol. 11, pp. 14–16 (1975 Jan. 9).

[2] B. Gilbert, "A New Wideband Amplifier Technique," *IEEE J. Solid-State Circuits*, vol. SC-3, pp. 353–365 (1968 Dec.).

[3] C. C. Todd, "A Monolithic Analog Compandor," *IEEE J. Solid-State Circuits*, vol. SC-11, pp. 754–762 (1976 Dec.).

[4] T. Yamaguchi, S. Takaoka, and K. Aizawa, "A New Configuration Using a Voltage Controlled Amplifier for a Dolby System Integrated Circuit," *IEEE Trans. Consumer Electron.*, vol. CE-25, pp. 723–729 (1979 Nov.).

[5] D. Baskind and H. Rubens, "Techniques for the Realization and Applications of Voltage-Controlled Amplifiers and Attenuators," presented at the 60th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 26, p. 572 (1978 July/Aug.), preprint no. 1378.

[6] D. Baskind, H. Rubens, and G. Kelson, "The Design and Integration of a High-Performance Voltage-Controlled Attenuator," presented at the 64th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 27, pp. 1018, 1020 (1979 Dec.), preprint no. 1555.

7 APPENDIX
QUIESCENT CURRENT DIVISION IN CURRENT-STEERED PGA CELLS

The current distribution shown in Fig. 2 is stated without formal justification. Here we show the current divisions to be valid. Since the circuit is considered in the quiescent state, i_1, i_2, i_3, and $i_4 = 0$. Let emitter currents of T_1, T_2, T_3, and T_4 be I_{e1}, I_{e2}, I_{e3}, and I_{e4}. Apply Kirchhoff's law to current distribution in Fig. 2:

$$I_{e1} + I_{e2} = I(1 + m) \tag{15}$$

$$I_{e3} + I_{e4} = I(1 - m) \tag{16}$$

$$I_{e2} + I_{e3} = I \; . \tag{17}$$

Apply Kirchhoff's law to the base-emitter voltages:

$$(V_{be1} - V_{be2}) + (V_{be3} - V_{be4}) = 0 \; .$$

Using Eq. (1),

$$\frac{KT}{q}\ln\left[\frac{I_{e1}}{I_{e2}}\right] + \frac{KT}{q}\ln\left[\frac{I_{e3}}{I_{e4}}\right] = 0 \; .$$

Hence

$$\frac{I_{e1}}{I_{e2}} = \frac{I_{e4}}{I_{e3}} = \lambda \; . \tag{18}$$

Substituting for λ in Eqs. (15) and (16),

$$I_{e2} = I\left(\frac{1 + m}{1 + \lambda}\right)$$

$$I_{e3} = I\left(\frac{1 - m}{1 + \lambda}\right) \; .$$

Hence from Eq. (17), $\lambda = 1$. Therefore

$$I_{e1} = I_{e2} \quad \text{and} \quad I_{e4} = I_{e3} \; .$$

We therefore derive I_{e1}, I_{e2}, I_{e3}, and I_{e4} from Eqs. (15) and (16):

$$I_{e1} = I\left(\frac{1 + m}{2}\right)$$

$$I_{e2} = I\left(\frac{1 + m}{2}\right)$$

$$I_{e3} = I\left(\frac{1 - m}{2}\right)$$

$$I_{e4} = I\left(\frac{1 - m}{2}\right)$$

The division of emitter currents is therefore as shown in Fig. 2. The output collector currents of T_1 and T_4 are almost equal to the emitter currents, provided that transistors have adequate current gain. Also near linearity between I_c and I_b (for modest signals) enhances linearity of the grounded base stage.

THE AUTHOR

Malcolm Hawksford was educated at the University of Aston in Birmingham, England, from 1965 to 1971. In 1968 he obtained a first class honors degree in electrical engineering, and that same year was awarded a BBC research scholarship to investigate the application of deltamodulation to color television. In 1972 he obtained a Ph.D. degree. In 1971 Dr. Hawksford became a lecturer at the University of Essex, England, in the Department of Electrical Engineering Science. During his time at Essex he actively pursued research projects within the field of audio engineering, where projects on power amplifier design, loudspeaker crossover design, analogue to digital conversion and music synthesis have been undertaken. He has presented papers at conventions of the Audio Engineering Society. He is currently a member of the AES, IEE and RTS.

LOW DISTORTION PROGRAMMABLE GAIN CELL USING CURRENT-STEERING CASCODE TOPOLOGY

1763(D-3)

M. J. Hawksford
University of Essex
Essex, United Kingdom

Presented at the 69th Convention 1981 May 12-15 Los Angeles

AN AUDIO ENGINEERING SOCIETY PREPRINT

LOW DISTORTION PROGRAMMABLE GAIN CELL USING CURRENT-STEERING CASCODE TOPOLOGY

Abstract

A programmable gain cell is described which operates in a current-steering mode. The technique is shown to offer good linearity and a precisely defined maximum gain that exhibits near zero distortion. The cellular structure is presented together with an application circuit using error feedforward distortion correction within the input stage, while similar techniques realise a precision current mirror but configured using error feedback.

Dr M J Hawksford
Audio Research Group
Department of Electrical Engineering Science
University of Essex
Wivenhoe Park
Colchester
Essex
ENGLAND

1. Introduction

The traditional circuitry that can be identified as the kernel of most analogue programmable gain amplifiers (PGA) use the translinear gain cell as first described by Barrie Gilbert[1], with further derivatives reported in references 2, 3, 4, 5 and 6. This basic cell is illustrated in Figure 1.

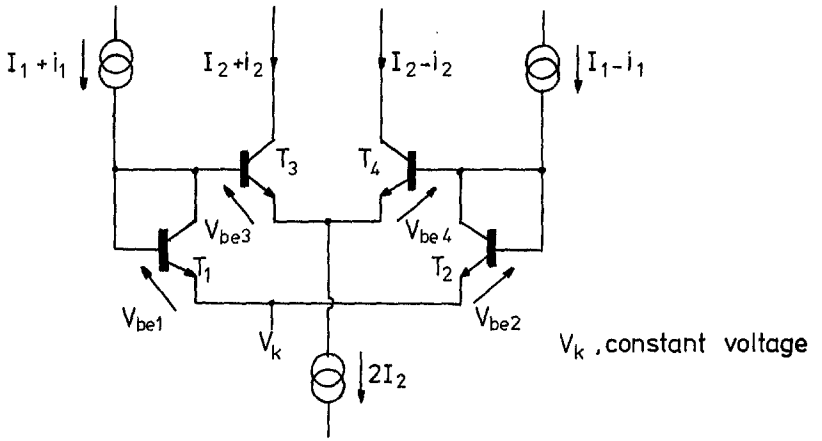


Figure 1 - Basic translinear gain cell

In elementary form, the cell is a differential current-in (i_1) differential current out (i_2) structure that requires all transistors to have the same parameters (matched physically and thermally).

If we assume,

$$I_e = I_0 \, e^{\frac{qV_{be}}{KT}} \qquad \ldots 1$$

and since by Kirchhoff

$$V_{be1} - V_{be2} = V_{be3} - V_{be4}$$

then,

$$\ln\left[\frac{I_1+i_1}{I_1-i_1}\right] = \ln\left[\frac{I_2+i_2}{I_2-i_2}\right]$$

which yields,

$$\frac{i_1}{i_2} = \frac{I_1}{I_2}$$

Equation 2 suggests a seemingly ideal solution where a linear relationship exists between i_1 and i_2 and the gain is determined by the ratio of the bias currents I_1 and I_2.

The maximum gain setting is not specified by equation 2 and in fact will ultimately depend upon the device adherence to the exponential relationship specified by equation 1. That is, once there is deviation from equation 1, non-linearity is introduced. Also unity gain ($I_1 = I_2$) is dependent upon device matching and as such is specified only as accurately as the transistors are matched.

The cell is also indirect. The signal current i_2 being derived by using T_3 and T_4 as a differential amplifier with T_1 and T_2 presented as non-linear load resistors to the respective input currents (I_1+i_1) and (I_2-i_2). The non-linear distortion is then minimised by matching the non-linear device transfer characteristics.

The gain cell to be described in this paper uses a current-steering topology. It effectively elliminates a stage of amplification compared with the cell in Figure 1 and most important: *It has a well-defined upper unity gain (current in - current out) which is essentially linear and independent of device characteristics and matching - a most useful attribute for an audio channel.*

This paper describes the theoretical basis of the current-steering gain cell and suggests an outline system topology that should enable high performance PGAs to be designed. In addition to the main cell, a precision linear current mirror is described possessing wide dynamic range. The current mirror is used here within the PGA, however, it should also find application in main-stream constant gain preamplifiers and power amplifier designs. Finally, the input circuitry proposed for the PGA includes error-correction feedforward to enhance transconductance, improve linearity and minimise Johnson noise.

2. *Current-Steering Gain Cell*

The elementary current-steering gain cell is illustrated in Figure 2. The cell consists of two pairs of matched transistors that should adhere to the exponential I_e/V_{be} relationship defined by equation 1. Justification of the quiescent current distribution is given in the Appendix (8).

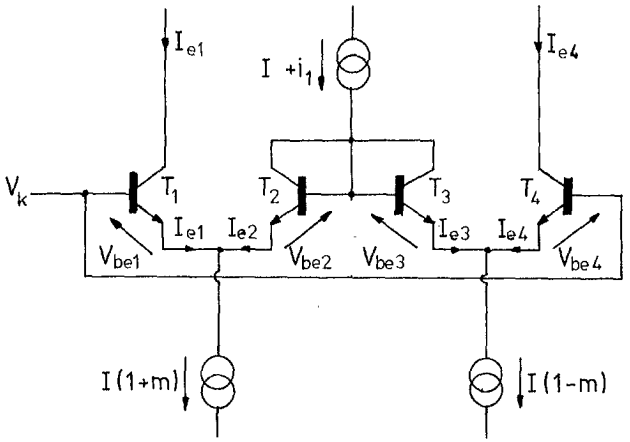


Figure 2 - Elementary current-steering gain cell

In Figure 2,

i_1 is input signal current $|i_1| < |I|$

i_2, i_3 output signal currents

I bias current

m gain control parameter, $-1 < m < 1$

V_k a constant bias voltage

Analysis

Applying Kirchhoff we obtain,

$$V_k - V_{be1} + V_{be2} - V_{be3} + V_{be4} = V_k$$

therefore

$$(V_{be1} - V_{be2}) + (V_{be3} - V_{be4}) = 0$$

Using equation 1 we observe;

$$(V_{be1} - V_{be2}) = \frac{KT}{q} \ln\left[\frac{I_{e1}}{I_0}\right] - \frac{KT}{q} \ln\left[\frac{I_{e2}}{I_0}\right]$$

therefore

$$(V_{be1} - V_{be2}) = \frac{KT}{q} \ln\left[\frac{I(\frac{1+m}{2}) - i_2}{I(\frac{1+m}{2}) + i_2}\right]$$

Similarly for V_{be3}, V_{be4}

$$(V_{be3} - V_{be4}) = \frac{KT}{q} \ln\left[\frac{I(\frac{1-m}{2}) + i_3}{I(\frac{1-m}{2}) - i_3}\right]$$

Thus,

$$\frac{I(\frac{1+m}{2}) - i_2}{I(\frac{1+m}{2}) + i_2} = \frac{I(\frac{1-m}{2}) - i_3}{I(\frac{1-m}{2}) + i_3}$$

where after simplification,

$$\frac{i_2}{i_3} = (\frac{1+m}{1-m}) \qquad \ldots 3$$

We also observe from Figure 2 (i_1 input current, i_2, i_3 output currents);

$$I + i_1 = \{I(\frac{1+m}{2}) + i_2\} + \{I(\frac{1-m}{2}) + i_3\}$$

therefore

$$i_1 = i_2 + i_3 \qquad \ldots 4$$

Equations 3 and 4 reveal the operation of the current steering PGA. The sum of the output currents is exactly equal to the input signal current (neglecting small base currents) and is independent of I_e/V_{be} characteristics. Also the current division is linearly related to m, the gain parameter. From equations 3 and 4, i_2 and i_3 are derived,

$$i_2 = (\frac{1+m}{2})\, i_1 \qquad \ldots 5$$

$$i_3 = (\frac{1-m}{2})\, i_1 \qquad \ldots 6$$

Thus the current gain of each cell is additionally-complementary (i.e. their sum is unity).

Equations 5 and 6 show that when m = 1 (or -1) then $i_2 = i_1$ (or $i_3 = i_1$) and on investigation Figure 2 reveals that the input current is then steered through T_1 and T_2 (or T_3 and T_4) such that device non-linearity has no effect, as the transistor pair acts as a cascode, assuming V_k is constant.

It is this latter mode of operation that is of greatest significance, since unity current gain results with excellent linearity due to grounded base operation of transistors T_1 and T_2 or T_3 and T_4.

However, one problem still remains: the rejection of gain control current under dynamic operation. The method by which control signal break-through is minimised is through the use of two techniques: a differential current mirror and two additionally-complementary gain cells driven from a differential input stage. These are the subjects of sections 3 and 5.

3. Generalised PGA Topology

This section describes the design of a basic PGA that uses two current-steering gain cells and two precision current mirrors. The circuitry is illustrated in the schematic of Figure 3.

The input stage consists of a long-tail pair circuit with local emitter degeneration formed by R_1. Since a non-linear fraction of the input signal ($V_1 - V_2$) is developed across the two base emitter junctions of the long-tail pair, two further long-tail pair stages are introduced to measure these error voltages and add corrective currents to each of the two output currents. Using feedforward error correction, greatly enhanced input stage performance with a significant reduction of non-linearity in the transconductance transfer characteristic is achieved.

Analysis

Referring to the input stage shown in Figure 3:

$$V_1 - V_2 = i_1R_1 + (V_{be1} - V_{be2})$$

where V_1, V_2 are input signals.

Applying equation 1 to each error amplifier in Figure 3:

$$V_{be1} = i_2R_2 + \frac{KT}{q} \ln\left[\frac{I_2+i_2}{I_2-i_2}\right] + I_2R_2$$

and

$$V_{be2} = -i_3R_2 + \frac{KT}{q} \ln\left[\frac{I_2-i_3}{I_2+i_3}\right] + I_2R_2$$

therefore

$$V_1 - V_2 = i_1R_1 + (i_2+i_3)R_2 + \frac{KT}{q} \ln\left[\left(\frac{I_2+i_2}{I_2-i_2}\right)\left(\frac{I_2+i_3}{I_2-i_3}\right)\right]$$

Assuming i_2, $i_3 < i_1$ such that i_2, $i_3 \ll I_2$

$$V_1 - V_2 \simeq \left[i_1 + \left(\frac{i_2+i_3}{R_1}\right)\left(R_2 + \frac{2KT}{qI_2}\right) \right] R_1$$

However, the output signal current i_0 is formed by,

$$i_0 = i_1 + i_2 + i_3 \qquad \text{(see Figure 3)}$$

Hence if

$$R_1 = R_2 + 2\frac{KT}{qI_2} \qquad \ldots 7$$

then

$$i_0 = \left(\frac{V_1 - V_2}{R_1}\right) \qquad \ldots 8$$

Since ΔV_{be1} and $\Delta V_{be2} < V_{in}/2$, then the error amplifiers operate well within their linear region. Thus, providing R_1 is selected according to equation 7, then compensation for V_{be1} and V_{be2} is achieved which results in the transconductance exhibiting excellent linearity with signal current.

The differential currents derived from the input stage provide signal currents for two current-steering gain cells, which in turn are both biased by identical gain control currents. The outputs of the two cells are suitably cross-coupled and when combined with two precision unity-gain current mirrors, produce additionally-complementary output currents I_{01}, I_{02}. To aid understanding of this process, the circuit diagram shown in Figure 3 should be observed with respect to the labelled currents. Under quiescent conditions $I_{x1} = I_{x2}$ and $I_{y1} = I_{y2}$, hence output currents I_{01}, I_{02} result which are essentially independent of the gain control currents.

The actual output currents generated when an input signal is applied are given by:

$$I_{01} = I_k + \left(\frac{V_1 - V_2}{R_1}\right)(1 + m) \qquad \ldots 9$$

$$I_{02} = I_k - \left(\frac{V_1 - V_2}{R_1}\right)(1 - m) \qquad \ldots 10$$

where these currents can be converted to voltages by using suitably chosen load resistors, R_L as shown in Figure 3.

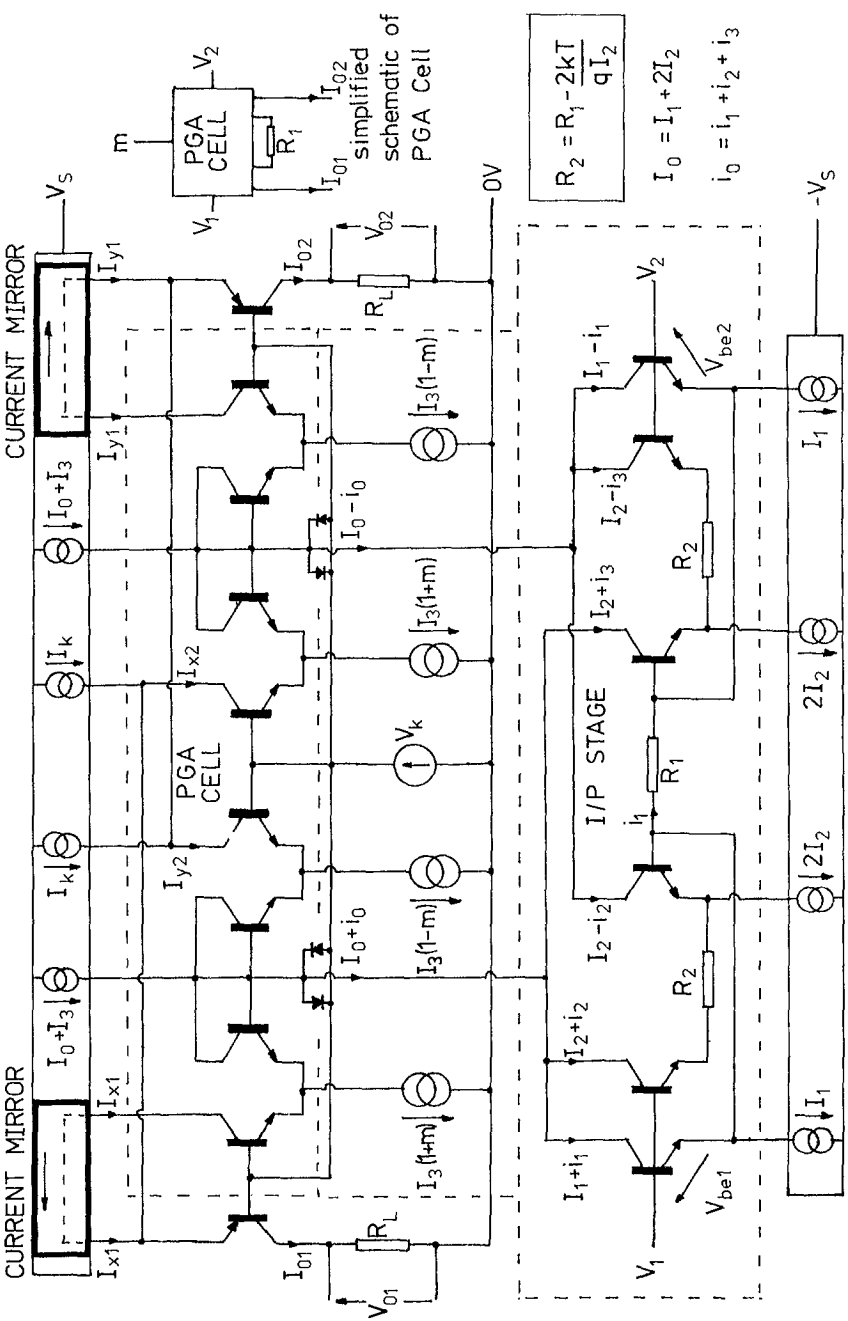


Figure 3 - Differential current-steering gain cells

(I/P stage example using error feedforward correction)

4. Linearisation Using Negative Feedback

Although adequate performance can be achieved with the system shown in Figure 3, it is possible to use the result that $I_{01} + I_{02}$ is independent of m. Thus, an output voltage can be derived which is independent of the gain selected and used as a feedback signal that is returned to the input stage. Consequently the current-steering gain cells need operate only with small signal currents which enhances linearity. The open-loop gain is determined by selecting R_1 and R_5, while R_3 and R_4 determine the feedback factor. A basic feedback schematic is illustrated in Figure 4 where the PGA cell is the circuit of Figure 3. The output voltage V_{01} and V_{02} are then given by;

$$V_{01} = V_1(1+m)\left[\frac{R_5(R_3+R_4)}{R_1R_3+R_1R_4+R_4R_5}\right] \quad \ldots 11$$

$$V_{02} = V_1(1-m)\left[\frac{R_5(R_3+R_4)}{R_1R_3+R_1R_4+R_4R_5}\right] \quad \ldots 12$$

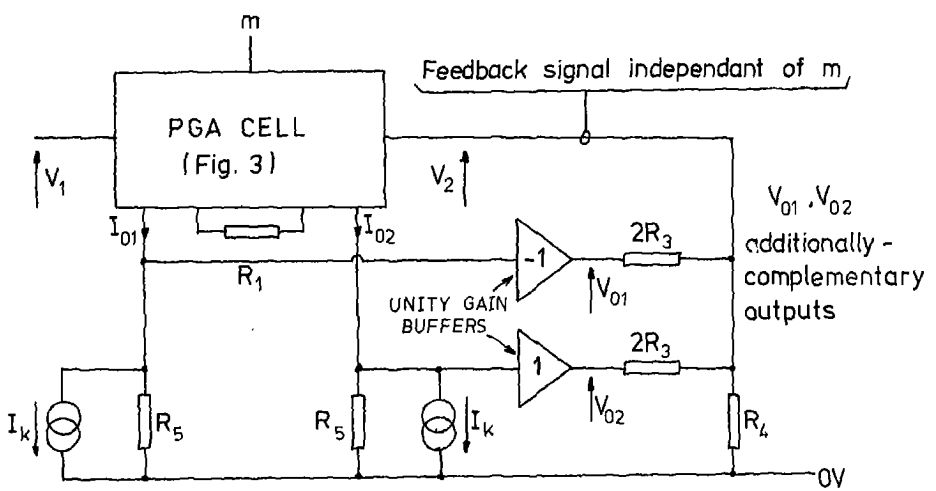


Figure 4 - Basic scheme for implementing overall negative feedback

5. Current Mirror

The two current mirrors used in Figure 3 can use an error correction feedback scheme to enhance linearity and minimise dependence upon I_e/V_{be}

characteristics. Figure 5 illustrates a basic current mirror often used in discrete amplifier design, while Figure 6 shows the enhanced design. Integrated current mirrors could be used, but they generally exhibit poor current gain linearity at currents in excess of a few milliamperes, especially at the extremes of their transfer characteristics.

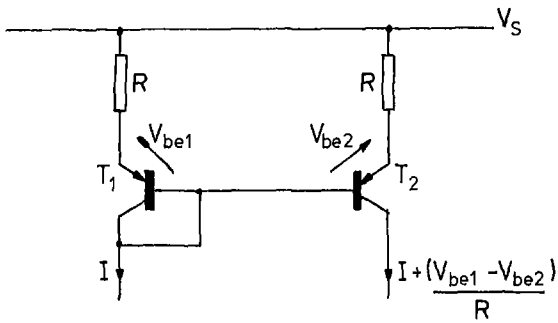


Figure 5 - Basic current mirror

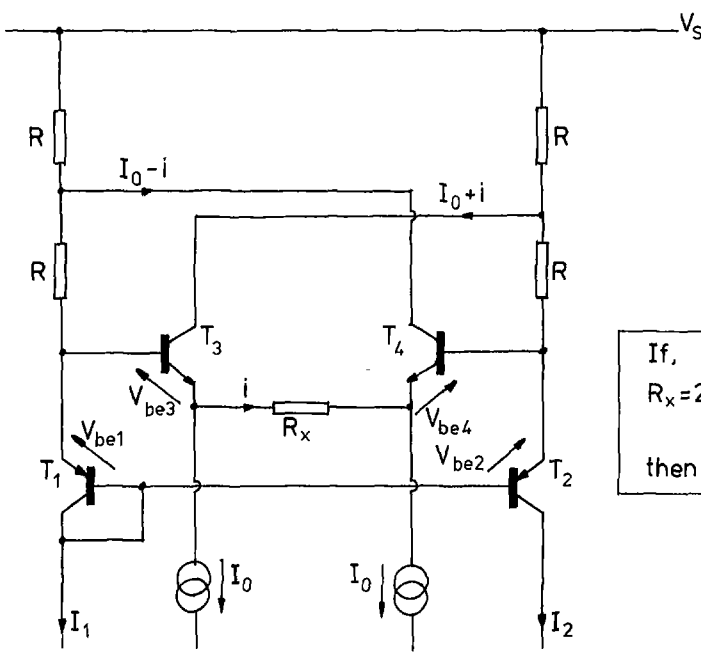


Figure 6 - Enhanced current mirror using error correction feedback

Analysis

Since the bases of T_1 and T_2 are at the same potential; by Kirchhoff

$$V_{be1} + I_1R + (I_1+I_0-i)R = V_{be2} + I_2R + (I_2+I_0+i)R$$

therefore

$$(V_{be1} - V_{be2}) + 2I_1R = 2iR + 2I_2R$$

But within the error amplifier T_3, T_4

$$(V_{be1} - V_{be2}) = (V_{be3} - V_{be4}) + iR_x$$

where,

$$(V_{be3} - V_{be4}) = \frac{KT}{q} \ln\left[\frac{I_0+i}{I_0-i}\right] \simeq (\frac{2KT}{qI_0})i \text{, for modest } i$$

therefore

$$\{(\frac{2KT}{qI_0}) + R_x\}i + 2I_1R = 2iR + 2I_2R$$

If,

$$R_x = 2R - \frac{2KT}{qI_0} \qquad \ldots 13$$

then

$$I_2 = I_1 \qquad \ldots 14$$

Since the error voltage $(V_{be1} - V_{be2}) \simeq 0$ due to T_1 and T_2 operating virtually with the same emitter current, the error amplifier is rendered linear. Consequently the correction system results in exceptional linearity over a wide range of the current transfer characteristic of the current mirror.

6. Conclusions

A topology for a programmable gain amplifier has been presented which can exhibit excellent gain stability and linearity. The cell offers the advantage of a precisely defined upper gain limit with the advantage that distortion is almost zero over a wide dynamic range.

The circuit requires transistor matching, but pair matching should be adequate providing transistors adhere closely to the logarithmic relationship between I_e and V_{be}, (e.g. LM394 at I_c < 1mA).

An application circuit is presented which provides additionally-complementary gains, a topology which allows overall negative feedback to be used to further enhance linearity. However, the open-loop characteristics should yield adequate performance and may be preferred by designers of the low feedback school.

Additional to the basic cell, an input transconductance stage is described which uses error correction feedforward to enhance linearity. Also a precision current mirror is presented which uses similar error correction but configured within a feedback loop. These circuits offer excellent linearity which is virtually independent of device characteristics.

7. References

1. Gilbert, B., 'Translinear Circuits: A Proposed Classification', IEE Electronic Letters, 9 January 1975, Vol 11, No 1, pp 14-16.

2. Gilbert, B., 'A New Wideband Amplifier Technique', IEEE Journal of Solid State Circuits, December 1968, SC-3, No 4, pp 353-365.

3. Todd, C.C., 'A Monolithic Analog Compandor', IEEE Journal of Solid State Circuits, December 1976, SC-11, No 6, pp 754-762.

4. Yamaguchi, T., Takaoka, S., Aizawa, K., 'A New Configuration Using a Voltage Controlled Amplifier for a Dolby System Integrated Circuit', IEEE Transactions on Consumer Electronics, November 1979, Vol CE-25, pp 723-729.

5. Baskind, D., Rubens, H., 'Techniques for the Realization and Application of Voltage Controlled Amplifiers', 60th Convention of the AES, May 2 1978, Preprint No 1378 (A-6).

6. Baskind, D., Rubens, H., 'The Design and Integration of High Performance Voltage Controlled Attenuators', 64th Convention of the AES, November 2 1979, Preprint No 1555 (D-2).

8. Appendix - Quiescent Current Division in Current Steered PGA Cells

The current distribution shown in Figure 2 is stated without formal justification. Here we show the current divisions to be valid. Since the circuit is considered in quiescent state i_1, i_2, i_3 and $i_4 = 0$. Let emitter currents of T_1, T_2, T_3 and T_4 be I_{e1}, I_{e2}, I_{e3} and I_{e4}. Apply Kirchhoff to current distribution in Figure 2:

$$I_{e1} + I_{e2} = I(1+m) \qquad \ldots A1$$

$$I_{e3} + I_{e4} = I(1-m) \qquad \ldots A2$$

$$I_{e2} + I_{e3} = I \qquad \ldots A3$$

Apply Kirchhoff to base emitter voltages:

$$(V_{be1} - V_{be2}) + (V_{be3} - V_{be4}) = 0$$

using equation 1,

$$\frac{KT}{q} \ln\left[\frac{I_{e1}}{I_{e2}}\right] + \frac{KT}{q} \ln\left[\frac{I_{e3}}{I_{e4}}\right] = 0$$

Hence

$$\frac{I_{e1}}{I_{e2}} = \frac{I_{e4}}{I_{e3}} = \lambda \qquad \ldots A4$$

Substituting for λ in equations A1 and A2

$$I_{e2} = I\left(\frac{1+m}{1+\lambda}\right)$$

$$I_{e3} = I\left(\frac{1-m}{1+\lambda}\right)$$

Hence from equation A3, $\lambda = 1$

Therefore

$$I_{e1} = I_{e2} \text{ and } I_{e4} = I_{e3}$$

We therefore derive I_{e1}, I_{e2}, I_{e3} and I_{e4} from A1 and A2

$$I_{e1} = I\left(\frac{1+m}{2}\right)$$

$$I_{e2} = I\left(\frac{1+m}{2}\right)$$

$$I_{e3} = I(\frac{1-m}{2})$$

$$I_{e4} = I(\frac{1-m}{2})$$

The division of emitter currents is therefore as shown in Figure 2. The output collector currents of T_1 and T_4 are almost equal to the emitter currents providing transistors have adequate current gain. Also near linearity between I_c and I_b (for modest signals) enhances linearity of the grounded base stage.

Topological Enhancements of Translinear Two-Quadrant Gain Cells*

MALCOLM O. J. HAWKSFORD

Department of Electronic Systems Engineering, University of Essex, Colchester, Essex, CO4 3SQ, UK

AND

P. G. L. MILLS

Tannoy Limited, Coatbridge, Strathclyde ML5 4TF, UK

A method is proposed for extending the performance regime of two-quadrant translinear gain cells that can both increase the gain-control range and lower distortion, particularly at extremes of attenuation. The general operating principles of translinear cells are reviewed and a laboratory design for an enhancement of the current-steering cell is presented. The design is supported by a range of measurements to validate the technique.

0 INTRODUCTION

The two-quadrant multiplier configured as a voltage-controlled amplifier (VCA) has found wide application as a dynamic gain element in audio systems, both for gain control and for program-controlled equalization. Consequently VCAs have been widely discussed in the literature [1]–[12]. In this paper the basic translinear, bipolar transistor gain cells are reviewed and an enhanced topology for the current-steering gain cell [12] is proposed, together with a prototype circuit and supporting performance data.

To realize gain control the audio circuit designer has available a range of methods that include motorized potentiometers, multiplying digital-to-analog converters (MDAC), FET attenuators, light-controlled resistors, pulse-modulation and switched-gain systems, and bipolar transistor arrays. However, this paper limits discussion to the bipolar array for implementing translinear gain elements as these are readily fabricated in integrated form and are compatible with large-scale analog systems.

Of particular importance when designing a VCA is the need to achieve an adequate performance regime over a wide range of the system's dynamic characteristic. A deficiency evident in several designs is the nature of distortion as a function of signal level and attenuation where, specifically at high attenuation, signal-to-distortion ratios deteriorate rapidly. This results from an inherent mechanism that yields a distortion residual at high attenuation in the output signal whose level with respect to the input remains approximately constant. Thus (at high attenuation) the distortion can be considerable. The enhanced topology presented directly addresses this problem and achieves a falling distortion with increased (high) attenuation.

To commence the study, the basic structures of translinear bipolar VCA cells are reviewed for tutorial value, and an approximate distortion analysis is presented that is directed at the VCA at high attenuation.

* Manuscript received 1988 May 4.

1 CLASSIFICATION OF TRANSLINEAR GAIN CELLS

The operation of a translinear cell is based on the logarithmic relationship between emitter current I_E and base–emitter voltage V_{BE} of a bipolar transistor, where

$$V_{BE} = \frac{kT}{q} \ln \left[\frac{I_E}{I_0}\right] \tag{1}$$

where

k = Boltzmann's constant
T = temperature, kelvin
q = charge on electron, coulomb
I_0 = saturation current, ampere.

Of particular concern is the temperature dependence of the transistor, both as a linear function of T and through the saturation current I_0, the latter being extremely temperature sensitive. To overcome this problem, the classic solution [1] is to use a transistor array of nearly identical devices operating under isothermal conditions, where a minimum of two transistors are required to compensate for I_0 and four transistors to compensate for T [see Eq. (1)].

Most of the VCA circuits are direct descendants of either the Gilbert cell or the log/antilog topologies as, for example, discussed in [6]. The dbx[1] VCA [8], which introduced the NPN/PNP array as a complementary cell, is also a direct descendant of the log/antilog configuration.

1.1 Two-Transistor Cells and Descendants

The basic two-transistor cell is illustrated in Fig. 1, where the operational amplifiers IC_1 and IC_2 suspend the transistor cell T_1, T_2 in a well-defined bias environment. Assuming IC_1 and IC_2 are ideal and ignoring base currents, then the operating conditions are

$$I_{E1} = V_i/R$$

$$I_{E2} = V_o/R$$

$$V_{CB1} = 0 \text{ V}$$

$$V_{CB2} = -V_g$$

where V_g is the gain-control voltage and V_{CB1} and V_{CB2} are the collector base voltages of T_1 and T_2, respectively.

Analysis

$$V_{BE1} - V_{BE2} = -V_g \ . \tag{2}$$

Assuming ideal behavior as stated in Eq. (1), then

$$\frac{kT}{q} \ln \left[\frac{I_{E1}}{I_{E2}}\right] = -V_g \tag{3}$$

where the dependence on saturation currents is eliminated provided T_1 and T_2 are matched and isothermal.

Hence eliminating I_{E1} and I_{E2} and rearranging,

$$\frac{V_o}{V_i} = e^{qV_g/kT} \ . \tag{4}$$

Eq. (4) shows the voltage gain of the VCA in Fig. 1 to be related to the gain-control voltage V_g through an exponential relationship; the dependence upon T should also be noted. Unfortunately, to operate correctly, the cell requires a value of $V_i > 0$ to prebias T_1 and T_2 to their active region, and this implies feedthrough of the gain-control function to the output, an undesirable characteristic for a two-quadrant cell.

To overcome this problem, two solutions have been proposed. One method consists of using two identical cells, as those in Fig. 1, but with a differential drive, as shown in Fig. 2,

$$V_{o1} = \left(V_i + \frac{v_i}{2}\right) e^{qV_g/kT}$$

$$V_{o2} = \left(V_i - \frac{v_i}{2}\right) e^{qV_g/kT} \tag{5}$$

$$V_o = V_{o1} - V_{o2} = v_i \, e^{qV_g/kT}$$

where

V_i = dc bias to bring cell into active region
v_i = signal component
V_{o1}, V_{o2} = respective outputs of two stages.

V_g also drives the noninverting inputs of the output operational amplifiers.

The second method, introduced in the dbx [8] cell, is to use a complementary four-transistor configuration, where the upper PNP transistors are also controlled by V_g, as shown in Fig. 3. Provided T_3 and T_4 have similar characteristics as T_1 and T_2 and isothermal conditions prevail, then for $V_i = 0$ V, $I_{C3} = I_{C1}$ and $I_{C4} = I_{C2}$. Thus gain-control feedthrough is compensated over the

[1] dbx is a trademark.

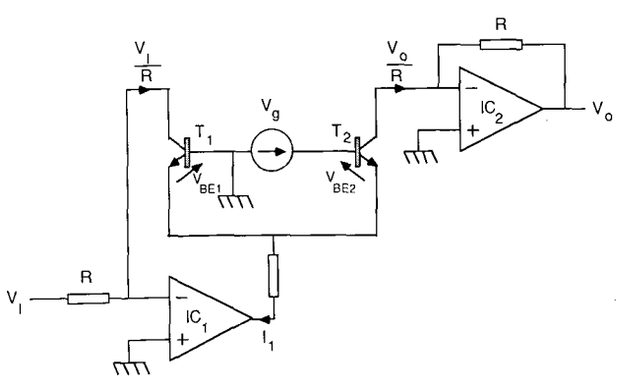


Fig. 1. Basic two-transistor log/antilog cell.

range of V_g. Of course, having decided to use a second pair of transistors, there is no reason why these cannot participate in the gain-control function, thus enhancing noise performance and dynamic range where, effectively, transistor pairs T_1/T_3 and T_2/T_4 appear in parallel. In Figs. 4 and 5 two variations are shown, one using an amplified diode while the second employs a floating supply such as a lead acid cell, a solar cell, or a switched-mode supply to eliminate bias current noise by not returning this current to ground. This latter circuit appears to offer much potential for low-noise operation, particularly where operating current levels are raised by the parallel connection of several transistor arrays. The insensitivity of gain to bias current level should also be noted.

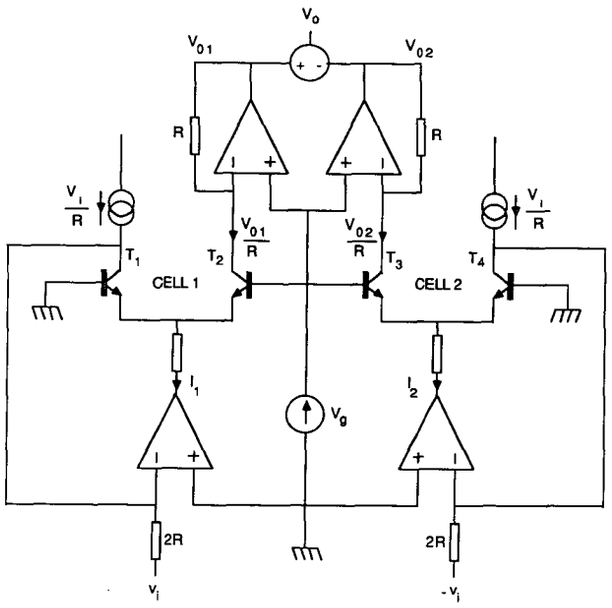


Fig. 2. Differential configuration of two-transistor log/antilog cell with bias.

Analysis of the Fig. 1 VCA revealed a linear relationship between emitter currents I_{E1} and I_{E2} where, from Eq. (3),

$$\frac{I_{E2}}{I_{E1}} = e^{qV_g/kT} .$$

Hence the tail current of T_1 and T_2 is also a linear function of I_{E1} and I_{E2} where, since $I_1 = I_{E1} + I_{E2}$, then

$$I_{E1} = I_1 \left(1 + e^{qV_g/kT}\right)^{-1} . \quad (6)$$

Noting the form of emitter currents I_{E1} and I_{E4} from Eq. 6, the voltage gain follows directly with reference to Fig. 6 as

$$\frac{v_o}{v_i} = 2g_m R \left(1 + e^{qV_g/kT}\right)^{-1} . \quad (7)$$

Consequently the Fig. 2 topology can be modified, whereby the tail currents form a differential input signal and the collector currents a differential output signal,

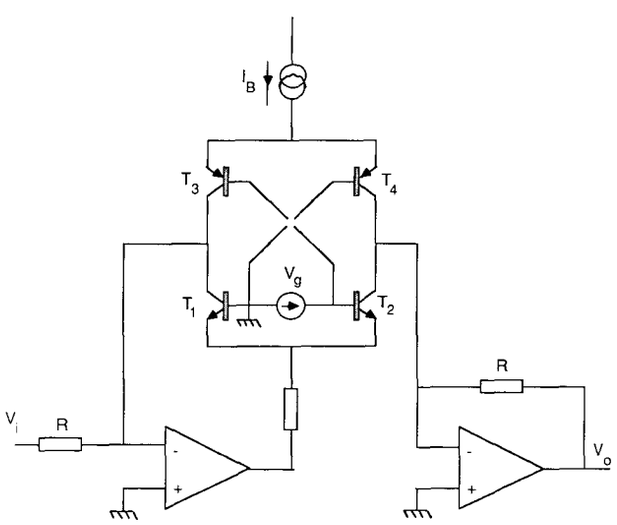


Fig. 3. Basic dbx voltage-controlled amplifier.

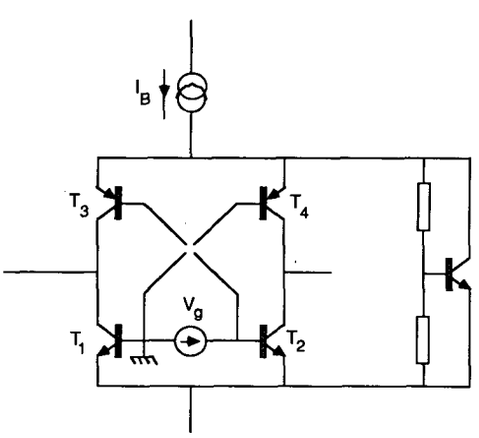


Fig. 4. dbx cell enhanced with amplified diode.

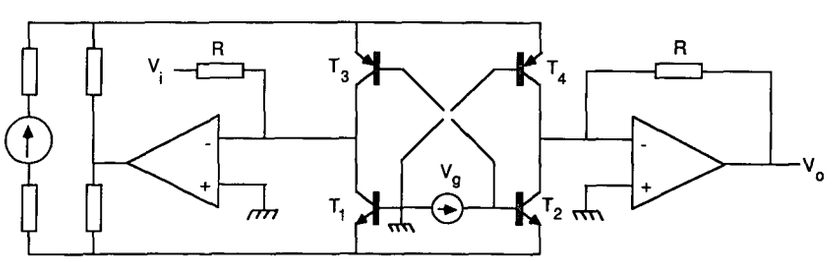


Fig. 5. Symmetrical dbx cell with floating supply.

as shown in Fig. 6 [see also Eq. (9)]. An important refinement of this reconfiguration is that for maximum gain T_1 and T_4 are now in common base (with T_2 and T_3 off); thus distortion is negligible.

1.2 Four-Transistor Cells

The two-transistor cells (and four-transistor derivatives) were shown to offer an exponential gain law and to retain a degree of temperature dependence. However, the use of four-transistor symmetrical circuits where the base–emitter voltages sum to zero can eliminate the temperature dependence and yield a linear gain law characteristic, where the desired gain law can be configured using nonlinear shaping circuits. The classic translinear circuit is the Gilbert multiplier [1] as shown in Fig. 7. Effectively, it is a differential amplifier where the base-to-base input voltage is predistorted by differentially driven diodes. However, operation is best understood by noting that the base–emitter voltages sum to zero and then applying Eq. (1) to each transistor whereby, assuming matching and isothermal operation, I_0 and T are eliminated.

Analysis

$$V_{BE1} - V_{BE2} - V_{BE3} + V_{BE4} = 0 \ . \tag{8}$$

Applying Eq. (1) to each (matched) transistor, then

$$\frac{I_{E1}}{I_{E2}} = \frac{I_{E3}}{I_{E4}} \ . \tag{9}$$

Neglecting base currents, the output V_o is expressed as

$$V_o = R(I_{E3} - I_{E4})$$

$$2I_g = (I_{E3} + I_{E4})$$

where

$$V_o = \left[\frac{RI_g}{V_i}\right] v_i \tag{10}$$

that is, the gain γ is given as

$$\gamma = \frac{RI_g}{V_1} \ . \tag{11}$$

The analysis shows a gain linearly dependent on I_g. Because all base–emitter voltages were summed to zero, T cancels, thus minimizing temperature dependence.

The current-steering gain cell, which is here allowed further study, was described earlier [12] and is shown in elementary form in Fig. 8. This cell also has the property of summing the four base–emitter voltages to zero, and is thus independent of T. Hence in this respect it differs fundamentally from the similar structure of Fig. 6. An analysis of the current-steering cell was given earlier [12], where it was shown that

$$\frac{i_0}{i_1} = 1 - \frac{I_g}{2I_1} \ . \tag{12}$$

Having reviewed a range of the basic topology for translinear-gain cells, we proceed by investigating the high-attenuation distortion of the current-steering cell and then introducing a modified structure for reducing distortion and extending the attenuation range to ≈140 dB at 1 kHz.

2 DISTORTION ANALYSIS OF CURRENT-STEERING CELL AT HIGH ATTENUATION

A measurement of distortion on the current-steering class of translinear circuits reveals a distortion residual that is approximately independent of attenuation. Consequently, at high attenuation, the signal-to-distortion ratio degrades, ultimately reaching an operation regime where the distortion is greater than the signal. This distortion mechanism reduces the effectiveness of the circuit and makes acceptable operation at high attenuation especially sensitive to misalignment, inherent transistor offsets, and input-stage finite common-mode gain.

Although the four-transistor current-steering cell illustrated in Fig. 8 is considered, similar discussion applies to other translinear circuits. Investigation has revealed three principal distortion contributions: finite effective emitter bulk resistance, dc base–emitter offsets, and finite common-mode gain in the input stage. The analysis proceeds by including emitter resistors r in each transistor that is both a representation of emitter

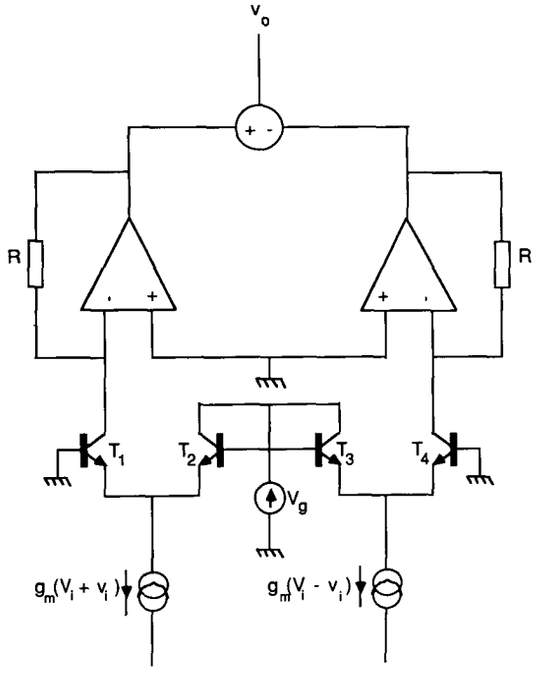


Fig. 6. Variation on log/antilog cell with input currents applied to T_1/T_2 and T_3/T_4 transistor pairs.

bulk resistance and a reflection of base bulk resistance. Also each transistor is assumed to exhibit a different saturation current, which results in an effective offset voltage V_{BEO} when the four-transistor base–emitter voltages are summed. The modified cell for analysis is shown in Fig. 9.

Analysis

$$I_g = 2I_1 - I_{01} - I_{04} \ . \tag{13}$$

$$\frac{kT}{q} \ln \left[\frac{I_{E2}I_{E4}}{I_{E1}I_{E3}}\right] + V_{BEO} + 2r(i_1 - i_0) = V_\varepsilon \tag{14}$$

where

$$i_0 = \frac{I_{01} - I_{04}}{2} \tag{15}$$

and V_ε is a correction voltage (see Fig. 9).

$$\lambda = -\frac{q}{2kT} [V_{BEO} + 2r(i_1 - i_0) - V_\varepsilon] \tag{16}$$

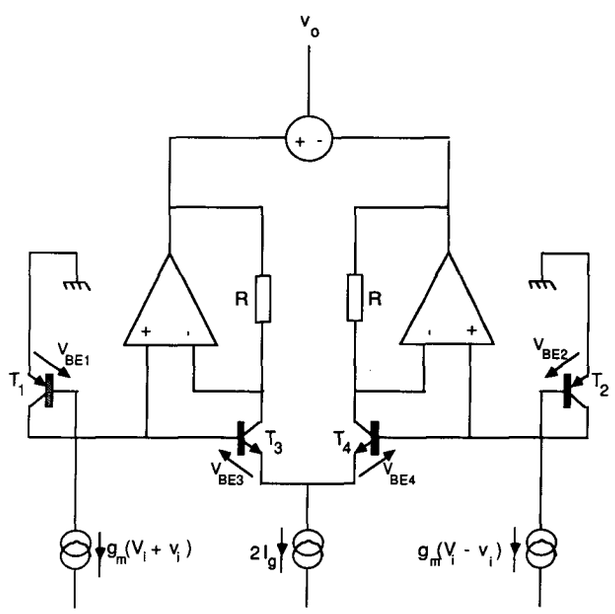


Fig. 7. Gilbert translinear gain cell [1].

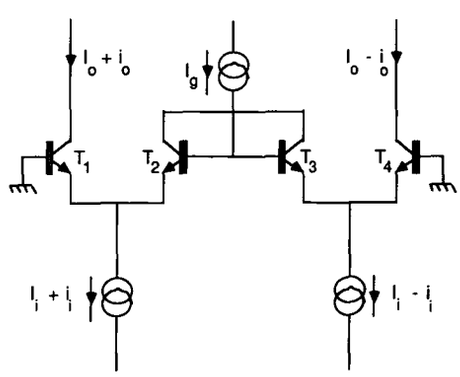


Fig. 8. Basic current-steering gain cell [12].

where

$$\ln \left[\frac{(I_1 + i_1 - I_{01})I_{04}}{(I_1 - i_1 - I_{04})I_{01}}\right] = 2\lambda \ . \tag{17}$$

Approximation

Let $I_{01} = \gamma(I_1 + i_1)$ and $I_{04} = \gamma(I_1 - i_1)$, where γ is the cell gain, that is,

$$\ln \left[\frac{(I_1 + i_1)(1 - \gamma)}{(I_1 - i_1)(1 - \gamma)} \frac{I_{04}}{I_{01}}\right] = 2\gamma \ .$$

Hence

$$\frac{(I_1 + i_1)I_{04}}{(I_1 - i_1)I_{01}} = e^{2\lambda} \ . \tag{18}$$

This simplification is justified for high gain ($\gamma \approx 1$) as there is minimal distortion in the output current while, for lower gains, $I_{01} << I_1 + i_1$ and $I_{04} << I_1 - i_1$.

Defining an input loading factor x,

$$x = \frac{i_1}{I_1} \tag{19}$$

and eliminating I_{01} and I_{04} using Eqs. (13), (15), (18), and (19),

$$i_0 = \frac{2I_1 - I_g}{2} \left(\frac{x - \tanh \lambda}{1 - x \tanh \lambda}\right) \ . \tag{20}$$

To deal with the nonlinearity and output offset current demonstrated by Eq. (20), incremental and target current gains γ_i and γ_t are introduced,

$$\gamma_i = \frac{\partial i_0}{\partial t i_1} \ , \qquad \gamma_t = \gamma_i \Big|_{\substack{V_{BEO} = 0 \\ r = 0}} \ .$$

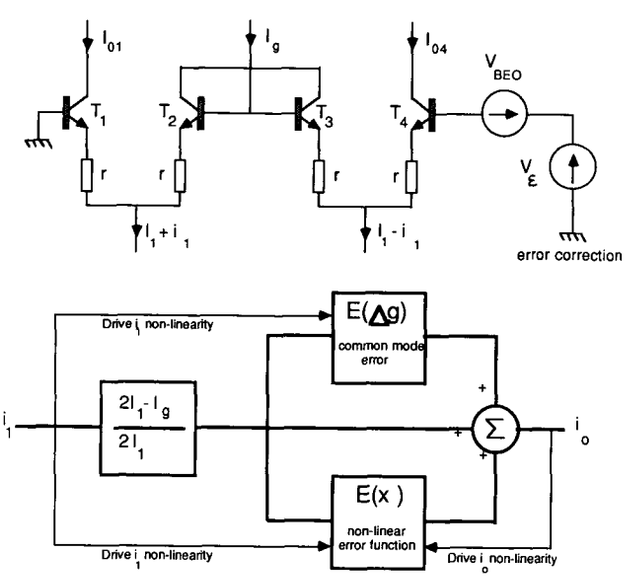


Fig. 9. Current-steering gain cell with approximate (low-level) error-function model [Eqs. (23), (27)].

Differentiating i_0 defined by Eq. (20),

$$\gamma_i = \operatorname{sech}^2 \lambda \left[\frac{1 + (1 - x^2)(1 - \gamma_i)\, qrI_1/kT}{(1 - x \tanh \lambda)^2} \right] \gamma_t \tag{21}$$

where, for $\lambda = 0$ and $r = 0\ \Omega$, the target current gain is

$$\gamma_t = \frac{2I_1 - I_g}{2I_1} \tag{22}$$

and takes the same form as the large-signal gain in Eq. (12).

To express the error in the incremental gain γ_i, an error function $E(x)$ is defined,

$$E(x) = \frac{\gamma_i}{\gamma_t} - 1$$

that is,

$$E(x) = \frac{(qrI_1/kT)(1 - x^2)(1 - \gamma_i) \operatorname{sech}^2 \lambda + \tanh \lambda\, [2x - \tanh \lambda(1 + x^2)]}{(1 - x \tanh \lambda)^2} \tag{23}$$

which, for $x \rightarrow 0$, simplifies to

$$E(0) = -\tanh^2 \lambda + \frac{qrI_1}{kT} (1 - \gamma_i) \operatorname{sech}^2 \lambda\ . \tag{24}$$

Eq. (23) describes an error function (see Fig. 9) that increases as $\gamma_i \rightarrow 0$ and is bounded by V_{BEO} and r. Even for small loading factors x, Eq. (24) shows a finite gain error that is dependent on λ. Although λ can be reduced by careful transistor matching, to minimize λ requires a corrective voltage V_ε to be added to the sum of the base-emitter voltages of the transistor array and is effectively in series with V_{BEO}, as shown in Fig. 9, where, observing Eq. (16),

$$V_\varepsilon = V_{BEO} + 2r(i_1 - i_0)\ . \tag{25}$$

This technique, which uses a constant voltage together with feedforward and feedback signals derived from input and output, respectively, is similar to that proposed by Bergstrom [13], though as this example demonstrates, it can be extended to other cell topologies.

Eq. (12) reveals the current gain to have significant sensitivity to I_1, particularly as $I_g \rightarrow 2I_g$, where small variations in either I_g or I_1 cause large fractional gain changes. At extreme attenuation, with T_1 and T_4 bias currents in the region 10^{-4} to $10^{-5} I_1$, slight distortion in the transconductance input circuitry that produces a small common-mode modulation of the bias current commensurate with T_1 and T_4 bias currents causes gain modulation of the cell. Also an imbalance between inverting and noninverting transconductance of the input stages produces a nonzero common-mode gain and has a similar effect.

To illustrate this distortion process, consider the following example where instantaneous tail currents of the T_1/T_2 and T_3/T_4 transistor pairs are i_{i1} and i_{i2}, respectively, and the noninverting and inverting transconductances are $(g_0 + \Delta_g)$ and $-(g_0 - \Delta_g)$, that is,

$$i_{i1} = g_0 v_i + (I_1 + \Delta g v_1)$$

$$i_{i2} = -g_0 v_i + (I_1 + \Delta g v_i)\ .$$

The equations for i_{i1} and i_{i2} reveal that under signal excitation a dynamic bias current $(I_1 + \Delta g v_i)$ is generated. If, for an idealized stage, $i_1 = g_0 v_i$, then the cell dynamic gain γ_d can be written following Eq. (12) as

$$\gamma_d = 1 - \frac{I_g}{2I_1\, [1 + (\Delta g/g_0)\, i_1/I_1]}\ . \tag{26}$$

Hence defining an error function $E(\Delta g)$ as

$$E(\Delta g) = \frac{\gamma_d}{\gamma_t} - 1$$

where γ_t is defined by Eq. 12, then assuming $\Delta g i_1 / g_0 I_1 << 1$,

$$E(\Delta g) = \frac{1}{2I_1/I_g - 1} \cdot \frac{\Delta g i_1}{g_0 I_1}\ . \tag{27}$$

Eq. (27) shows that for high attenuation, where $I_g \rightarrow 2I_1$, the gain is extremely sensitive to Δg. Consequently measures should be taken to ensure a negligible common-mode error resulting from both linear and nonlinear distortions in the input circuitry.

The key to cell enhancement is therefore to limit the attenuation per cell and to maintain an effective operating current in all transistors. In the next section a two-stage gain cell is described which offers potential improvements in distortion performance together with a falling distortion with increased attenuation.

3 ENHANCED TWO-STAGE CURRENT-STEERING CELL

The enhanced current-steering cell employs an extended two-stage topology with attenuation divided equally between each stage. Consequently each stage now operates over a more restricted attenuation range, which offers the following principal advantages:

1) The first stage operates with the full input signal, but because its attenuation is now restricted and transistors do not operate at their limits, a relatively low distortion is returned with a low sensitivity to finite common-mode gain in the input stage [see Eq. (27)].

2) The second stage operates under identical bias conditions but accepts a much lower input signal. Consequently second-stage distortion is dramatically reduced where the first stage is the dominant distortion generator.

3) However, the distortion at the output of the first stage is attenuated by the second stage. Thus, overall, the distortion now progressively falls with increased attenuation, even though at high attenuation the first stage tends to produce a constant distortion component referred to input level.

The primitive two-stage gain cell based on the current-steering cell is shown in Fig. 10 together with two equal gain-control currents of magnitude I_g. The cell $\{T_1 \ldots T_4\}$ operates conventionally, but with collectors of T_1 and T_4 feeding into the second stage $\{T_5 \ldots T_8\}$. However, for both stages to operate under similar bias conditions at a given attenuation, a current $I_g/2$ must also be summed to compensate for the loss of current in T_1 and T_4 when $I_g > 0$. Under these conditions, the expression for current gain becomes

$$\frac{i_0}{i_1} = \left(\frac{2I_1 - I_g}{2I_1}\right)^2 . \tag{28}$$

The circuit of Fig. 10 requires minimal extra voltage headroom where a typical value for V_B (see Fig. 10) is between 1 and 2 V. It is also relatively simple to generate the extra control current sources. Although the circuit requires an extra four transistors, this method of extension is more effective than combining arrays $\{T_1 \ldots T_4\}$ and $\{T_5 \ldots T_8\}$ in parallel.

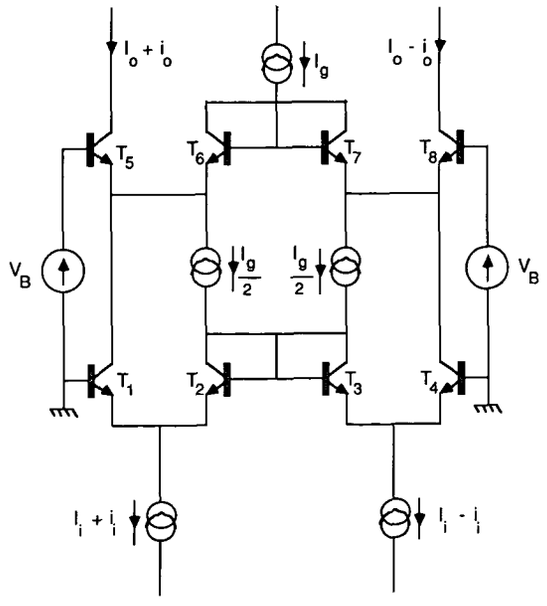


Fig. 10. Primitive two-stage current-steering gain cell.

To validate the modified current-steering cell, a prototype circuit was constructed and is shown in Fig. 11, while Fig. 12 illustrates the power supply and biasing arrangement. To minimize circuit noise through extra circuitry requirements and low bias resistor values, a dual supply rail of ± 30 V was used from which internal voltages of ± 15 V were derived.

A novel feature of the circuit was to use the gain-control voltage $V_{control}$ to drive a servo amplifier IC_6, which set the quiescent output voltages of IC_3 and thus determined the collector currents of T_5 and T_8. The servo amplifier then drove three grounded-base stages whose output currents I_{C1}, I_{C2}, and I_{C3} established the required gain-control currents and second-stage input bias compensation.

The complete VCA illustrated in Fig. 13 was assembled on a printed circuit board and metal plates attached to the LM394 arrays to aid isothermal operation. The circuit also includes switches SW_1 and SW_2 to facilitate dc alignment, hence the mimimization of gain-control feedthrough. The circuit alignment procedure is outlined in the Appendix.

4 MEASURED PERFORMANCE OF TWO-STAGE CURRENT-STEERING CELL

The enhanced circuit of Figs. 11 and 12 was assessed using a range of distortion and noise measurements, and the results are presented in Table 1. As anticipated, the distortion at maximum gain is of low level, a characteristic of the current-steering cell which effectively becomes a grounded-base stage. However, also encouraging is the distortion at 20 kHz, +15 dBV input, where for an attenuation of +100 dB a distortion of −122 dB referred to input was recorded, while for 20 kHz, +0 dBV, the distortion could not be resolved at +100 dB attenuation with available instrumentation.

5 EXTENSION OF CASCADE TO dbx ARRAY

In principle, the cascade technique can also be applied to other cells, such as the dbx array. A primitive circuit is shown in Fig. 14, which uses two complementary transistor arrays. Ideally, to minimize noise, both cells could be powered from independent floating supplies, as suggested earlier. Also, such an approach would appear to be particularly beneficial when the cells are operated in so-called class AB [8] mode (particularly for second stage), which is used to minimize noise generation from within the transistor array. However, these comments are made purely as a suggestion for further development. They are not supported at this stage with experimental verification; but, from the results obtained from the current-steering cell, improvements in distortion performance are to be anticipated.

6 CONCLUSION

This report has presented a tutorial review to outline a number of basic approaches to the translinear mul-

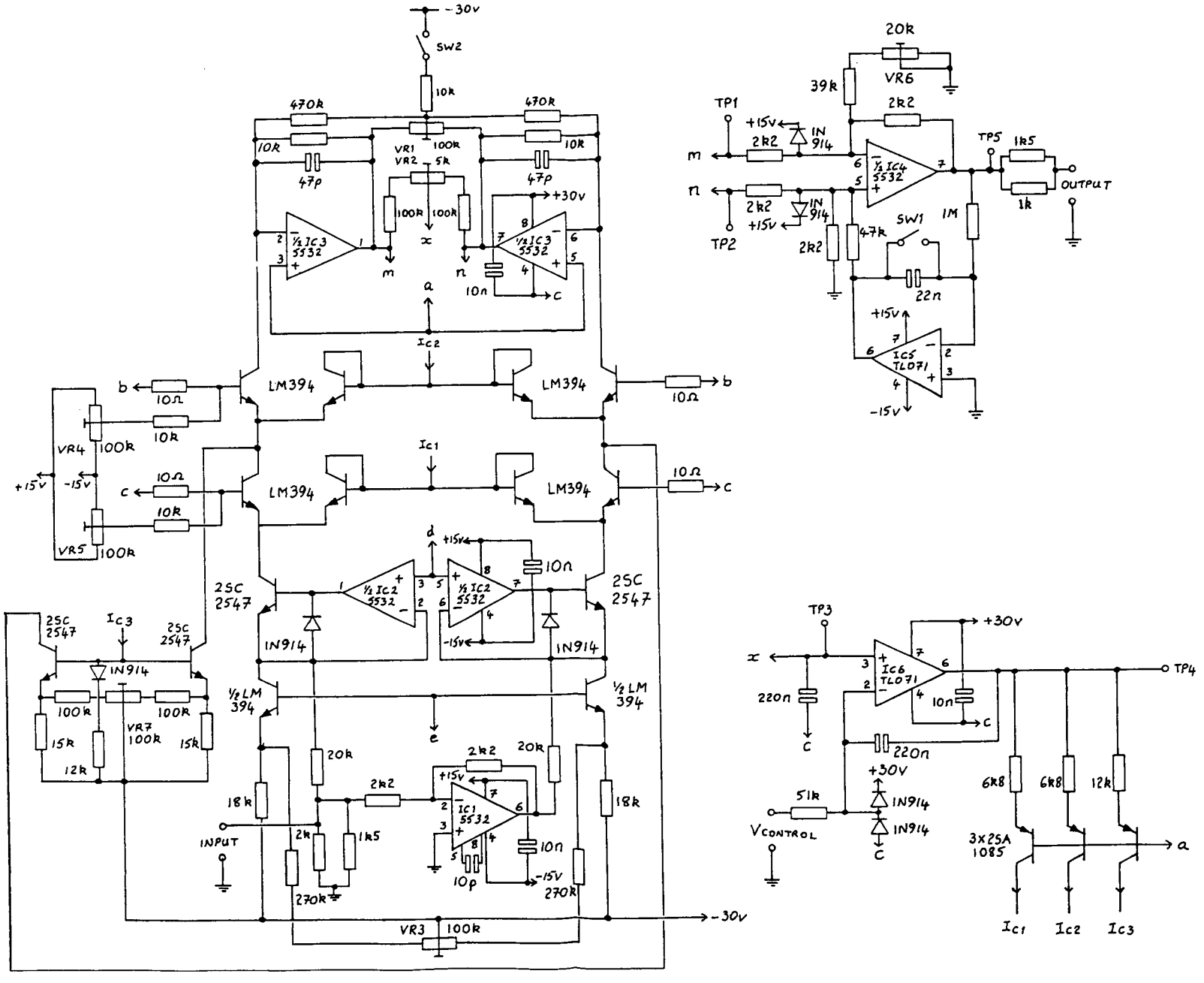


Fig. 11. Voltage-controlled amplifier, main circuit.

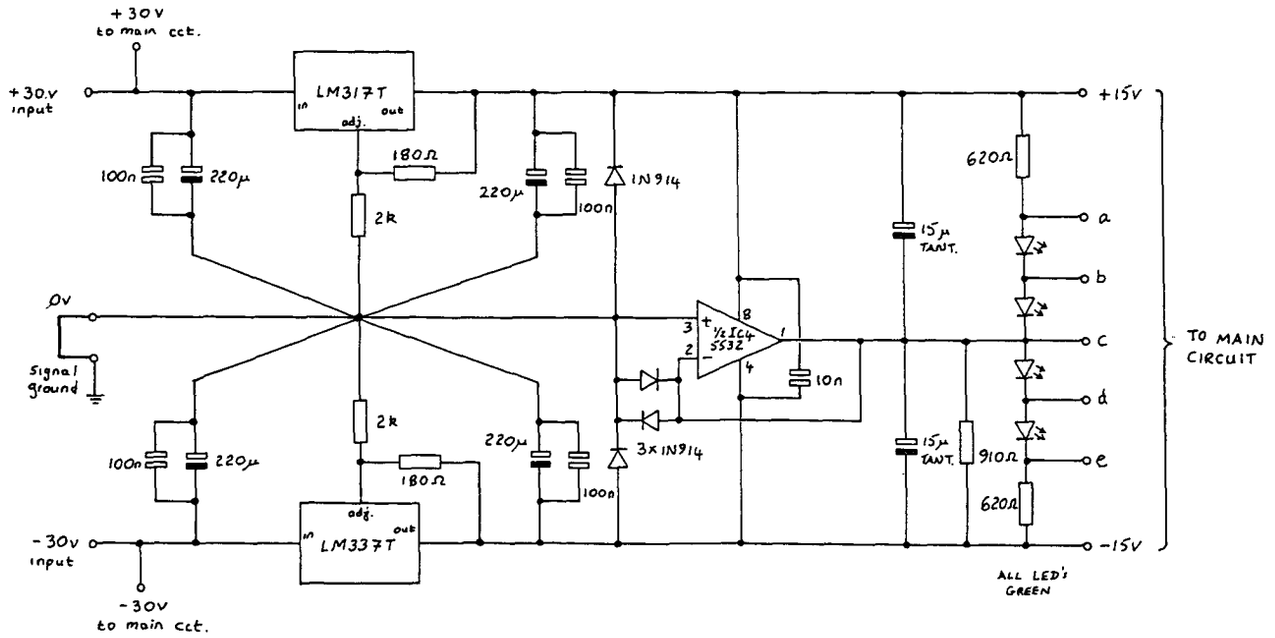


Fig. 12. Voltage-controlled amplifier, power supply.

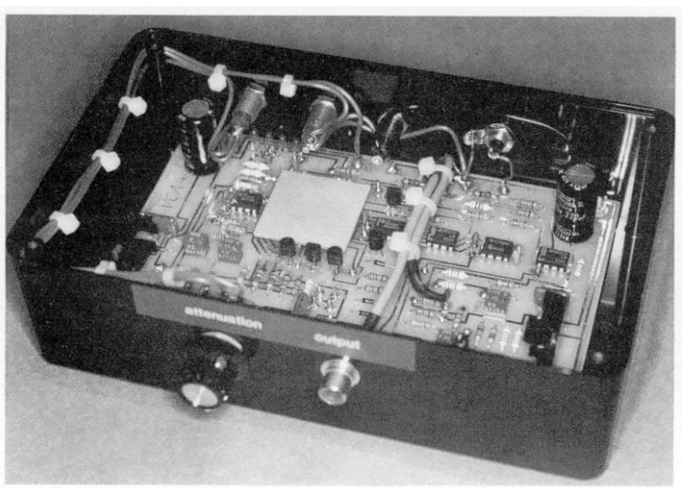


Fig. 13. Two-stage cascade prototype voltage-controlled amplifier.

Table 1. Voltage-controlled amplifier performance data.

Maximum gain		0 dB	
Maximum attenuation		140 dB at 1 kHz 115 dB at 20 kHz	
Signal-to-noise ratio (20 Hz–20 kHz unweighted)			
Maximum gain		−105 dB	
Minimum gain		−110 dB	
Distortion (THD) at maximum gain			
0-dBV input		1 kHz—not measurable; 20 kHz—0.0008%	
+15-dBV input		1 kHz—0.0004%; 20 kHz—0.0025%	
Distortion with attenuation (referred to input)			
Input level	Attenuation (dB)	1-kHz distortion (dB)	20-kHz distortion (dB)
0 dBV	+20	−94	−92
	+40	−105	−104
	+60	−115	−112
	+80	−125	−120
	+100	Not measurable	Not measurable
+15 dBV	+20	−82	−80
	+40	−89	−88
	+60	−98	−97
	+80	−108	−107
	+100	−126	−122

tiplier cell, together with an extension using two (or more) arrays connected in a cascade. To validate the technique, an experimental circuit was introduced from which a range of measurements was derived. The two-stage approach exhibited a significant improvement and showed in particular a useful reduction of distortion at high attenuation. For a given number of transistors (eight in this example) the cascade arrangement appears to offer a useful improvement over the parallel option. However, there is no fundamental reason why arrays of both serial and parallel connections should not be formed to yield further enhancement. With the experimental circuits, it proved desirable to operate from relatively high supply rails since this saved additional circuitry and the use of reduced resistor values which, in general, would otherwise have impaired the noise performance.

The report attempted an approximate distortion analysis of the current-steering cell operating at high attenuation where the need for zero common-mode bias current modulation was highlighted. The analysis also supported the desirability for using cells with more limited attenuation configured within a multistage cascade and suggested how distortion reduction may be introduced.

Although the circuits of Figs. 11 and 12 are semidiscrete prototypes, the system has potential for integrated-circuit development and for forming the basis of an effective VCA. Also, extensions to complementary NPN/PNP current-steering cells would appear feasible and allow a useful reduction in supply rail voltage, particularly if a floating gain cell supply was configured.

7 ACKNOWLEDGMENT

The authors would like to offer their appreciation to Ben Duncan of Band J Sound for his interest in this project and for drawing their attention to the Bergstrom patent [13].

8 REFERENCES

[1] B. Gilbert, "A Precise 4-Quadrant Multiplier with Subnanosecond Response," *IEEE J. Solid-State Circuits*, vol. SC-3, pp. 365–373 (1968 Dec.).

[2] Sansen and Meyer, "Distortion in Bipolar Transistor Variable Gain Amplifiers," *IEEE J. Solid-State Circuits*, vol. SC-8 (1973 Aug.).

[3] B. Gilbert, "A High-Performance Monolithic Multiplier Using Active Feedback," *IEEE J. Solid-State Circuits*, vol. SC-9, pp. 364–373 (1974 Dec.).

[4] B. Gilbert, "Translinear Circuits: A Proposed Classification," *IEE Electron. Lett.*, vol. 11, pp. 14–16 (1975 Jan. 9).

[5] B. Gilbert, "A New Technique for Analog Multiplication," *IEEE J. Solid-State Circuits*, vol. SC-10, pp. 437–447 (1975 Dec.).

[6] D. H. Sheingold, *Nonlinear Circuits Handbook* (Analog Devices, 1976).

[7] C. C. Todd, "A Monolithic Analog Compander," *IEEE J. Solid-State Circuits*, vol. SC-11, pp. 754–762 (1976 Dec.).

[8] D. Baskind and H. Rubens, "Techniques for the Realization and Application of Voltage-Controlled Amplifiers and Attenuators," presented at the 60th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 26, p. 572 (1978 July/Aug.), preprint 1378.

[9] D. Baskind, H. Rubens, and G. Kelson, "The Design and Integration of a High-Performance Voltage-Controlled Attenuator," presented at the 64th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 27, pp. 1018, 1020 (1979 Dec.), preprint 1555.

[10] T. Yamaguchi, S. Takaoka, and K. Aizawa, "A New Configuration Using a Voltage Controlled Amplifier for a Dolby System Integrated Circuit," *IEEE Trans. Consumer Electron.*, vol. CE-25, pp. 723–729 (1979 Nov.).

[11] B. Gilbert and P. Holloway, "A Wideband Two-Quadrant Analogue Multiplier," in *Proc. IEEE Int. Solid-State Circuits Conf.* (1980 Feb.), pp. 200–201.

[12] M. J. Hawksford, "Low-Distortion Programmable Gain Cell Using Current-Steering Cascode Topology," *J. Audio Eng. Soc.*, vol. 30, pp. 795–799 (1982 Nov.).

[13] G. Bergstrom, "Signal Correction for Electrical Gain Control Systems," US patient 4,234,804 (1980 Nov. 18).

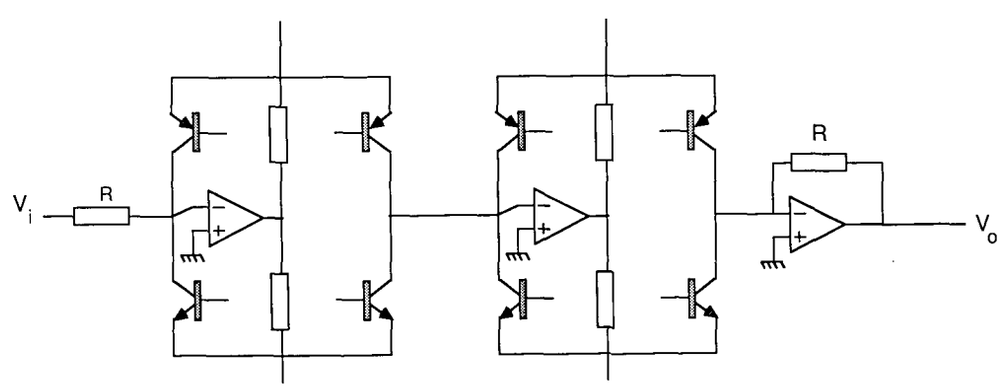


Fig. 14. Cell cascade using two dbx arrays.

APPENDIX
BASIC ADJUSTMENT PROCEDURE FOR TWO-STAGE VCA PROTOTYPE

1) Select minimum gain ($V_{control}$ = 0 V), SW_1 closed, SW_2 closed, VR_4, VR_5 midway. Adjust VR_1 to achieve $V_m = V_n$ at IC_3 outputs (T_{P1}, T_{P2}); then adjust VR_6 for zero output voltage (T_{P3}).

2) Select maximum gain, SW_1 closed, SW_2 open. Adjust VR_3 for $V_m = V_n$ at IC_3 outputs.

3) Select 12-dB attenuation. Adjust VR_7 for $V_m = V_n$ at IC_3 outputs.

4) Apply sine-wave input at maximum gain and adjust VR_2 to give zero ac output at x (T_{P3}).

Fine tuning of gain feedthrough over the range of $V_{control}$ can be achieved by iterative adjustment of VR_4, VR_5, VR_7, and the above procedure.

The circuit operates normally with both SW_1 and SW_2 open (where SW_1 selects an output servo to minimize dc offsets at the output, though it can be left closed if response down to dc is required). In practice, alignment can be minimized by the use of matched resistors (0.1% or better) and matched transistor arrays. The former was not used in prototype but is seen as an integral part of integrated circuit fabrication.

Dr. Hawksford's and Mr. Mills's biographies were published in the March issue.

Tandem Quadruplet VCA Topology

M.O.J.Hawksford
Dept. Electronic Systems Engineering
University of Essex

Tandem quadruplet VCA topology

M. O. J. Hawksford, University of Essex

0 Introduction

Voltage controlled amplifiers currently find a wide application in professional audio systems [1] where they function as remote controlled gain elements and a means of achieving programmable, parametric equalization. Most modern devices are based upon bipolar transistors [2 to 6] where the key element is the base emitter I_E/V_{BE} characteristic which for an ideal device adheres to the law,

$$V_{BE} = \frac{k\,T}{q} \ln\left[\frac{I_E}{I_S}\right] \qquad \ldots 1$$

where
- V_{BE}, base emitter voltage
- I_E, emitter current, ampere
- k, Boltzmann's constant, 1.38E-23 joule/kelvin
- T, absolute temperature, kelvin
- q, charge on electron, 1.602E-19 coulomb
- I_S, saturation current, ampere

VCA cells operate with matched transistors using either pairs or quadruplet arrays. Cells based upon pair matching, where the two series V_{BE} voltages add and are equated to a constant gain control voltage, produce a logarithmic gain law, while a 4 transistor cell where $\sum V_{BE} = 0$ produce a linear gain law with lower temperature sensitivity. Also, the incorporation of complementary working, offer significant topological advantage, by easing the design with respect to bias current absorption and also reducing the number of amplifiers required to both condition the input voltage and derive the output voltage, especially when operating in a single-ended mode as opposed to differential working.

In this paper, we explore a new cell topology based upon the complementary emitter follower configuration, where the parallel connection of base emitter junctions of NPN/PNP transistors can yield a near constant slope impedance with signal current over the class A operating range and can be programmed by modulating the cell bias current. Also, by cascading a number of such cells in series, the input voltage can be fully developed across silicon junctions without recourse to potential division via a resistor, as such higher signal to noise ratios are possible that do not require the adoption of class AB biassing together with the associated problem distortion and modulation noise, there is also lower sensitivity to noise in the biasing circuitry.

1 Parallel, complementary basis cell

The basis cell of the new VCA system is shown in Figure 1 and consists of a

voltage bias source V_B, and two complementary transistors T_1, T_2 configured in an emitter follower mode where, including the bias voltages, the base emitter junctions are effectively in parallel. We define v_c as the instantaneous voltage across the cell, i_c the output current and $I_B + i_c/2$, $I_B - i_c/2$ as the respective, instantaneous currents in T_1 and T_2. Although the signal current distribution between T_1 and T_2 is shown as symmetrical, the current I_B may also include a common mode component and thus effectively change with signal level from its quiescent value (ie $i_c = 0$) of I_{BO}. The cell transconductance is defined,

$$g_m = \frac{\partial i_c}{\partial v_c}$$

where our aim is to seek the best condition for constancy of g_m to enable the cell to be exercised over a maximum signal space. In passing, we note that the cell input current is effectively zero and that although the collector currents of T_1, T_2 include dynamic biasing (that is, $I_B = f(i_c)$), that the difference of the collector currents is equal to i_c.

System equations

applying Kirchhoff,

$$v_c = - V_B + V_{BE1}$$

$$v_c = V_B - V_{BE2}$$

and from equation 1,

$$V_{BE1} = \frac{k\,T}{q} \ln \left[I_B + \frac{i_c}{2} \right]$$

$$V_{BE2} = \frac{k\,T}{q} \ln \left[I_B - \frac{i_c}{2} \right]$$

where assuming transistors T_1 , T_2 are matched thermally and parametrically

$$2\,v_c = \frac{k\,T}{q} \ln \left[\frac{I_B + i_c/2}{I_B - i_c/2} \right] \qquad \ldots 3$$

$$\text{and } 2\,V_B = \frac{k\,T}{q} \ln \left[\frac{I_B^2 - i_c^2/4}{I_s^2} \right] \qquad \ldots 4$$

Forming the derivative $\partial i_c/\partial V_c$ from equation 3 and noting the definition for g_m in equation 2.

$$g_m = \frac{qI_B}{kT}\left[\frac{1 - \left(\frac{i_c}{2I_B}\right)^2}{1 - \frac{i_c}{I_B}\frac{\partial I_B}{\partial i_c}}\right] \quad \ldots 5$$

To explore the behaviour of g_m with i_c, consider the following biasing methods:

(i) Constant voltage biasing, V_B = constant

Consider the cell bias voltage V_B to be constant, where in equation 4, $\partial V_B/\partial i_c = 0$ which yields $\partial I_B/\partial i_c = 0.25 i_c/I_B$, thus equation 5 gives,

$$g_m = \frac{2q}{kT} I_B$$

However, assuming constant V_B and $I_B = I_{BO}$ when $i_c = 0$, then from equation 4, $I_B^2 - (i_c/2)^2 = I_{BO}^2$, giving

$$g_m = \frac{2q}{kT} I_{BO} (1 + (i_c/(2I_{BO}))^2)^{0.5}$$

If we define the quiescent transconductance (that is at $i_c = 0$) as g_{mo},

$$g_{mo} = \frac{2q}{kT} I_{BO} \quad \ldots 6$$

and a loading factor x as,

$$x = \frac{i_c}{2I_{BO}} \quad \ldots 7$$

$$g_m = g_{mo}(1 + x^2)^{0.5}$$

(ii) Constant current biasing, $I_B = I_{BO}$

Constant current biasing (that is I_B remains independant of i_c) can be achieved using the bias circuit in Figure 2 where diodes D_1, D_2 have identical characteristics to transistors T_1, T_2 and are thermally matched; this also has the desirable characteristic of eliminating the thermal dependence on I_s and T as expressed in equation 4. Assume that the diodes are driven by differential currents $I_{BO} + i_c/2$, $I_{BO} - i_c/2$, whereby

$$V_B = \frac{k\,T}{q} \ln \left[\frac{I_{BO}^2 - i_c^2/4}{I_s^2} \right]$$

hence equating with V_B from equation 4, $I_B = I_{BO}$. Since I_B is now constant, $\partial I_B / \partial i_c = 0$, where from equation 5, 6 and 7

$$g_m = g_{mo} \left(1 - x^2\right)$$

Once more the cell exhibits non linearity, but a comparison with g_m for constant voltage biasing, now shows the distortion is in the opposite sense.

(iii) Intermediate biasing

The findings in (i) and (ii) suggest a more optimum performance by seeking an intermediate biasing condition between the two extrema, hence we modify the current drive in Figure 2 to $I_{BO} + \alpha\, i_c/2$, $I_{BO} - \alpha\, i_c/2$ where α is a constant, giving

$$V_B = \frac{k\,T}{q} \ln \left[\frac{I_{BO}^2 - \left(\alpha\, i_c/2\right)^2}{I_S^2} \right]$$

then again equating V_B with equation 4,

$$I_B^2 = I_{BO}^2 - (1 - \alpha^2)\,(i_c/2)^2$$

Differentiating and substituting into equation 7, observing the definition of equation 6, 7 yields

$$g_m = g_{mo}(1 - \alpha^2 x^2)(1 + (1 - \alpha^2)x^2)^{1/2} \qquad 8$$

This result is a compromise solution between the constant voltage and constant current bias modes, where for illustration, we define a distortion factor $D(x, \alpha)$,

$$g_m = g_{mo}(1 + D(x, \alpha))$$

whereby,

$$D(x, \alpha) = (1 - \alpha^2 x^2)(1 + (1 - \alpha^2)x^2)^{1/2} - 1 \qquad \ldots 9$$

In Figure 3, a family of $D(x, \alpha)$ is plotted against x ranging from 0 to 1 and α ranging from 0.5 to 0.6 in steps of 0.05, where acceptable linearity is demonstrated for $\alpha \approx 0.65$.

2 Error correction for transistor bulk resistance

VCA cells that use bipolar transistors require the individual devices to adhere to an accurate logarithmic relationship where the transistor should exhibit good *logarithmic conformity* . Appropriate transistor selection and choice of bias

currents together with the minimisation of signal dependent changes in collector to base voltage (to eliminate contributions from slope impedances) can improve logarithmic conformity, although the ohmic voltage drops associated with emitter and base bulk resistances set an ultimate bound on performance. The effect of bulk resistances can be represented by modifying equation 1 as follows

$$V_{BE} = \frac{kT}{q} \ln\left[\frac{I_E}{I_S}\right] + I_E\left[r_{ee'} + \frac{r_{bb'}}{1+\beta}\right] \quad \ldots 10$$

where the base bulk resistances $r_{bb'}$ can be reflected to the emitter using the I_C / I_B current gain β and combined with the emitter bulk resistance $r_{ee'}$ to form a composite resistance r_{bk},

$$r_{bk} = r_{ee'} + \frac{r_{bb'}}{1+\beta} \quad \ldots 11$$

The key to correction for the effects of bulk resistance [6] is to introduce a voltage proportional to emitter current that is appropriately scaled and of opposite polarity to the internal ohmic voltage drops.

8 transistor error correction cell

The basic cell of Figure 1 can be modified by introducing a further 2 transistors in series with T_1, T_2 and then compounding this structure with a similar cascaded circuit as shown in Figure 4. If we neglect the contribution from base currents and assume all transistors to be matched, then for top half of circuit,

$$v_c + 2V_B - V_{BE1} - 2\left(I_B + \frac{i_c}{2}\right) r_{bk} - V_{BE3} - \left(I_B - \frac{i_c}{2}\right) r_x$$

$$+ \left(I_B + \frac{i_c}{2}\right) r_x + V_{BE7} + 2\left(I_B - \frac{i_c}{2}\right) r_{bk} + V_{BE5} - 2V_B = v_c$$

hence if, $r_x = 2\, r_{bk}$ and $V_{BE1} = V_{BE3}$, $V_{BE5} = V_{BE7}$

then $v_c = V_{BE1} - V_{BE5}$

a similar analysis can be performed for the lower half of the circuit where the optimum selection of r_x can again desensitise performance to the contribution of bulk resistance. The bias voltages V_B are multiplied by 2, but the circuit of Figure 2 is readily modified to accommodate this factor.

3 Tandem connection of follower cells

Section 1 and 2 have demonstrate the means by which a complementary follower configuration can be approximately linearized to have a near constant transconductance with signal current, where defining a current I_{BO} in a bias circuit with signal current, voltage V_B can be produced which determines the cell transconductance to be proportional to the current I_{BO}.

However, as with all bipolar cells, the input voltage range is restricted which implies a limitation in signal to noise ratio (SNR). In this section we investigate a tandem connection where a number of cells N are connected in cascade such that each cell sees only a fraction of the total input voltage, which can now be of greater level. A basic structure is shown in Figure 5 where N pairs of cells are used. The input signal V_{in} is divided by an equi-value resistor network to produce intermediate voltages to which are added the appropriate bias voltages V_B using unity-gain summing amplifiers. We may therefore assume that with appropriate transistor matching, that the resistance of each cell r_c is identical and that each carries a nominal signal current i_c. A positive attribute of this configuration is the insensitivity to noise components appearing on the intermediate driving nodes of each quadruplet, as signals at these nodes introduce equal and opposite signal currents in adjacent cells such that when currents are summed, the differential errors cancel. Thus, the noise contribution of each intermediate unity-gain amplifier introduced between resistor divider networks and each cell is minimised.

Signal to noise ratio

Assume that each cell can accommodate a peak signal level $\hat{v}_c$ where $\hat{v}_c = I_{BO}/g_{mo} = kT/q$, ($\approx 25mV$) before the onset of unacceptable distortion, and that at a particular state of bias, the total cell noise source has a mean square value $\overline{e_c^2}$ in series with the cell resistance r_c (where $r_c = 1/g_m$).

$$\therefore \text{ Peak SNR (1 cell)} = \left(\frac{\hat{v}_c}{r_c}\right)^2 \left(\frac{r_c^2}{\overline{e_c^2}}\right) = \frac{\hat{v}_c^2}{\overline{e_c^2}}$$

If there are N cells in series, the peak input voltage is now $N\hat{V}_c$ and if all cell currents are summed, the peak signal current is $N\hat{V}_c/r_c$. However, the components of noise currents add on a *power basis*, therefore

$$\text{mean square noise current} - N\frac{\overline{e_c^2}}{r_c^2}$$

whereby, the total SNR becomes

$$SNR_{(N\ cells)} = N\frac{\hat{v}_c^2}{\overline{e_c^2}}$$

However, there is an upper bound to the number of cascaded cells that is determined by the peak input signal $\hat{V}_{in}$, where

$$N_{opt} = \text{int}\left[\frac{\hat{V}_{in}}{\hat{V}_c}\right]$$

If the optimum number N_{opt} is exceeded, then the loading on each cell is correspondingly reduced, reducing both distortion *and* SNR. It is also expedient to compare the performance of the cascaded cell structure with that of a parallel connection of N cells as shown in Figure 6. In fact, both cell configurations will offer potentially the *same* SNR providing we assume the associated biasing sources and amplifiers to be noiseless. However, when we introduce noise sources into the biasing amplifiers (that includes the noise from V_B), then the following points should be noted:

(i) In the parallel connection, each cell sees the same external noise source developed across the cell resistance r_c, therefore through coherent addition and the effective system resistance r_c/N, the external noise sources undergo significant amplification.

(ii) In the series connection, the outputs of the intermediate amplifier ($x_2, x_3, x_4; y_2, y_3, y_4$) when summed via the individual assumed identical cells, cancel as any two adjacent amplifiers are driven in an opposite sense by the noise component e_n.

(iii) In the series connection, it is only the external noise sources presented at the nodes x_1, x_5 and y_1, y_5 which contribute to a net output current, but these are produced across an effective resistance r_c of the first cell and similarly for the N^{th} cell.

(iv) In the series connection, although the intermediate voltages do not contribute directly to the output, they are necessary to steer each cell into its optimum operating range to minimise distortion.

4 Conclusion

The paper has introduced a quadruplet cell that functions similarly to an emitter follower. A theoretical study demonstrated that by using an optimized, dynamic biasing system that the complementary NPN/PNP parallel connection of transistor could yield improved linearity compared with either constant voltage or constant current biasing. The cell offers a linear gain law where transconductance is directly proportional to a current I_{BO} applied to the bias circuit, which also offers the advantage of a low temperature sensitivity providing the cells use matched transistors under isothermal operation.

A tandem connection of identical cells was shown to give an improved SNR and offer (compared with a parallel connection) lower sensitivity to noise in associated amplifiers and bias circuits. Although the present paper has not presented complete circuitry, the system approach introduced sufficient detail to demonstrate the feasibility of a VCA using the follower cell, where in particular the role of error correction for ohmic voltage drops in the bulk resistance of transistor was discussed thus allowing each transistor to achieve more accurate logarithmic conformity. Similarly, when implementing a cell, circuitry should be used to maintain both a low and a constant collector base voltage and thus minimise distortion products from transistor slope impedances.

References

1 Duncan, B.,"VCA's investigated", Studio Sound, vol 31, nos 6, 7, 8, 9, pt 1 June, pt 2 July, pt 3 August, pt 4 September 1989

2 Gilbert B., "Translinear circuits: A proposed classification", IEE Electronic. Letters, vol.11. pp.14-16, January 9, 1975

3 Gilbert, B., "A new technique for analog multiplication", IEEE Journal of Solid State Circuits, vol.SC-10, no.6, December 1975, pp.437-447

4 Baskind, D. and Harvey, R., "Techniques for the realisation and application of voltage controlled amplifiers and attenuators", 60th Convention of the AES, May 1978, preprint 1378 (A-6)

5 Baskind, D., Rubens, H., Kelson, G., "The design and integration of a high performance voltage controlled attenuator", Convention of the AES, November 1979, preprint 1555 (D-2)

6 Hawksford, M.J., "Topological enhancements of translinear two-quadrant gain cells", JAES, vol. 37, no. 6, June 1989, pp. 465-475

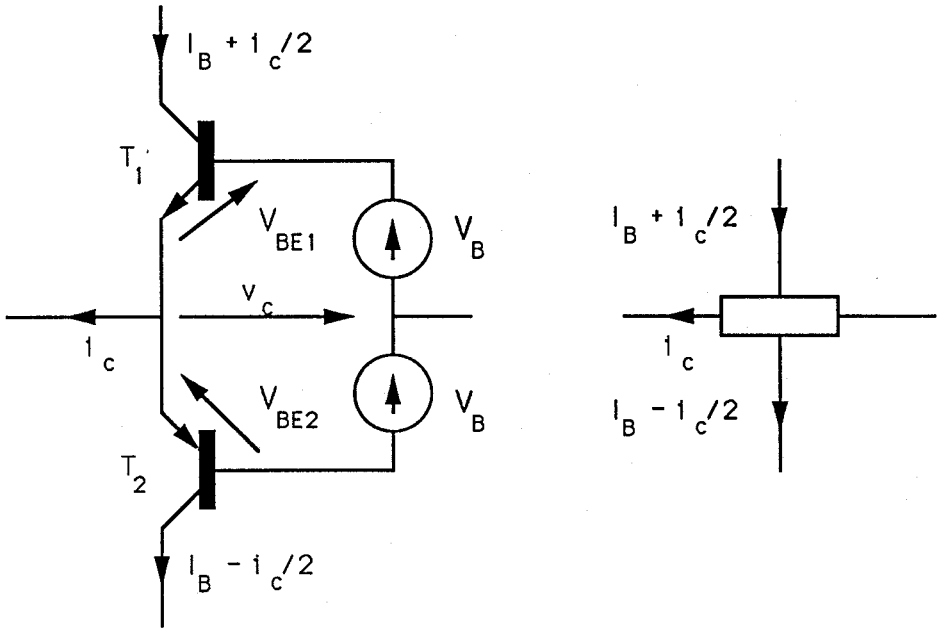


Fig. 1 Basic cell formed from complementary emitter follower

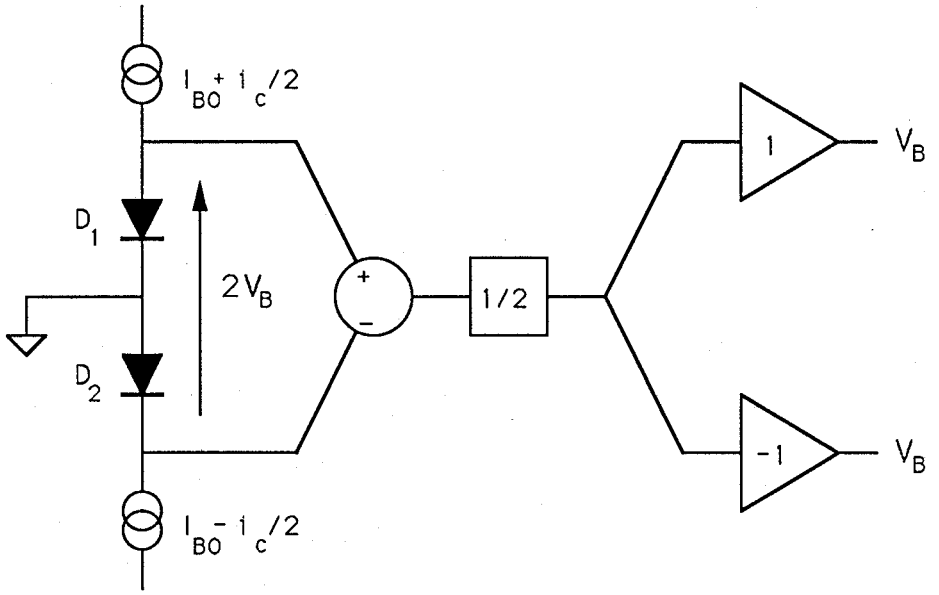


Fig. 2 Derivation of dynamic bias voltage, V_B.

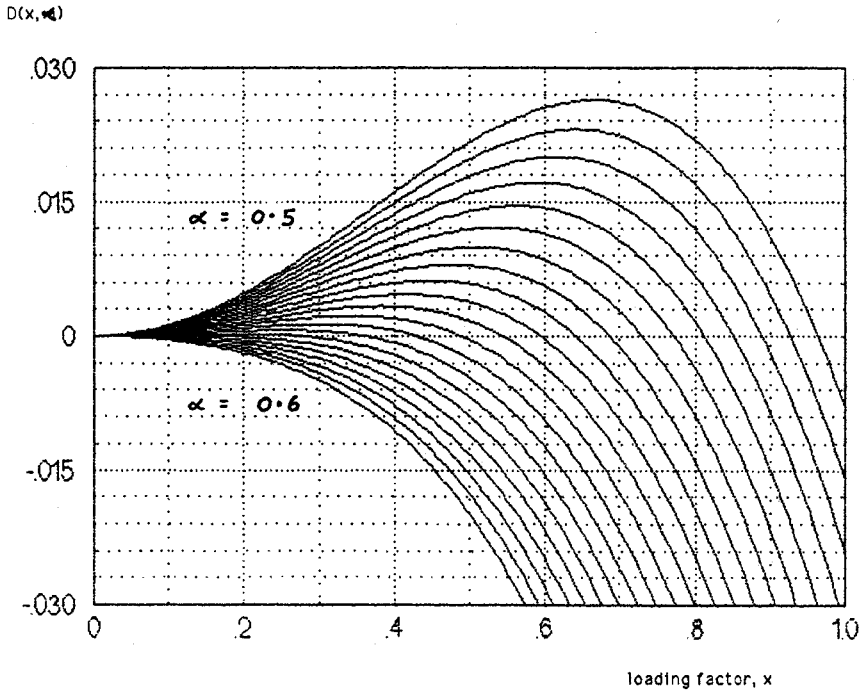


Fig. 3 Family of distortion plots D(x,∝) against cell loading factor, x; for ∝ = 0.5 to 0.6 in steps of 0.05 .

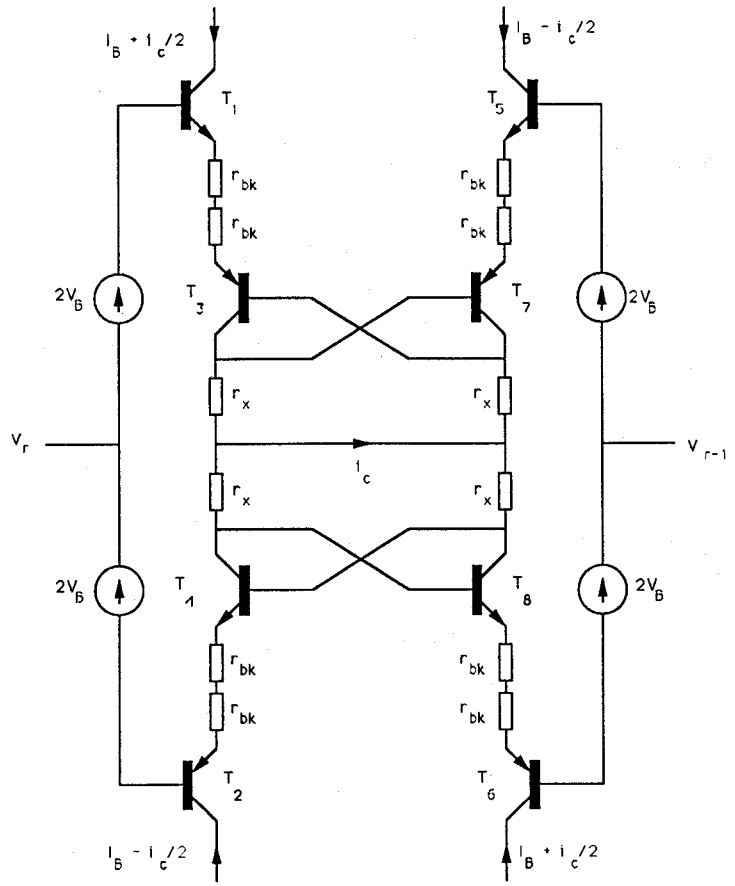


Fig. 4 Cascaded cell with error correction to compensate for base and emitter bulk resistance.

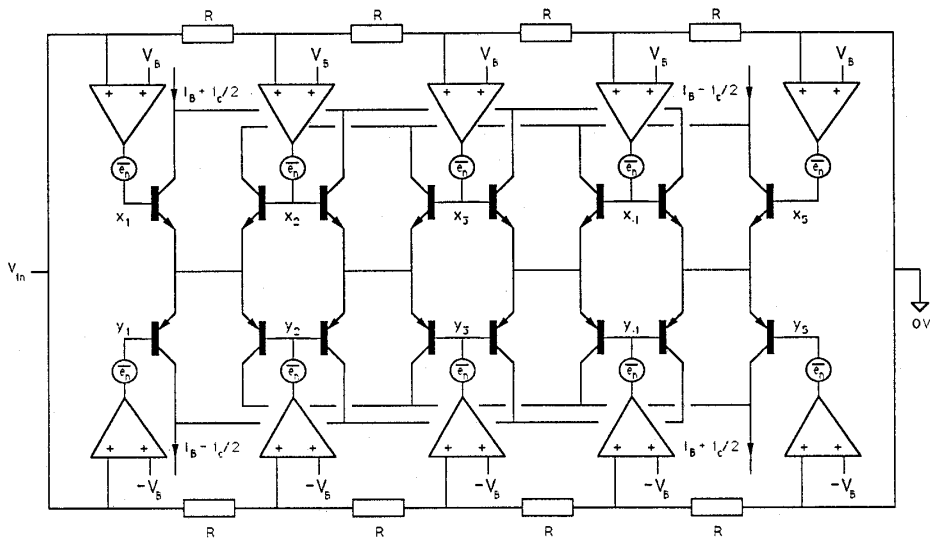


Fig. 5 Series connection of 4 quadruplet gain cells.

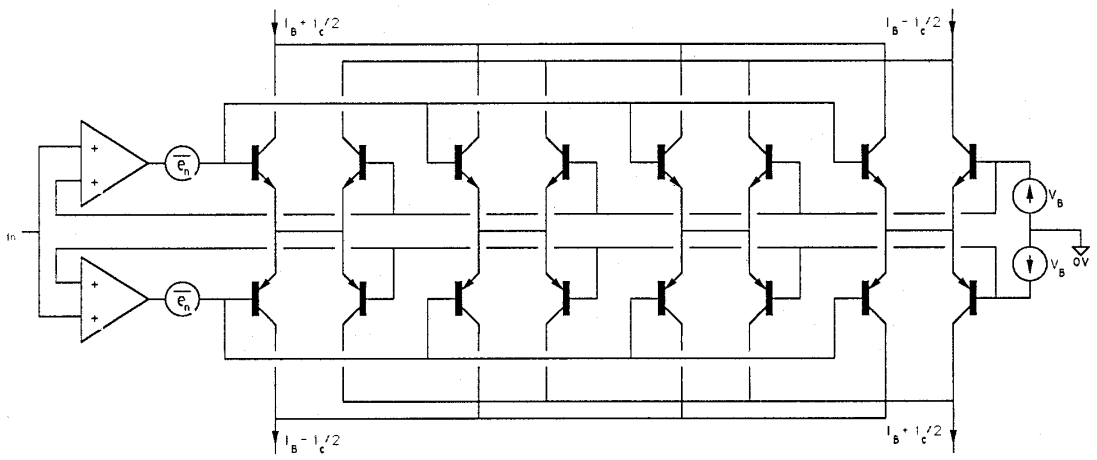


Fig. 6 Parallel connection of 4 quadruplet gain cells.

Voltage-controlled amplifier systems

Malcolm Hawksford looks at the design of audio systems using voltage controlled amplifiers.

The voltage controlled amplifier (VCA) is widely accepted in audio systems, particularly in mixing consoles where it enables remote control of both gain functions and more specialised parametric equalizers. Its advantage is that it enables signals to be processed in an essentially analogue environment but has no need for long signal routing to control surfaces.

Computer systems can be overlaid to yield sophisticated control and function automation while allowing signals to remain in the analogue domain. Such considerations have formed the principal thrust for refinement of circuit topology. Lucrative studio applications, each of which may be using 100 or more programmable gain elements, have made the VCA a front line target for competitive manufacture.

In professional audio, the specification required of the VCA is considerable: noise, bandwidth and distortion should be comparable with the best operational amplifiers yet the VCA should offer the added function of gain control.

It should also satisfy the need for accurate calibration, repeatability and temperature insensitivity.

We shall consider only electronic solutions based upon bipolar technology, as these techniques are highly refined. Motorised and precision-switched networks, although well matched on performance grounds, are often too expensive and are space-inefficient for many applications.

A basic, though non-ideal, VCA cell is shown in **Fig. 1**. The circuit consists of two transistors in a differential, long-tail pair configuration where the gain is controlled by the tail current $2I_E$. The input signal V_i is attenuated by the resistor network R_1, R_2 to produce an inter-base voltage V_{EB} that is restricted to a few millivolt. Inevitably this limits the signal to noise ratio (SNR). The output voltage V_o is derived using a differential trans-resistance amplifier operating on the collector currents of T_1, T_2 such that

$$V_o = R(I_{c1} - I_{c2}) = 2RI$$

In this example, I is the collector signal current that also circulates through each base emitter junction, where the difference process, in association with two essentially identical but differentially-driven transistors (T_1, T_2) is fundamental to many VCA circuits.

At the heart of the VCA system design equations is the logarithmic relationship

between emitter current I_E and base emitter voltage V_{BE}.

$$V_{BE} = \frac{kT}{q} \ln\left[\frac{I_E}{I_s}\right] \quad \ldots 1$$

where V_{BE} is base emitter voltage, I_E is emitter current; k is Boltzmann's constant, 1.38E-23J/Kelvin; T is temperature Kelvin; q is charge on an electron, 1.602E-19 coulomb, and I_s is saturation current.

Although this equation excludes several parameters found in more general purpose transistor models[1], it nevertheless embraces the target to which a transistor should adhere if it is to be a successful candidate for a voltage controlled amplifer.

In other words, the logarithmic law is the key [illegible] metric element.

In practice, although equation 1 is a simplification, it can be approximated with suitable transistor selection, appropriate biasing and by maintaining the collector-base voltage at a low but constant voltage to minimise the contribution from slope impedances[2].

The basic cell of Fig. 1 includes a number of structural details necessary for a VCA.

In its present mode of operation, it is not yet an acceptable candidate, even though it can offer a basic understanding of VCA operation. A positive attribute of the cell is the output difference circuit which implies that the output signal is a function of the difference of V_{BE1} and V_{BE2}.

This feature not only compensates for the base emitter offset voltage of a single transistor but, for perfectly matched transistors, it eliminates the dependence upon the saturation current I_S which is a highly temperature sensitive transistor parameter.

Using equation 1 and assuming the same I_S for [illegible] T_1 and T_2,

$$V_{BE1} - V_{BE2} = \frac{KT}{q} \ln \frac{I_{E1}}{I_{E2}} \quad \ldots 2$$

The two-transistor cell remains temperature-dependent as T (in degrees Kelvin) appears in the equation; however, it is orders of magnitude less problematic than the temperature variation with I_s. The incorporation of transistors that are both parametrically and thermally matched is therefore the second key to VCA design. It necessitates the use of monolithic arrays using transistors that have accurate geometric tolerances and device proximity, fabricated on a common substrate to aid isothermal operation.

Unfortunately, when the circuit of Fig. 1 is driven from the base as a differential amplifier, such that $I_{E1} = I_E + i$ and $I_{E2} = I_E - i$, equation 2 shows a non-linear relationship between the output signal current i and the inter-base voltage $V_{BE2} - V_{BE2}$. Indeed, if re_1, re_2 are the small signal V_{BE}/I_E slope resistances of T_1, T_2, then using transistor amplifier theory[3], we can show

$$V_{BE1} - V_{BE2} = i(r_{e1} + r_{e2})$$

where

$$r_{e1} = \frac{kT}{q(I_E + i)}, r_{e2} = \frac{kT}{q(I_E - i)} \quad \ldots 3$$

that is, the cell transconductance g_m only becomes linear for $i << I_E$,

$$g_m = \frac{2q}{kT} I_E$$

which is a severe restriction on dynamic range. The small signal expression for g_m reveals a gain proportional to I_E and since $I_E > 0$ the cell of Fig. 1 can form the basis of a two-quadrant multiplier. Thus by controlling the tail current, appropriate I_E/V_{BE} slope resistances for T_{E1}, T_2 are established which programs the cell transconductance.

However, for a successful VCA application, the problem of non-linearity must be addressed in order to attain both acceptable distortion performance an[illegible] SNR. A number of topological variants will be presented, which yield the following family of gain laws:

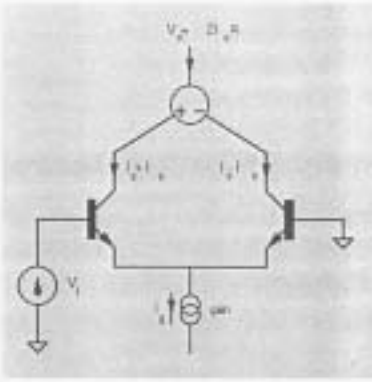
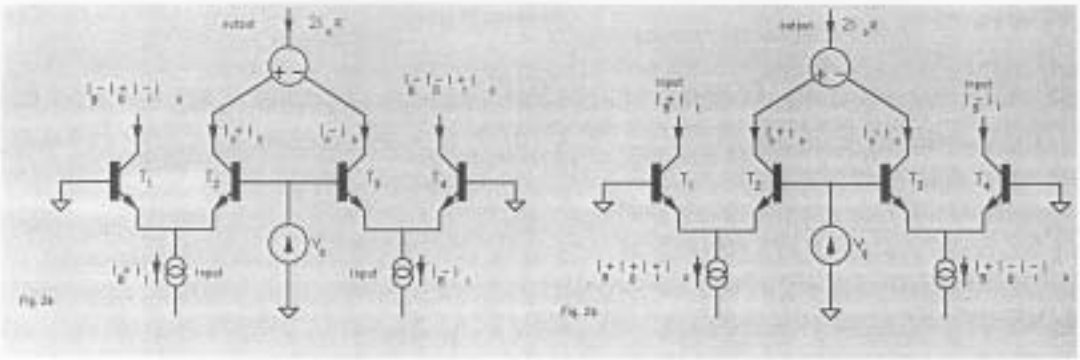


Fig. 1. Basic 'long-tailed pair' two-transistor gain circuit. The circular symbol in the collector circuit represents a differential trans-resistance amplifier in which the output voltage responds to the difference in input current.

(i) Pseudo-logarithmic gain law, but with a maximum gain of unity;
(ii) True logarithmic gain law with capability of gain > 1;
(iii) Linear gain law but with a maximum gain of unity;
(iv) True linear gain law with capability of gain > 1.

It will be shown that groups (i) and (ii) are variants on Fig. 1 but incorporate dual, 2 transistors cells where the base emitter voltages of each pair sum to a constant voltage, while categories (iii) and (iv) use 4 transistors with $V_{BE} = 0$ yielding theoretical temperature independence. We shall also show how a fully complementary topology elegantly solves the

Fig. 2. Two differentially-driven current division cells with input applied to emitters (Fig. 2a) and collectors (Fig. 2b).



bias absorption problem and simplifies the design of both input and output circuitry.

Two-transistor basis cells

Although the topology of Fig. 1 is non-linear, it can be linearised by exchanging the roles of signal and gain control. If the inter-base voltage is equated to a gain control voltage V_g such that

$V_g = V_{BE1} - V_{BE2}$.

then rearranging equation 2,

$$\frac{I_{E1}}{I_{E2}} = e^{\frac{qV_g}{kT}} = \gamma$$

where γ is a gain parameter incorporating an exponential law. This is a useful result especially where a decibel gain control law is required, such that if

Gain (dB) = $20\log_{10}(\gamma)$

then,

$$\text{Gain (dB)} = \left[\frac{20q}{kT\ln(10)}\right]V_g$$

The expression for γ shows that a constant inter-base voltage (and strictly constant temperature) produces a constant current division ratio between T_1 and T_2, even for large input signals, providing the transistors exhibit good I_E/V_{BE} logarithmic conformity over the working range of emitter currents.

Figure 2 illustrates the two methods of incorporating this constant current division ratio into a VCA that depends upon the entry node for the input current.

However, to enable bipolar operation (as $I_i > 0$), two system extensions are required: firstly, a constant bias current I_B is superimposed upon the input signal current to enable the signal component to go negative and secondly, an additional, identical two-transistor cell is incorporated again with a bias current I_B but now with the signal current of inverted polarity.

The output signal is then derived as a function of the gain cell output currents using a differential, transresistance amplifier, which simultaneously cancels the output bias components and also prevents gain control feedthrough under perfect symmetry.

Thus using the current annotation in Fig. 2 and the result $I_{E1}/I_{E2} = \gamma$, then, for Fig. 2a,

$$\frac{i_o}{i_i} = \frac{2\gamma}{1+\gamma}$$

and for Fig. 2b

$$\frac{i_o}{i_i} = \gamma \qquad \text{(true logarithmic law)}$$

The gain cell of Fig. 2a produces a pseudo logarithmic law that approaches a maximum gain of unity; indeed observation of the circuit shows that as V_g increases, T_2 and T_3 become common-base stages offering excellent bandwidth and distortion characteristics.

The Fig. 2b cell results in a true logarithmic law together with current gain > 1, although there is an additional complication of forcing the collector currents of T_1, T_4 to equal the input currents $I_B + i_i$, $I_B - i_i$. This can be solved using the arrangement of Fig. 2c, where a precision operational amplifier is included in a feedback system to adjust the tail current until the desired collector current is established.

Four-transistor basis cells

The use of four transistor cells where $\Sigma V_{BE} = 0$ enables the kT/q term to cancel and the cell to become theoretically temperature-independent. The gain law is also transposed from logarithmic to linear.

As an example, a new system arrangement is shown in **Fig. 3** where the gain control signal is now a current I_g. Using the result $V_{BE1}-V_{BE2}+V_{BE3}-V_{BE4} = 0$, and noting the annotation of transistor currents, it can be shown that[4]

$$\frac{i_o}{i_i} = 1 - \frac{I_g}{2I_B}$$

Although the gain is proportional to I_g there is a restriction that $0 < I_g < 2I_B$ as the maximum gain is unity, whereon transistors T_1, T_4, as with the Fig. 2a topology, operate as grounded base stages and yield excellent linearity.

An alternative four-transistor cell of low temperature sensitivity[5, 6, 7, 8, 9], again uses the principle that $\Sigma V_{BE} = 0$, but allows a current gain > 1. This cell can be considered an extension of the topology in Fig. 1 in which the additional T_3, T_4 transistors effectively predistort the input signal applied to T_1, T_2 and achieve overall linearity.

Although the process of linearization can be explained by considering the interplay of the transistor I_E/V_{BE} slope resistances, it is simpler to demonstrate by analysis. From the condition

$V_{BE1} - V_{BE2} + V_{BE3} - V_{BE4} = 0$

and using equation 1,

$$\frac{I_{E1}}{I_{E2}} = \frac{I_{E4}}{I_{E3}}$$

when appropriate currents are substituted, it follows that

$$\frac{i_o}{i_i} = \frac{I_g}{I_B}$$

Cell noise performance

Although this has demonstrated how a gain cell can be linearized to operate over a wide range of signal currents, of equal importance is the cell noise performance. Ideally this should approach that of the best operational amplifiers and not exhibit significant noise modulation artefacts with either gain control or signal level.

Noise components in the output signal arise both from the gain cell and associated signal conditioning circuitry. Examination of all the topologies presented reveals at their core a differential amplifier with no local emitter degeneration.

This means that the internal noise sources of each transistor can experience significant amplification.

Because the noise amplification factor as well as the actual noise sources depend upon the gain and state of cell biasing, it is common for the output noise to vary as a function of gain and signal level. This produces modulation noise, where the maximum output noise may not occur at maximum gain, depending upon topology (in practice, cells that offer a maximum current gain of unity also offer low noise at unity gain).

The linearization of the gain cell is crucial in obtaining an acceptable signal to noise ratio (SNR), where it is expedient to select transistors not only with respect to parametric matching but also for noise performance.

However, the noise contribution from associated circuitry should not be neglected as it

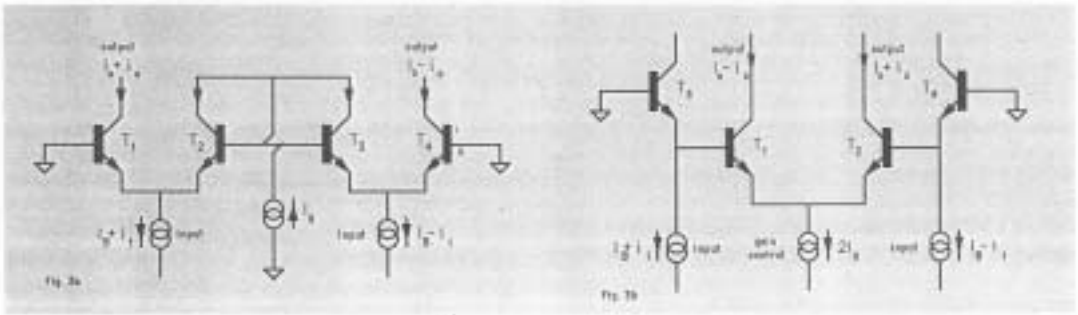


Fig. 3a. Current steering gain cell, after Hawksford[4, 13].
Fig. 3b. Four-transistor cell after Gilbert[5, 6], with linear gain law and current gain capability >1.

Fig.4. Differential trans-resistance output stage.

can often represent a limit on the absolute noise performance, especially at certain strategic gain settings.

There is not space in this present article to do justice to the subject of VCA noise, but an example may illustrate some of the design problems typically encountered. VCA circuitry can be sub-divided into four principal areas:

(i) input trans-conductance amplifier, often with single ended to differential transformation;

(ii) gain cell topology;

(iii) output differential trans-resistance amplifier;

(iv) gain control signal conditioning.

Consider the differential output stage shown in Fig. 4 that incorporates three amplifiers where each introduces noise. IC_1, IC_2 are trans-resistance stages that convert the gain cell output currents to voltages while simultaneously maintaining a well-defined collector voltage. Resistor R_3 must convey a current up to $2I_0$, therefore a minimum positive supply $> 2\ I_0R_3$ is required. If IC_1, IC_2 have negligible input noise currents, then the mean-square output noise voltage from this cell is $e_n^2 + 4kTR_3F$ and the signal power is $(i_0R)^2$. The best SNR is achieved therefore when $4kTR_3F > e_n^2$.

The second stage IC_3 converts the differential signal from IC_1, IC_2 to single-ended mode, where if the differential gain = 1 (ie $R_2 = R_1$) then the output noise from this amplifier is 6dB above the input noise of IC_3. In the context of a unity gain VCA, this can represent an unacceptable reduction in noise performance.

There are clear problems associated with using too many cascaded amplifier stages including that of absorbing the gain cell output bias currents. These difficulties also extend to the input stage, especially if the more typical, single-ended mode of operation is required.

All of the circuits so far considered are non-complementary and require asymmetric biasing systems to function. The cells must be suspended in a well-defined electrical environment both to maintain appropriate biasing and to minimise collector-base voltages and their variations.

This reduces junction heating and prevents the slope impedance transistor parameters from introducing distortion[3].

Because of this, it is common practice to make $VCB = 0V$ and to strive to maintain this voltage constant with signal current variations. For example, the use of collector resistors on the gain cell output transistors would be poor practice since it would allow large changes in collector-base voltages to occur.

One solution to these difficulties is to introduce a complementary topology using accurately matched NPN and PNP transistors fabricated on a common substrate. A configuration pioneered by dbx[10, 11] is shown in Fig. 5.

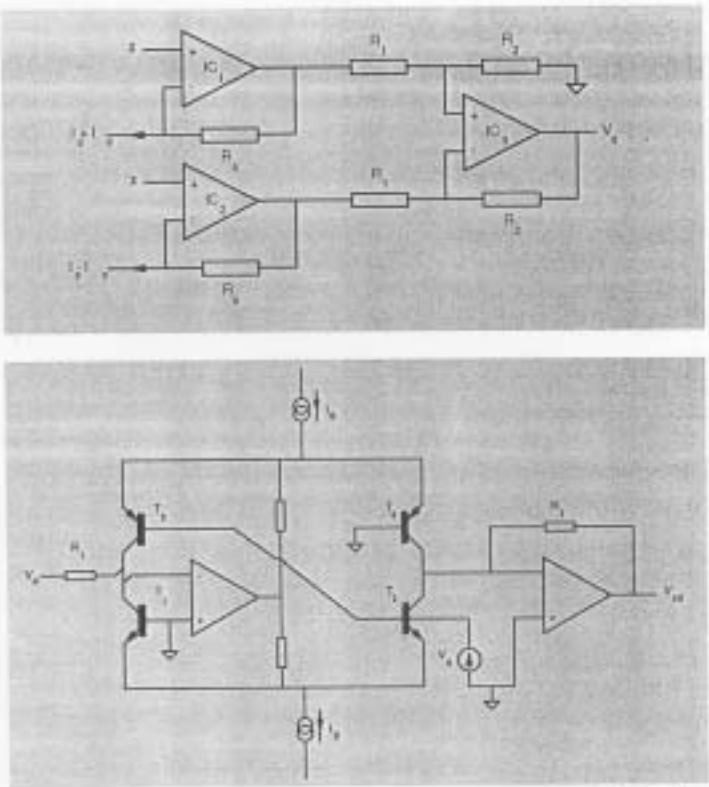


Fig.5. Basic dbx[10, 11] gain cell with support circuitry.

To understand the operation of this circuit, compare the operation of T_1, T_2 with that of T_1, T_2 in Fig. 2a which yielded a logarithmic gain law. Then, turn the T_1, T_2 circuit upside down and substitute PNP devices, thus forming a similar but complementary circuit which can now absorb the bias currents from T_1, T_2 as the biasing of T_3, T_4 and of T_1, T_2 are essentially the same.

The input transconductance amplifier IC_1, by its virtual earth operation, locates the collectors of T_1, T_3 near zero volt as does the transresistance amplifier IC_2 for the collectors of T_2, T_4. Signal current entering R_1 is then steered via feedback control symmetrically into the collectors T_1, T_3 while output current from T_2, T_4 flows via the trans-resistance amplifier IC_2 into R_2.

The control voltage V_g, which is applied to both T_1, T_2 and T_3, T_4 transistor pairs, controls the current division between collectors T_1, T_3 and T_2, T_4, which determines the system gain. The overall VCA gain then becomes,

$$\frac{V_o}{V_i} = \frac{R_2}{R_1}T$$

The topology of Fig. 5 is most elegant and neatly solves many associated VCA problems while reducing the number of cascaded operational amplifiers to two.

If differential operation is required, two separate VCA systems could be used, though gain tracking is paramount. The SNR depends mainly on the gain cell and cell loading factor (that is, the ratio of signal current to bias current[12]) and amplifier noise and reaching a compromise between cell loading and levels of distortion.

Operation can be enhanced by incorporating a number of cells in parallel, although this practice is similar to increasing the junction area.

However, a main performance limitation is the bulk (ohmic) resistance associated with the cell transistors. These have greatest effect at high collector currents where the base emitter slope impedances (r_e) become small.

Bulk resistances are a primary source of non-linearity in gain cells, where the presence of both emitter and base bulk resistance degrades logarithmic conformity by adding an ohmic related voltage in series with the ideal logarithmic device. The main VCA cell equation then becomes:

$$V_{BE1} - V_{BE2} = \frac{kT}{q}\ln\left[\frac{I_{E1}}{I_{E2}}\right] + (I_{E1} - I_{E2})r_{bulk}$$

Consequently, in a current division cell, the

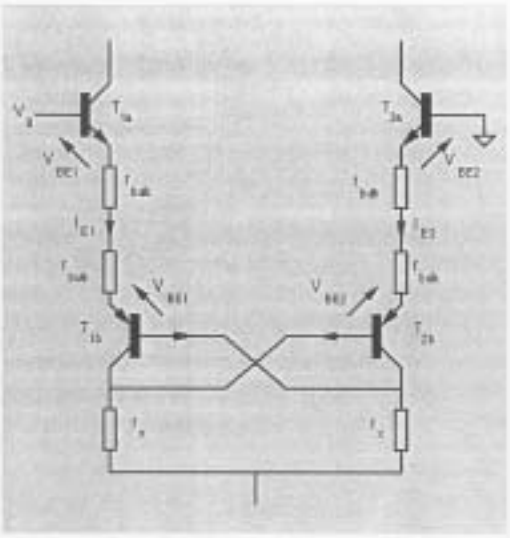


Fig.6. Modified current division cells with bulk resistance error correction circuitry. All transistors are assumed to have parametrically and thermally identical β and to have ic/ib current gains.

term $\ln[I_{E1}/I_{E2}]$ is no longer held constant and the current division ratio becomes signal dependent, thus introducing distortion.

However, a method of distortion correction exists where a correction signal V is added to the inter-base voltage to compensate for the term $(I_{E1}-I_{E2})r_{bulk}$. The exact nature of the correction depends upon topology but can in principle be introduced to all the gain cells considered here.

Thus by the dual strategy of selecting transistor arrays to have low bulk resistance (possibly by using parallel transistors) and by using 'ohmic-loss error correction', the inherent distortion can be maintained at low level over a wide signal and gain control range.

To illustrate one example of a distortion correction topology, consider an adaptation of the basic current division cell of Fig. 2a, as shown in **Fig. 6**.

The modified topology includes transistor bulk resistances r_{bulk} and assumes all transistors to be matched and NPN/PNP devices to be complementary. The means of cancelling the effect of bulk resistance can be demonstrated by applying Kirchhoff's voltage law, whereby

$$V_g - V_{BE1} - 2I_{E1}r_{bulk} - V_{BE1} - \left[I_{E2} - \frac{I_{E2}}{1+\beta} + \frac{I_{E1}}{1+\beta}\right]r_c + \left[I_{E1} - \frac{I_{E1}}{1+\beta} + \frac{I_{E2}}{1+\beta}\right]r_c + V_{BE2} + 2I_{E2}r_{bulk} + V_{BE2} = 0$$

Hence, if the collector resistance r_c is given by

$$r_c = \frac{2r_{bulk}}{1 - \frac{2}{1+\beta}}$$

then

$$V_g = I\,(V_{BE1} - V_{BE2}).$$

Thus, by cancelling the ohmic voltage drops, the internal base emitter voltages become directly equated to the gain control voltage V_g, although we now observe a factor of 2 as the base emitter junctions of $T_{1a,b}$ and $T_{2a,b}$ are in series.

The modified topology can readily be applied to other forms of gain cell. It would be particularly appropriate in complementary guise for inclusions in the dbx cell of Fig. 5. Alternatively, a weighted sum of input and output signals can be superimposed upon the gain control signal[13], allowing the ohmic related voltage drops to be cancelled.

Conclusion

Our discussion has considered a number of factors encountered in the design of a VCA, where the importance of linearity and low noise have been emphasised. We have shown that complementary topologies and error correction for the effects of bulk resistances are also effective tools in furthering the art.

In practice, most bipolar VCAs are variants on the type of topologies presented here where, for example, so called 'class AB biasing' is used with the dbx complementary type of circuit. Such techniques increase distortion but can reduce noise, especially at low signal level.

The maintenance of low distortion at high attenuation is also important where performance can degrade due to transistor current starvation and common mode signal components within a differential input stage[15]. Indeed, a two-stage current steering topology has been proposed where an experimental circuit topology is presented[15]. Because of its non-complementary structure it is also a good illustration of the problems encountered with power supply design and the maintenance of low noise performance.

Most of the problems in VCA design are now well understood and modern fabrication can yield low cost devices of acceptable performance. However, the balance of technology is rapidly changing with the advent of sophisticated and affordable digital processing. Apart from some specialised equipment, the day of the analogue VCA may soon pass into the history books. Even the most sophisticated devices still have, and will probably continue to have, measurable deficiencies, as exposed in Ben Duncan's acclaimed series in Studio Sound[14]. ■

References

1. Gross, I.E., *Modeling the bipolar transistor*, Tektronix, Inc.
2. Hawksford, M.J., *Reduction of transistor slope impedance dependent distortion in large-signal amplifiers*, JAES, vol.36, no.4, April 1988, pp.213-222.
3. Ritchie, G.R., *Transistor circuit techniques: discrete and integrated*, Van Nostrand Reinhold, ISBN 0-442-30533-8, 1985.
4. Hawksford, M.J., *Low-distortion programmable gain cell using current-steering cascode topology*, JAES, vol.30, no.11, November 1982, pp.795-799.
5. Gilbert, B., *A precise 4-quadrant multiplier with subnanosecond response*, JSCC, vol.SC-3, no.4, December 1968, pp.365-373.
6. Gilbert, B., *A high-performance monolithic multiplier using active feedback*, JSCC, vol.SC-9, no.6, December 1974, pp.364-373.
7. Gilbert B., *Translinear circuits: A proposed classification*, *IEE Electronic* Letters, vol.11, pp.14-16, January 9, 1975.
8. Gilbert, B., *A new technique for analog multiplication*, IEEE Journal of Solid State Circuits, vol.SC-10, no.6, December 1975, pp.437-447.
9. Gilbert, B., and Holloway, P., *A wideband two-quadrant analogue multiplier*, Solid State Circuits Conference, February 1980, pp.200-201.
10. Baskind, D. and Harvey, R., *Techniques for the realisation and application of voltage controlled amplifiers and attenuators*, 60th Convention of the AES, May 1978, preprint 1378 (A-6).
11. Baskind, D., Rubens, H., Kelson, G., *The design and integration of a high performance voltage controlled attenuator*, Convention of the AES, November 1979, preprint 1555 (D-2).
12. Hawksford, M.J., *Distortion correction circuits for audio amplifiers*, JAES, vol. 29, no. 7/8, July/August 1981, pp. 503-510.
13. Hawksford, M.J., *Topological enhancement of translinear two-quadrant gain cells*, JAES, vol. 37, no. 6, June 1989, pp. 485-475.
14. Duncan, B., *VCAs investigated*, Studio Sound, vol 31, nos 6, 7, 8, 9, pt 1 June, pt 2 July, pt 3 August, pt 4 September.

LINEARIZATION OF CLASS D OUTPUT STAGES FOR HIGH-PERFORMANCE AUDIO POWER AMPLIFIERS

S. Logan, M.O.J. Hawksford

University of Essex, England

A system is described which uses a quantized feedback technique to reduce the distortion generated by a class D PWM amplifier, thus making it more suitable for use in a high-performance audio system. The technique is applicable to digital PWM and applies local error correction in a noise shaping loop to reduce output-stage distortion and thus achieve a performance closer to that defined by the digitally defined PWM code.

0 INTRODUCTION

The advantages of using a class D amplifier are well known and recent research has demonstrated the feasibility of realising an almost completely-digital audio system. Pulse-Width-Modulation offers many advantages as a method of digital-to-analogue conversion and it is particularly suitable for use with a class D power amplifier. PWM signals may now be generated with great accuracy (1-9) and advances in DSP technology allow this process to be incorporated into a power amplifier. Much research into the design of class D power amplifiers has been carried out (11,12) and many of the system impairments that place limits on the quality of sound reproduction have been identified. However, as DSP techniques now exist that can generate low distortion, time quantized PWM code, these performance limitations are mainly a function of output-stage induced distortion. In this paper we seek to show how a local error correcting feedback loop can be applied to the power amplifier output stage to raise its performance closer to that required by very high quality music reproduction systems.

1 DEVELOPMENT

Several methods of correcting for the power amplifier errors were considered, however it was decided to avoid analogue processing that modified pulse amplitude and to use a correction technique that altered only the widths of the incoming PWM pulses, as this was potentially a more efficient and cost-effective process and could be implemented by mainly digital circuitry. As such, the correction system shown in Figure 1 was considered initially.

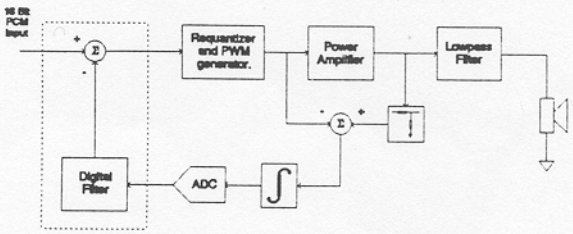


Figure 1: A digital correction system.

The error introduced by the power amplifier is measured by subtracting the input signal from a scaled version of the output signal. The error signal is digitized and filtered, before being subtracted from the incoming PCM signal. This method allows digital processing of the error signal and offers the potential for very high performance to be achieved. This system requires DSP hardware (enclosed by the dashed line in figure 1) to perform the filtering of the error signal and the subtraction from the incoming data. However, as the correction system was envisaged as a relatively simple circuit which could be added to an existing power amplifier without a major redesign, this form of digital correction method was felt to be overly complicated. Consequently, a correction system has been developed which uses an analogue delay line to re-time the

Advanced A-D and D-A Conversion Techniques and their Applications, 6-8 July 1994, Conference Publication No. 393, © IEE, 1994

edges of the PWM waveform. It will be shown that this system can give a worthwhile reduction in distortion (around 50 dB) while being relatively straightforward and cost-effective to implement. The structure of the delay-line error correction system is shown in Figure 2. If the incoming PWM pulses propagate along a tapped delay line then the tap from which the output is taken may be varied dynamically to allow the waveform edges to be re-timed.

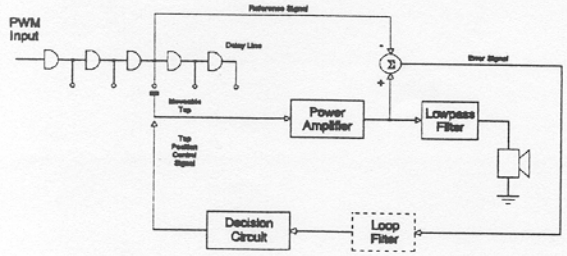


Figure 2: The delay-line correction system.

To aid in system visualization, the output from the delay line is shown as being derived via a moveable tap. In practice all the taps would be connected to a logic switching circuit and the required tap would be selected using an address decoder. This system allows direct control of pulse-edge timing over the incoming PWM signal which realizes the advantages of digital correction without the complexity of DSP hardware. The method of operation of the delay-line correction system is as follows:
The incoming PWM signal propagates along a delay line which has a number of taps. The output from a selected delay-line taps is then passed to the power amplifier which switches its output voltage. However, due to switching errors and power supply droop etc., there is an addition of distortion to the output signal. This error introduced by the power amplifier is measured by subtracting the input signal from a scaled version of the output signal. The derived error signal is processed by the loop filter and then passed on to the decision circuit, which selects the delay-line tap from which the incoming PWM signal will be read. If the delay-line tap is moved as the edges of the incoming PWM signal are passing through the delay line then the edges may be re-timed to compensate for the error introduced by the power amplifier. The loop filter in this system is a single integrator (although other filters are being investigated) and the decision circuit is a low-resolution (typically 5 bit) analogue-to-digital converter with an associated address decoder. The analogue delay line is readily available and consist of an LC delay line with digital buffer gates giving equal-increment taps.

2 INVESTIGATIONS

Software was developed to simulate the operation of a class D power amplifier in order to evaluate the correction system. Sampled sinewave test signals are generated which consist of three frequencies of equal amplitude. These are sampled at 44.1 kHz and represented as double-precision floating-point values. The test signals are 8 times oversampled and requantized to 7 bits using 4th-order noise shaping (10). The requantized signals are then used to generate a trailing-edge modulated pulse train, which is fed into the simulated power amplifier. For the purpose of this simulation, digital linearization was not performed as it is only the output stage that is under investigation. The power amplifier introduces errors into the pulse train, which the correction system under test attempts to remove. The output waveforms before and after correction are compared to assess the effectiveness of the correction system. FFT analysis can be performed to investigate the effects of the correction system on the frequency spectrum of the incoming signals. The power amplifiers used in the tests operate in class AD, and have been specified to allow 95% modulation of the output signal.

3 RESULTS

The power amplifier errors used in the tests consisted of an exponential droop in the pulse amplitudes and a linear slewing of the pulse

edges, resulting in the distortions shown in figure 3.

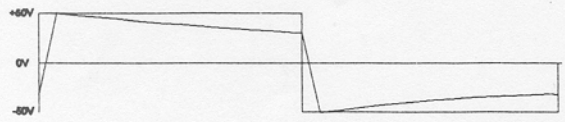


Figure 3: Simulated pulse errors.

Figure 4 shows the result of applying the correction system to an amplifier with these distortions. The error signal from the corrected power amplifier (solid line) is considerably smaller than that from the uncorrected one. This test shows a reduction in output signal mean-squared error of ≈ 38 dB.

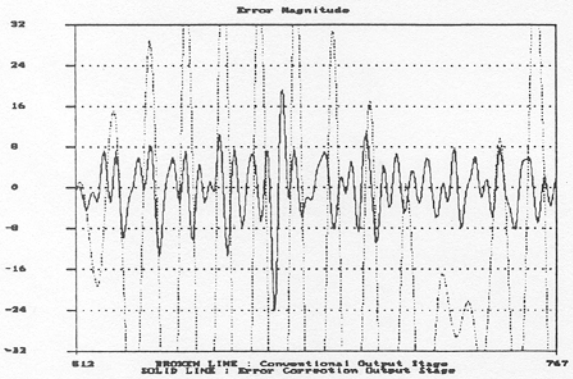


Figure 4: Error waveforms, showing the improvment due to the correction system.

3.1 Effects of correction step size

Correction is applied in discrete steps by moving the delay-line tap, as described in section 1. The size of the correction steps is determined by the inter-tap delay of the analogue delay line. For the initial tests, an inter-tap delay of 20 ns was chosen, as this corresponds to approximately a 1 bit change in the value of the incoming signal. It was found that this correction step size was too coarse and that the system was unable to correct effectively for small errors. Further tests were then carried out with smaller correction step sizes. The results of these tests are shown in Figures 5-7 (results are scaled to show the corrected output signal more clearly). It can be seen that the output-stage errors are further reduced as the correction step size is reduced.

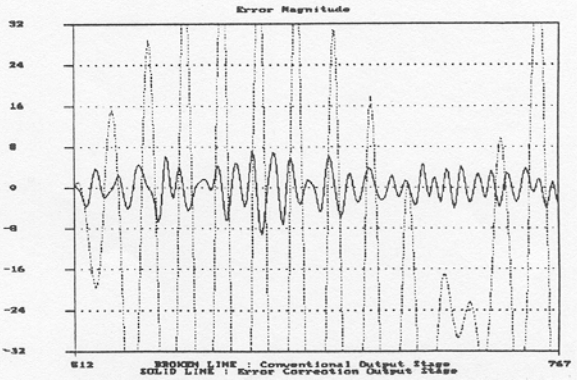


Figure 5: Error waveforms, 10 ns step size.

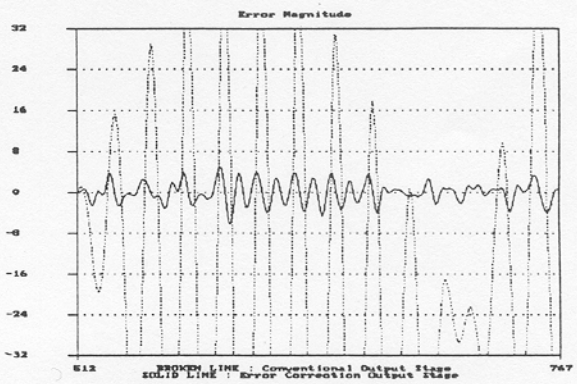


Figure 6: Error waveforms, 5 ns step size.

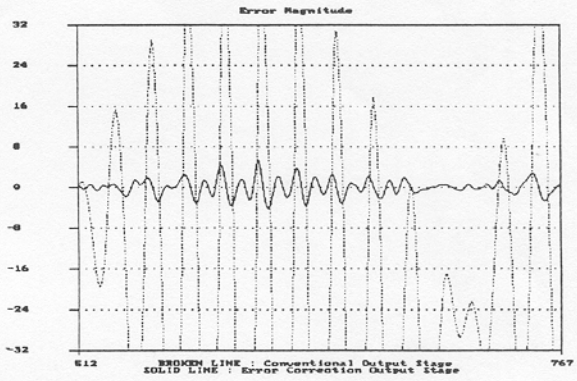


Figure 7: Error waveforms, 2.5 ns step size.

It was found that the amount of error reduction was related to the correction step size, as illustrated in figure 8. The graph shows a plot of mean-squared output-signal error reduction (in dB) against delay-line step size. An almost linear relationship is observed.

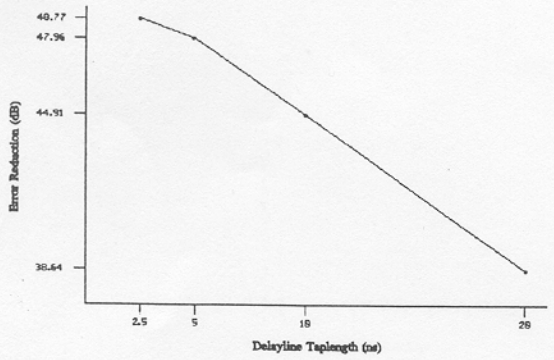


Figure 8: The relationship between mean^2 error reduction and correction step size.

Figure 9 shows the spectrum (4096 point FFT) of the uncorrected power amplifier error signal. Large error components are visible at the three input frequencies. Figure 10 shows the FFT of the corrected power amplifier error signal (2.5 ns system). It can be seen that there is a considerable reduction (between 25 and 54 dB) in the amplitudes of the three large error components.

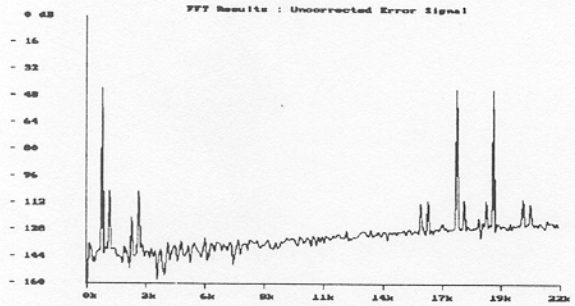


Figure 9: Uncorrected power amplifier error signal FFT plot.

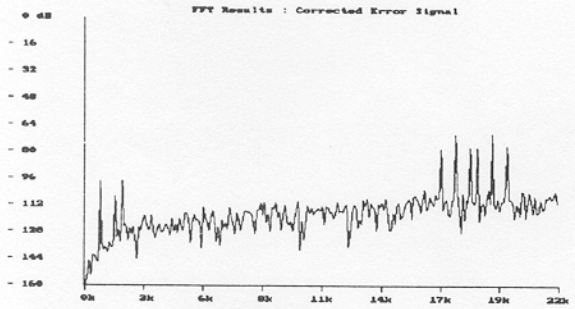


Figure 10: Corrected power amplifier error signal FFT plot (2.5 ns system).

The results show that the penalty for the improved distortion performance is an increase in the noise floor. There are also some additional intermodulation products introduced by the correction system and future work will be directed towards removing these, possibly by investigating higher-order loop filters.

3.2 Tapered delay line

When the delay line tap-length is reduced to allow finer correction, the range of errors which can be handled is reduced. Therefore the number of taps in the delay line must be increased to compensate. The original 20 ns system had a correction range of ± 10 taps. This increases to ± 40 taps when the step size is reduced to 2.5 ns. It was felt that the circuit required to implement this number of taps would be overly complicated, and so the number of taps should be reduced. A system was proposed which uses a tapered delay line, with fine 2.5 ns steps near the centre with more coarse 5 ns steps at the ends. The results of this system are shown in figure 11 below.

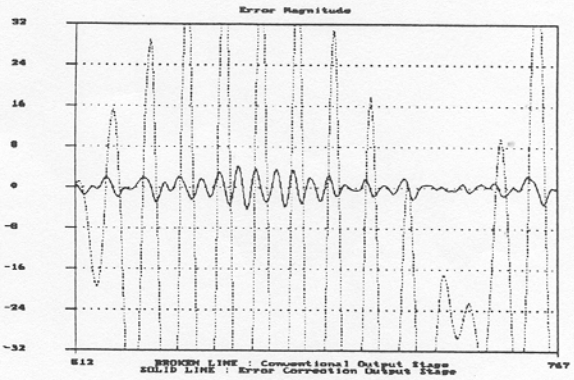


Figure 11: Error waveforms, tapered system.

Comparing this results to figure 7 shows that a similar performance is obtained while reducing the size of the delay line from ± 40 taps to ± 15. This allows a considerable reduction in circuit complexity.

3.3 Additional tests

The ability of the correction system to remove power-supply ripple was investigated, as this is a problem to which class D amplifiers are particularly susceptible. Figure 12 shows the result of applying the correction system to an

amplifier with 100 Hz power supply ripple. It can be seen that the error varies over a large range and this exercises the whole dynamic range of the correction system. The system performs well under these conditions, and the corrected error signal remains small throughout this test.

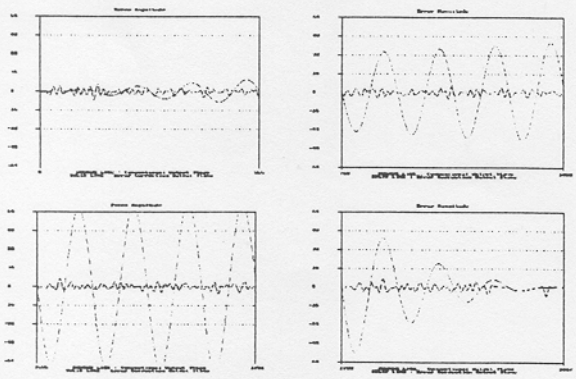


Figure 12: Error waveforms, 100 Hz ripple.

Figures 13 and 14 show the FFT plots of the two error signals where the advantage from the correction system is apparent.

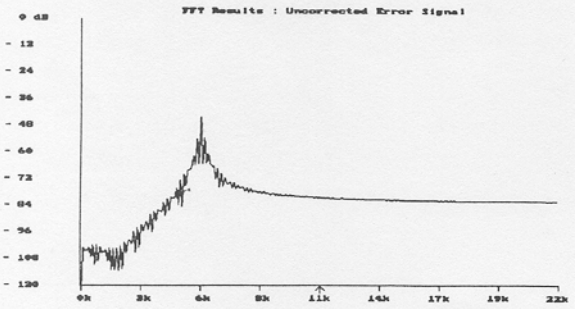


Figure 13: Uncorrected amplifier error signal FFT plot.

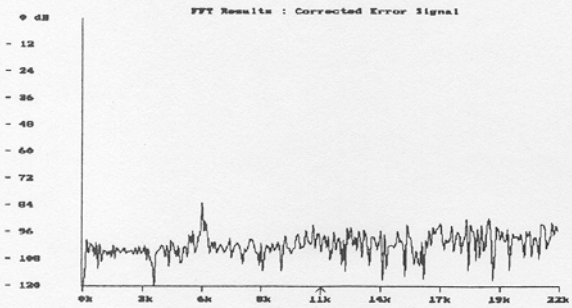


Figure 14: Corrected amplifier error signal FFT plot.

A test was performed using 100 kHz power-supply ripple, to simulate the operation of a switched-mode power supply. A single frequency input signal was used in this test, as this shows the effects of the power supply ripple more clearly. The output waveforms before and after correction are shown in Figure 15. The FFT plots of the corrected and uncorrected error signals are shown in Figures 16 and 17. Again, an encouraging result is obtained.

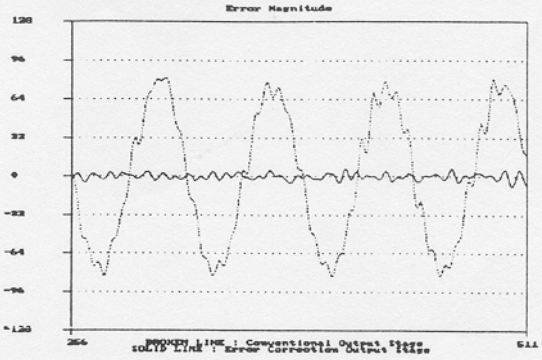


Figure 15: Error waveforms, 100 kHz ripple.

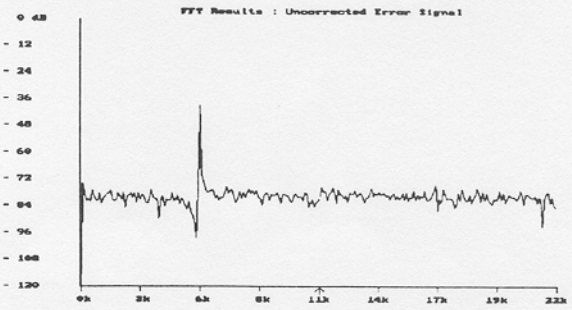


Figure 16: Uncorrected amplifier error signal FFT plot.

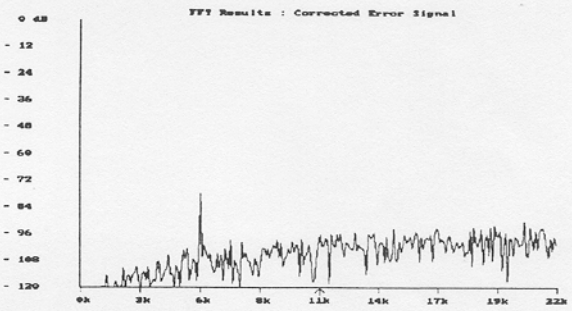


Figure 17: Corrected amplifier error signal FFT plot.

4 Conclusion

The problems encountered in the design and construction of high-quality class D power amplifiers are well known. Notwithstanding, class D amplifiers offer the advantages of higher efficiency and smaller size. The audio signal can also be retained in the digital domain up to the point of power conversion,

thus minimising analogue related imperfections and possibly maximizing system signal-to-noise ratio. A correction system has been developed which uses a digitally controlled analogue delay line to re-time the edges of the incoming PWM signal to reduce the errors introduced by the power amplifier. This system does not require any additional digital signal processing hardware and should therefore be relatively straightforward to implement. A time-domain error reduction of 50 dB has been demonstrated by simulation using this system. Also, it has been shown that the correction ability of the system is directly related to the size of the correction steps used. Correction step sizes down to 2.5 ns have been investigated, but available computing resources precluded taking these investigations any further. It was found that the number of delay-line taps could be reduced by employing a tapered delay line. This allowed a 60% reduction in the number of taps, while retaining similar performance to the uniform system. It should be noted that the errors used in these tests were relatively large. Ideally more realistically sized errors should be used, but this would again require smaller correction step sizes. If more powerful computing resources were available, then the accuracy of the simulations could be improved. Smaller step sizes, higher simulation sampling ratios and smaller power-amplifier errors should allow for a more accurate system appraisal. Nevertheless, encouraging results have been obtained and it is anticipated that future work will improve upon these as there are several extensions yet to be investigated.

References

1 Hawksford, M.O.J., "Multi-level to 1 bit transformations for applications in digital-to-analogue converters using oversampling and noise shaping", *Proc. Institute of Acoustics*, vol 10, pp 129-143, November 1988

2 Hawksford, M.O.J., "Dynamic model based linearization of quantized pulse-width-modulation for applications in digital-to-analogue conversion and digital power amplifier systems", *JAES*, vol 40, no 4, pp 235-252, April 1992

3 Craven, P., "Towards the 24-bit DAC: Novel noise-shaping topologies incorporating correction for the nonlinearity in a PWM output stage", *JAES*, vol 40, no 4, pp 291-313, May 1993

4 Sandler, M.B., "Digital-to-analogue conversion using pulse width modulation", *IEE Electronics & Computing Journal*, vol 5, no 6, pp 339-348, December 1993

5 Mellor, P.H., Leigh, S.P. and Cheetham, B.M.G., "The implementation and performance enhancement of a completely digital power amplifier", *Proc. Institute of Acoustics*, vol 12, pt 8, pp 67-75, November 1990

6 Mellor, P.H., Leigh, S.P. and Cheetham, B.M.G., "Improved sampling process for a digital, pulse-width-modulated, class D power amplifier", *IEE Colloquium on Digital Audio Signal Processing*, London, May 1991

7 Goldberg, J. and Sandler, M.B., "Comparison of PWM modulation techniques for digital power amplifiers", *Proc. Institute of Acoustics*, vol 12, pt 8, pp 57-65, November 1990

8 Bowman, R.G., Hiorns, R.E. and Sandler, M.B., "Design and performance of a noise shaping, pulse-width-modulated digital-to-analogue converter", *IEE Colloquium on Digital Audio Signal Processing*, London, May 1991

9 Goldberg, J.M. and Sandler, M.B., "Noise shaping and pulse-width modulation for an all-digital audio power amplifier", *JAES*, vol 39, no 6, pp 449-460, June 1991

10 Hawksford, M.O.J., "Chaos, oversampling and noise shaping in digital-to-analog conversion", *JAES*, vol 37, no 12, pp 980-1001, December 1989

11 Attwood, B.E., "Design parameters important for the optimization of very high-fidelity PWM(class D) audio amplifiers", *JAES*, vol 31, no 11, pp 842-853, November 1983

12 Hancock, J.N., "A class D amplifier using MOSFETs with reduced minority carrier lifetime", *JAES*, vol 39, no 9, pp 650-662, September 1991

Dynamic Model-Based Linearization of Quantized Pulse-Width Modulation for Applications in Digital-to-Analog Conversion and Digital Power Amplifier Systems*

M. O. J. HAWKSFORD, *AES Fellow*

Department of Electronic Systems Engineering, University of Essex, Colchester, Essex, CO4 3SQ, UK

Quantized pulse-width modulation (PWM) offers an efficient means of converting digital data to analog either at low signal levels for DAC systems or at higher levels in power amplification. However, although a number of techniques exist, they are flawed by varying degrees of dynamic nonlinearity inherent to the conversion process, especially at lower sampling rates. The basic nonlinear mechanisms of PWM are described and a family of model-based solutions to the linearization problem is presented that retains the advantage of a uniform sampled digital format. As such it is possible to design PWM converters that exhibit vanishing levels of distortion even under broadband high-level signal excitation.

0 INTRODUCTION

Pulse-width modulation (PWM) was invented by A. H. Reeves, who also invented pulse-code modulation (PCM) [1]–[3] and the capacitor microphone. The technique of PWM has had a long history of theoretical and practical development, where it has found wide application in power-efficient amplifiers and power-supply systems [4], [5]. Indeed, it is somewhat appropriate that both PCM and PWM originate from the same inventor, as this paper seeks to find an optimum solution to combining the two systems within the digital domain. Recently PWM has been identified as a means of achieving digital-to-analog conversion (DAC) and has been used in MASH conversion systems in association with noise shaping and oversampling. Such systems require the process of generating PWM to be performed in the digital domain, where an early study was undertaken by Sandler [6], with evolutionary extensions subsequently reported in numerous related papers [7]–[10].

One approach is to mimic the analog methods of implementing PWM using a digital ramp function and comparator. However, if the technique of natural sampling is implemented, then there is significant computational complexity in calculating the intersection of ramp and finely time-quantized, oversampled, and band-limited data as well as extremely high clock rates to time the pulse transitions. More recently it has been recognized that oversampling and noise shaping with multilevel quantization can be used to reduce the resolution required of the PWM signal, where an outline system was proposed [11, sec. 8.3] by the author in 1985. Later work by Sander and coworkers [12], [13] has developed this theme and supported the feasibility of more practical switching rates in association with acceptable signal-to-noise ratios. Theoretically the target performance of PWM is enhanced by an increase in the sampling rate, although the design of the switching output stage is then more problematic and efficiency degrades due to the extra signal transitions where both voltage and current occur simultaneously in the output transistors. Also edge slewing, jitter, and power-supply compliance and its associated transient response degrade performance. For high efficiency the sampling rate should be low, but for low distortion and ease of signal reconstruction using low-pass filtering, the rate should be high. However, it will also be shown that an increase in sampling rate requires an increase in system accuracy,

* Manuscript received 1991 September 19; revised 1991 December 14.

thus negating some of the reported advantage of noise shaping and excessive oversampling.

The principal thrust in modern PWM systems is to convert uniformly time-sampled and amplitude-quantized audio data directly into an efficient PWM code that exhibits a low level of nonlinearity. Ideally, only modest oversampling should be used, which is sufficient to gain a signal recovery and possible noise-shaping advantage, but limited so as not to degrade efficiency. Also, it is advantageous if the bulk of signal processing occurs at the uniform sampling rate to avoid the need for excessive oversampling ratios to mimic the process of natural sampling.

In this paper we address directly the principal distortion-generating mechanisms of PWM and demonstrate that by use of dynamic (or time-varying) filtering, model-based linearization can be achieved to a high degree of accuracy that generates a low residue, even in the presence of high-level and energetic signals. The process is first demonstrated using a symmetrical nonrecursive filter architecture with constrained recursive coefficient adaptation forward and backward of the present output sample, which is suitable for finely resolved PWM. The technique is then applied to more coarsely quantized structures, which also enables a noise-shaping advantage to lower the resolution of the PWM samples.

Two recent papers by Mellor et al. [14], [15] have demonstrated significant reductions in PWM distortion by modifying the uniform sampling process. As it is well documented that naturally sampled PWM offers enhanced linearity over uniform sampling, Mellor proposes using linear interpolation operation between adjacent input samples to estimate the PWM pulse transitions so as to approximate those of a naturally sampled system. This is a fundamental proposition in the evolution of digital PWM that paves the way to a low-distortion, all-digital power amplifier system. However, research now suggests that there exist more optimal strategies for linearizing uniformly sampled digital PWM, and in this paper an approach argued from the spectral domain is presented.

1 NONLINEAR MECHANISMS IN PWM

The discussion in this paper is restricted to a particular class of PWM, where the input data are assumed to be uniformly sampled and the pulse width is calculated at each sampling instance. This technique simplifies signal processing compared with natural sampling, where in digital architectures there is a significant complication in calculating the intersection of a finely interpolated audio signal with that of a linear staircase function, as illustrated in Fig. 1.

Initially we shall assume audio data in four times oversampling format, although alternative oversampling ratios can also offer advantage in association with noise shaping. The uniformly sampled PWM process maps each data sample to an equivalent-area rectangular pulse of constant amplitude but variable width. We consider a symmetrical pulse distribution, as shown in Fig. 2. The width of the pulse τ is then determined as

$$\tau = \frac{x(n)}{A} T_s \tag{1}$$

where A is the pulse amplitude, $x(n)$ the amplitude of the nth input sample, and T_s the sampling period of

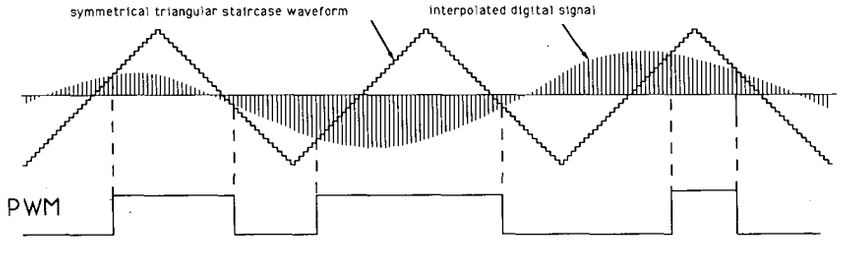


Fig. 1. Natural sampling showing interpolated digital signal and problem of calculating intersections.

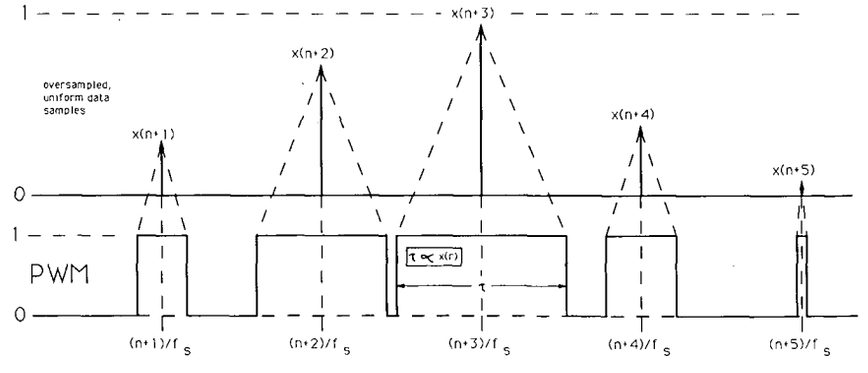


Fig. 2. Oversampled symmetrical PWM using uniform sampling where pulse width is directly proportional to sample amplitude.

the PWM sequence. In this example $0 < x(n) < A$ and corresponds directly to a pulse width range of $0 < \tau < T_s$. To accommodate a bipolar PWM format, a dc level of $A/2$ is subtracted from the square wave and a corresponding dc offset of $A/2$ is added to the input sequence, transforming $x(n)$ to a range of $-A/2$ to $A/2$.

The nonlinear distortion inherent in this class of PWM process arises from two elements.

1) The spectrum of the input sequence is replicated about the sampling frequency and its harmonics. Consequently the baseband signal must adhere to conventional uniform sampling theory to prevent aliasing distortion.

2) Each sample is then replaced by a constant-amplitude rectangular pulse with an area proportional to the sample amplitude $x(n)$. Although this guarantees exact dc coding, the broad spectrum of each individual pulse is dynamically modified as a function of the pulse width. Hence following subsequent summation over all pulses, it is this dynamic spectral modulation that is the root of nonlinearity in PWM.

To examine the mechanism of dynamic spectral modulation, the Fourier transform $F_n(f)$ of the nth pulse, illustrated in Fig. 2, is expressed as

$$F_n(f) = \frac{\tau A}{T_s}\left[\frac{\sin(\pi f\tau)}{\pi f\tau}\right] e^{-j2\pi nfT_s} \tag{2a}$$

where τ is calculated from Eq (1). For a low-PWM modulation index, $\pi f\tau << \pi/2$,

$$F_n(f) \approx \left[\frac{\tau A}{T_s}\right] e^{-j2\pi nfT_s} \tag{2b}$$

and the modulation process is approximately linear. However for higher modulation indices the $\sin(\pi f\tau)/\pi f\tau$ factor modifies the spectrum of each individual sample, introducing a frequency-dependent gain modulation that generates nonlinear distortion products over a broad spectrum. In essence, each pulse width has a unique Fourier transform that exhibits differential gain errors with increasing frequency. In Fig. 3 $\sin(\pi f\tau)/\pi f\tau$ is plotted as a function of f for a family of τ, where gain modulation is evident. To illustrate this distortion mechanism, Fig. 4(a) shows the output spectrum of a symmetrical uniformly sampled PWM system for a periodic input sequence $x(n)$ calculated over 1024 samples. Here

$$x(n) = A[0.5 + 0.15\sin(k_0 n) + 0.15\sin(k_1 n) + 0.15\sin(k_2 n)] . \tag{3}$$

The sampling rate is 176 kHz and the three equiamplitude superimposed input signal frequencies are 171.875 Hz, 17.1875 kHz, and 20.45313 kHz. However, to expose the intermodulation distortion in the

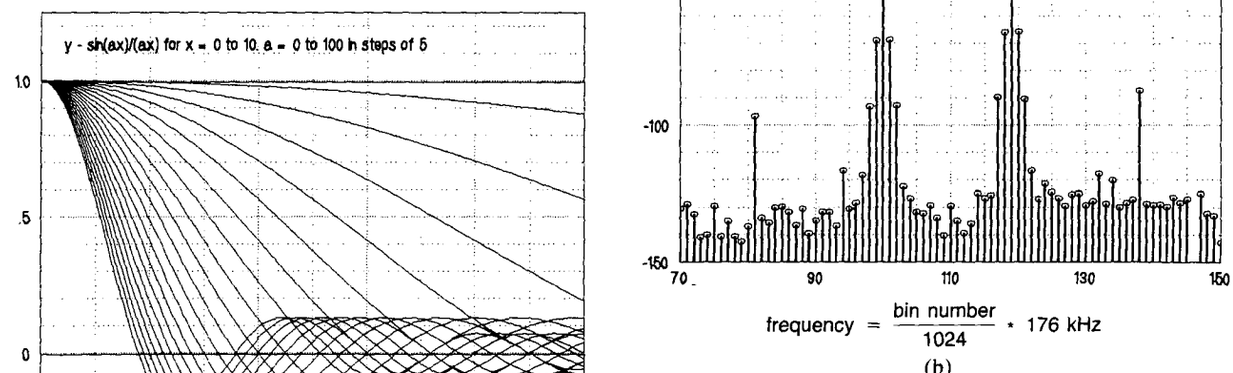


Fig. 3. Family of sinc(x) functions representing dynamic modulation of pulse spectrum in PWM.

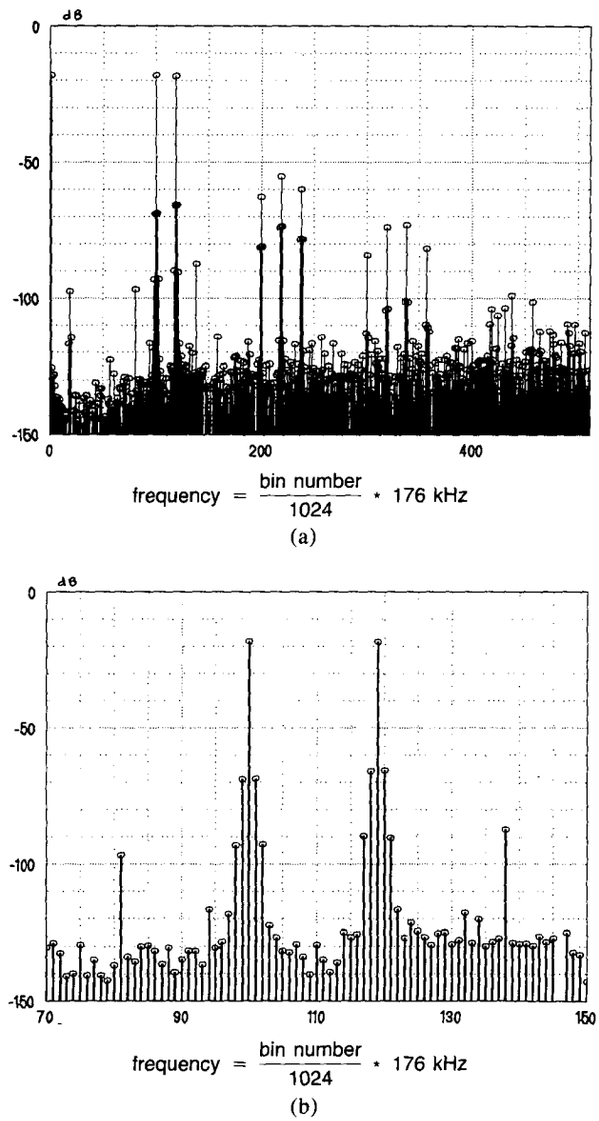


Fig. 4. (a) Noncorrected PWM spectrum computed over 1024 samples. Sampling rate 176 kHz; correction band 0–22 kHz; input frerquencies 171.875 Hz, 17.1875 kHz, and 20.45313 kHz. (b) Spectrum of (a) with expanded frequency scale to show intermodulation products.

PWM output spectrum, an expanded-frequency-scale plot is shown in Fig. 4(b). The resultant spectrum $F_p(f)$ is computed over 511 samples and corresponds to

$$F_p = \frac{1}{NS} \sum_{n=0}^{NS-1} F_n(f) \ . \tag{4}$$

It is observed that dynamic spectral modulation occurs after the sampling function. Hence the higher frequency distortion products do not undergo aliasing into the baseband, which enables an opportunity for linearization.

2 INTERLEAVED AND DYNAMIC FIR CORRECTION OF THE PWM PROCESS

To attain linearization, the uniformly sampled digital signal must be modified to compensate for the change in spectral shape with pulse width, although the frequency range over which this is achieved can be limited to the 0–22-kHz passband. The main constraint to be imposed upon this process is the restriction to uniformly spaced samples for the modified sequence, whereas the pulse width takes, by its nature, intersample values. The method of correction is to effectively time-interleave multiple finite-impulse-response (FIR) filters between the oversampled PCM data and the pulse-width modulator whose frequency responses are individually matched to the inverse of each corresponding pulse $x(n)$ in the PWM sequence. The overall filtering process is dynamic as the transfer function of each pulse is a function of pulse width, which can vary between zero and the sampling period, although while an individual filter is contributing to an output pulse, its local coefficients are held constant. In practice, the filtering process can be considered as a single filter, but where the coefficients shift with the input data. Thus a coefficient contributing to the current output is also uniquely linked to its corresponding input data sample. However, the exact structure and operation of the correction process become clearer to comprehend when the simulation program segments in Secs. 2.1 and 2.2 are examined.

To introduce the corrective procedure, a symmetrical five-tap FIR filter is shown in Fig. 5, where the incremental time delay T_s is matched to the system sampling

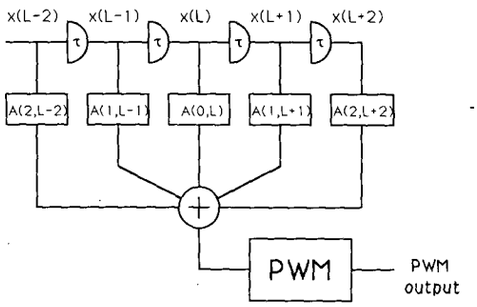


Fig. 5. Five-coefficient FIR filter.

period $1/f_s$ s. The input to the filter is a sampled-data sequence $x(n), 0 < x(n) < 1$, corresponding to a PWM sample width τ, $0 < \tau < 1/f_s$, and the PWM sample is assumed symmetrical about the sample instant. This symmetrical distribution is critical to the present system as it enables a symmetrical FIR filter to be used and therefore prevents complications arising from dynamic phase modulation.

The five coefficients of an FIR filter associated with the current sample n are defined as a_2, a_1, a_0, a_1, a_2. To express the dynamic form of the filter, the corresponding coefficients uniquely linked to individual input samples are represented as $A(2, n-2), A(1, n-1), A(0, n), A(1, n+1), A(2, n+2)$, where n is the current output sample. Thus the notation $A(x, y)$ means coefficient a_x associated with sample $x(y)$.

The aim, therefore, is to associate a unique FIR filter with each PWM mapped input sample $x(n)$ and to compute corresponding filter coefficient sets $\{A(0, n), A(1, n), A(2, n)\}$, where $a_0 = A(0, n)$, $a_1 = A(1, n)$, and $a_2 = A(2, n)$, such that the transfer function of the FIR filter $E_{x(n)}(f)$ and the Fourier transform of the corresponding PWM sample $F_{x(n)}(f)$ closely approximate a (constant) target function $T(f)$ that is matched over the passband 0 to f_u Hz, where f_u is typically 22 kHz for a digital audio system.

However, to maintain constancy of pulse area, the sum of the FIR filter coefficients is unity, where the central coefficient a_0 is given by

$$a_0 = 1 - 2 \sum_{r=1}^{\lambda} a_r \ . \tag{5}$$

That is, where the number of independent coefficients $\lambda = 2$, it is required to calculate two coefficients for a five-tap filter.

Although the target function $T(f)$ could be unity, we choose here the transfer function of a 50% PWM sample corresponding to the quiescent state, $x(n) = 0.5$,

$$T(f) = \frac{\sin(0.5\pi f T_s)}{0.5\pi f T_s} \tag{6}$$

where $T(f)$ is normalized to unity at dc.

The transfer function $E_{x(n)}(f)$ of the five-tap FIR equalization filter associated with the sample $x(n)$, as shown in Fig. 5, can be expressed as

$$E_{x(n)} = a_0 + 2a_1 \cos(2\pi f T_s) + 2a_2 \cos(4\pi f T_s) \tag{7}$$

and the normalized transform $F_{x(n)}(f)$ of a PWM sample corresponding to $x(n)$ is

$$F_{x(n)}(f) = \frac{\sin[x(n)\pi f T_s]}{x(n)\pi f T_s} \ . \tag{8}$$

Consequently the matching task that approximates the

product of the transform of the nth PWM sample and the transform of the nth associated correction FIR filter to the target transform, performed over a frequency band 0 to f_u Hz, can be stated as

$$E_{x(n)}(f)F_{x(n)}(f) \rightarrow T(f)|_{0 \text{ to } f_u} . \quad (9)$$

Substituting and rearranging,

$$a_1[\cos(2z) - 1] + a_2[\cos(4z) - 1] \rightarrow \frac{x(n)\sin(z/2)}{\sin[x(n)z]} - 0.5|_{0 \text{ to } f_u} \quad (10)$$

where

$$z = \frac{\pi f}{f_s} . \quad (11)$$

This matching task could be solved by an optmization procedure to minimize an error function within the 0 to f_u Hz frequency band. However, because the functions exhibit similar curvature over the lower frequency region, a simpler approach (for a_1 and a_2) is to force equality in the function of Eq. (10) at two frequencies, where we select f_u Hz and $0.5f_u$ Hz. Hence solving two simultaneous equations based on Eq. (10) and modifying the functional form to prevent errors due to small differences, the solution can be stated as

$$a_1 = -\frac{S_1 - 4S_2 + 4\sin^2(z)(S_2 - 1) + 3}{D_1}$$

$$a_2 = \frac{S_1 - 4S_2 + 4\sin^2(z/2)(S_2 - 1) + 3}{D_2}$$

where

$$D_1 = 4\left\{\sin^2(z) - 4\sin^2\left(\frac{z}{2}\right)[1 - \sin^2(z)]\right\}$$

$$D_2 = 4\cos^2\left(\frac{z}{2}\right) D_1$$

$$S_1 = \frac{2m\sin(z/2)}{\sin(mz)}$$

$$S_2 = \frac{2m\sin(z/4)}{\sin(mz/2)} . \quad (12)$$

Here m is the normalized input data, $0 < m < 1$, on which the coefficient estimation is based, and the solution is referred to as equation set (12).

To demonstrate the effectiveness of this transform matching process, two sets of Fourier transforms are shown in Fig. 6. The sets include:

1) The target transform corresponding to a pulse, where $m = 0.5$

2) Two extreme case example transforms of pulses, where $m = 0$ and 1

3) The corresponding compensated transforms for $m = 0$ and 1.

In these examples the sampling rate f_s was 176 kHz and the coefficients were calculated for exact matching at 22 and 11 kHz, respectively. The results clearly show how the compensated and target curves are closely matched over the band 0–22 kHz, even though the PWM sample widths vary over the extreme range from 0 to $1/f_s$. A detailed comparison of tabulated data also confirms equality at both 22 and 11 kHz.

In a pratical implementation of this system, a number of techniques can be adopted for determining a_1 and a_2 for each normalized value of m. First, a direct solution of equation set (12) can be implemented for each new sample m. Second, a discrete look-up table can be computed for a range of m and a ROM used to store coefficient values. Finally, a polynomial approximation can be formed for a_1 and a_2 as a function of m, which offers the advantage of efficient signal interpolation. For systems with fine quantization of data (approxi-

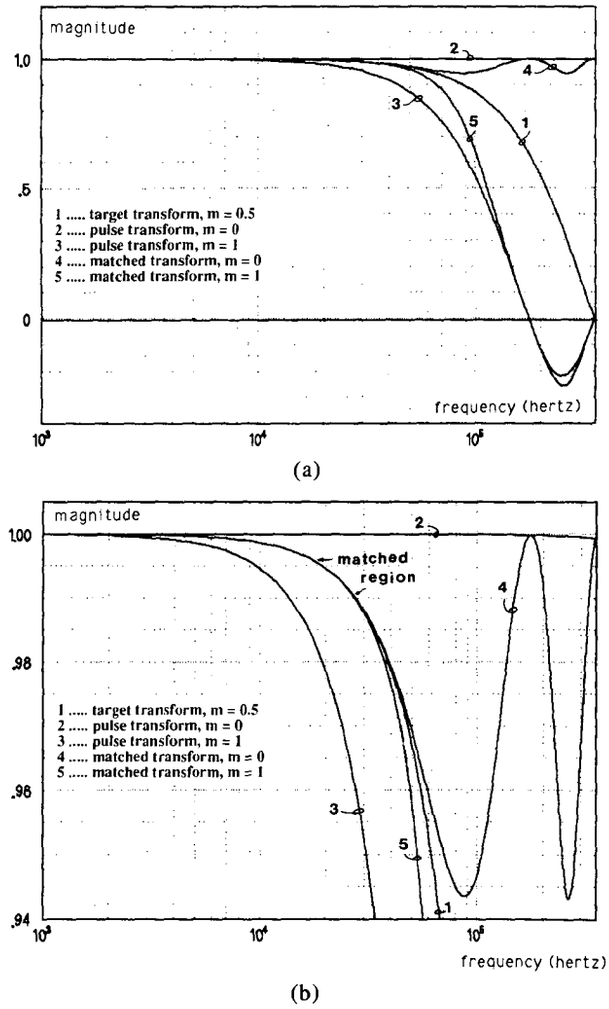


Fig. 6. PWM transform matching functions for $m = 0$ and $m = 1$. Sampling rate 176 kHz; matching frequencies 11 and 22 kHz. (a) Full-range vertical scale. (b) Expanded vertical scale.

mating a continuous function), the polynomial method is the simplest. However, in more coarsely quantized systems with, say, 1024 values of m, the ROM technique using predetermined values is the most efficient. Thus knowing the oversampling ratio and modulation index, the appropriate coefficients can be loaded into the FIR filter, thus making the filter response dynamic or sample-value dependent.

To demonstrate the variation of coefficients with modulation index m, Fig. 7 shows plots of a_1 and a_2 against m for a PWM system with four times oversampling (that is, $f_s = 176$ kHz). Approximate polynomial matched generating functions for a_1 and a_2 are also given as

$$a_1 = -0.00598x^4 + 0.000692x^3 - 0.0554x^2 + 0.0000680x + 0.0141$$

$$a_2 = 0.00151x^4 - 0.000180x^3 + 0.00345x^2 - 0.0000158x - 0.000925 .$$

However, there is a further level of complexity in the correction process that arises from pulse-width-dependent distortion and the associated dispersive response of the dynamic FIR filter. When the coefficients a_1 and a_2 are superimposed on adjacent samples in a more general $x(n)$ sequence and subsequently mapped into a PWM format, they, by their nature, each modify the width of the four adjacent PWM samples, which in turn modify the correction coefficients already calculated and associated with these samples. This nonlinear interdependence of coefficients is a recursive process, which because of the symmetric and two-sided form of the FIR impulse response, propagates both forward and backward of the current sample, although the number of iterations can be constrained.

There are a number of methods for calculating the coefficient set for a specific data sequence, and we present here two examples. Both methods rely on an interactive process to converge to a final solution of coefficients. Hence in the initial discussion we consider samples to be approximately continuous within computer precision.

Fig. 8 shows the general system architecture of the dynamic FIR filter used to determine the output sample sequence $y(L)$ that drives directly the pulse-width modulator. Each sample is associated with five coefficients, which also interact with their adjacent samples and where each coefficient set is calculated from a nonlinear dependence on both the central and the surrounding samples. Consequently as the data samples are shifted at the uniform PWM sampling rate, the coefficients having influence on the output sample also change, resulting in a dynamic FIR filter.

2.1 Zigzag Algorithm

In this example the input data $x(L)$ propagate through a 30-stage shift register, where the current output is taken at $L = 18$ and the clock rate $f_s = 1/T_s$ Hz. The input data sequence $x(L)$ is transformed to the output sequence $y(L)$ via a coefficient matrix $A(r, L)$,

$A(0, L)$ coefficient a_0 associated with sample L
$A(1, L)$ coefficient a_1 associated with sample L
$A(2, L)$ coefficient a_2 associated with sample L.

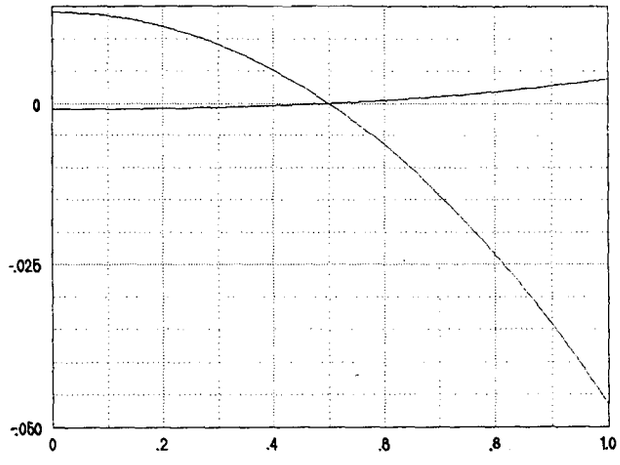


Fig. 7. Plots of coefficients a_1 and a_2 against m for $f_s = 176$ kHz and $f_{opt} = 0-22$ kHz.

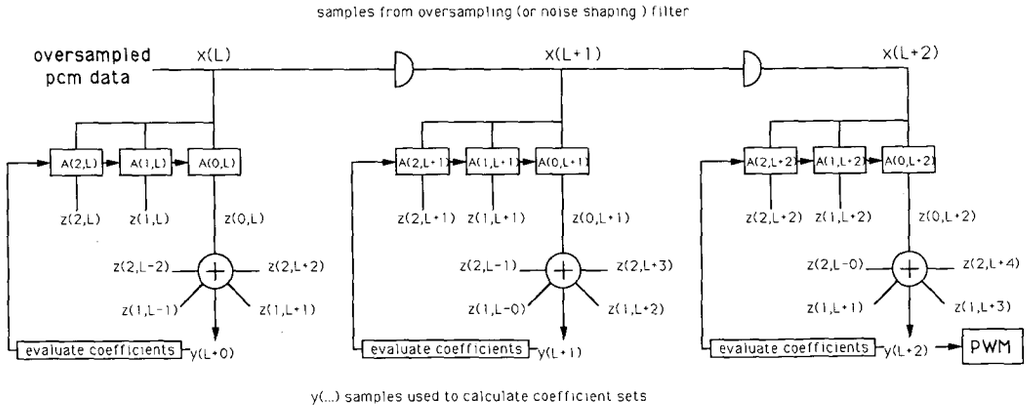


Fig. 8. Interconnected coefficient sets, where each set is calculated from knowledge of surrounding samples.

From Eq. (5),

$$A(0, L) = 1 - 2[A(1, L) + A(2, L)] . \quad (13)$$

As $a_2 << a_1 << a_0$, the influence of a sample on its surrounding samples decays rapidly, so only a limited number of samples need be considered either side of the output sample $L = 18$. In the present example, an input vector spans $L = 1$ to 30, where the zigzag procedure operates as follows:

1) All coefficients $A(0, L)$ are set to 1 and coefficients $A(1, L)$, $A(2, L)$ are set to zero.

2) The coefficient set associated with $L = 18$ is calculated.

3) The coefficient set associated with $L = 17$ is calculated using $A(1, 18)$.

4) The coefficient set associated with $L = 19$ is calculated using $A(1, 18)$ and $A(2, 17)$.

5) The procedure is repeated following a zigzag trajectory either side of $L = 18$, moving progressively outward, and calculating the output matrix $y(L)$ using the latest coefficient estimates via the expression

$$\begin{aligned} y(L) = {} & x(L-2) * A(2, L-2) + x(L-1) \\ & * A(1, L-1) + x(L) * A(0, L) \\ & + x(L+1) * A(1, L+1) \\ & + x(L+2) * A(2, L+2). \end{aligned} \quad (14)$$

6) Steps 2) to 5) are then repeated, say four times, allowing convergence to the final coefficient set.

An important attribute of this algorithm is that the input samples remain unchanged and the output samples $y(L)$ are used only to determine the latest coefficient sets. The recursive procedure that runs both forward and backward of the output sample allows a convergence of the coefficients surrounding $L = 18$, hence forming an estimate of the signal required to drive the pulse-width modulator.

The process is encapsulated in the following subroutines, where the coefficients are estimated using equation set (12), with appropriate constants determined by the oversampling ratio:

```
REM coefficient optimization using zigzag algorithm
FOR L = 1 TO 20
A(L, 0) = 1 : A(L, 1) = 0 : A(L, 2) = 0
NEXT
FOR Q=1 TO 6
FOR P=0 TO 7
L = 10 - P : Y(L) = X(L-2)*A(L-2, 2) + X(L-1)*A(L-1, 1) +
X(L)*A(L, 0) + X(L+1)*A(L+1, 1) + X(L+2)*A(L+2, 2)
GOSUB 10000: REM subroutine to evaluate A(L, 0), A(L, 1) and
A(L, 2) on Y(L)
L = 10 + P : Y(L) = X(L-2)*A(L-2, 2) + X(L-1)*A(L-1, 1) +
X(L)*A(L, 0) + X(L+1)*A(L+1, 1) + X(L+2)*A(L+2, 2)
GOSUB 10000: REM subroutine to evalute A(L, 0), A(L, 1) and
A(L, 2) on Y(L)
NEXT : NEXT
FOR L = 20 TO 1 STEP -1
X(L) = X(L-1)
NEXT
REM last evaluation of Y(18) forms PWM drive
RETURN

10000 : REM direct coefficient evaluation subroutine based on
equation set (12)
E1 = 2*Y(L)*SIN(.5*Z)/SIN(Y(L)*Z) : E2 = 2*Y(L)*SIN(.25*Z)/
SIN(.5*Y(L)*Z)
A(L, 1) = (-E1 + 4*E2 - 4*(E2 - 1)*SIN(Z)^2 - 3)/Z1
A(L, 2) = (E1 - 4*E2 + 4*(E2 - 1)*SIN(Z/2)^2 + 3)/Z2
A(L, 0) = 1 - 2*(A(L, 1) + A(L, 2))
RETURN
```

2.2 Single-Sided Unidirectional Shifting Algorithm

A more efficient procedure uses a single unidirectional scan of the data $x(L)$ to determine coefficients, but shifts both data and coefficients on the system clock. Consequently when the data are scanned, the new coefficient estimates can use the past coefficient estimates alongside their corresponding input sample values. After each shift function, the data are scanned from $L = 1$ to $L = 16$, calculating new coefficients via Eq. (12). The new data samples cause the latest input coefficients to adapt rapidly, but the changes become progressively smaller at higher sample values, as effectively the 16th sample has been through 16 iterations. The output is again taken from $L = 18$, but now the coefficient sets linked with samples $L = 16-20$ do not change, meaning that the pulse area contribution of each input sample is preserved precisely through Eq. (13). As processing is always based on the actual input samples $x(n)$ and Eq. (13) is applied rigorously after final convergence, effects of computational errors are noncumulative. The procedure is described succinctly in the following subroutine.

```
REM coefficient optimization using single-sided shifting algorithm
FOR L = 20 TO 1 STEP -1
X(L) = X(L-1) : A(L, 0) = A(L-1, 0) : A(L, 1) = A(L-1, 1): A(L,
2) = A(L-1, 2)
NEXT
XX = A(0, 0)*X(0) + A(1, 1)*X(1) + A(2, 2)*X(2)
GOSUB 10000: REM subroutine to evaluate A(L, 0), A(L, 1), and
A(L, 2) on XX
XX = A(0, 1)*X(0) + A(1, 0)*X(1) + A(2, 1)*X(2) + A(3, 2)*X(3)
GOSUB 10000 : REM subroutine to evaluate A(L, 0), A(L, 1), and
A(L, 2) on XX
FOR L = 2 TO 18
XX = A(L-2, 2)*X(L-2) + A(L-1, 1)*X(L-1) + A(L, 0)*X(L) +
A(L+1, 1)*X(L+1) + A(L+2, 2)*X(L+2)
GOSUB 10000 : REM subroutine to evaluate A(L, 0), A(L, 1), and
A(L, 2) on XX
NEXT
REM last value of XX forms PWM drive
RETURN
10000 : REM direct coefficient subroutine based on equation set (12)
E1 = 2*XX*SIN(.5*Z)/SIN(XX*Z) : E2 = 2*XX*SIN(.25*Z)/
SIN(.5*XX*Z)
A(L, 1) = (-E1 + 4*E2 - 4*(E2 - 1)*SIN(Z)^2 - 3)/Z1
A(L, 2) = (E1 - 4*E2 + 4*(E2 - 1)*SIN(Z/2)^2 + 3)/Z2
A(L, 0) = 1 - 2*(A(L, 1) + A(L, 2))
RETURN
```

2.3 Response of Linearization Process to Impulsive Data

Table 1 shows the system response to a single impulsive input. First the linearization algorithm is conditioned by an input sequence $x(L) = 0.5$, which sets coefficients a_1 and a_2 to zero. A single sample is then entered where $x(L) = 0.83$. (Note that a unity modulation index corresponds to a maximum pulse duration of T_s.) Ten samples on either side of the impulse are computed, and the distribution of coefficients is also

shown to demonstrate the bilateral dispersive response of the linearization algorithm.

2.4 Spectral Response of Linearization Process

Consider a PWM sequence that is periodic over NS samples and where the modulation index of sample r is $m(r)$ and is normalized to a range of $0 < m(r) < 1$ corresponding to a pulse width $0 < \tau < T_s$. Noting the spectral weighting function described by Eqs. (2) and (4), the Fourier transform of this sequence is given by

$$F(f) = \frac{1}{\text{NS}} \sum_{r=0}^{\text{NS}-1} m(r) \left\{ \frac{\sin[m(r)\pi fT]}{m(r)\pi fT} \right\} e^{-j2\pi rfT} .$$

Harmonic X of the fundamental frequency is $f_x = (X/\text{NS})f_s$, and since $T_s = 1/f_s$, then $f_xT_s = X/\text{NS}$, whereby the transform becomes

$$F\left(\frac{Xf_s}{\text{NS}}\right) = \sum_{r=0}^{\text{NS}-1} \frac{\sin[\pi X m(r)]}{\pi X} e^{-j2\pi rX/\text{NS}} . \quad (15)$$

Eq. (15) can be used to compute the Fourier transform of the pulse sequence $m(r)$ that includes the PWM distortion, resulting from sample reconstruction using rectangular pulses of constant amplitude but variable width, that are symmetrically arranged about their sampling instants. However, if the data sequence $m(r)$ has been predistorted by the linearization algorithm, the effects of nonlinearity should be virtually negated. A data sequence, identical to that described in Sec. 1, has been processed by the algorithm described in Sec. 2.2, where the output spectrum [computed using eq. (10)] is shown in Fig. 9. Comparing this result with the noncorrected transform shown in Fig. 4 reveals a significant reduction in intermodulation products and thus demonstrates the effectiveness of the linearization process.

2.5 Optimum Estimation of Equalizer Target Function

Eq. (6) showed a target function normalized to $m = 0.5$, while in Fig. 7 an example set of coefficients was shown for $f_s = 176$ kHz. These results reveal asymmetry in the distribution of a_1 and a_2, where the moduli are greater for $m = 1$ compared with $m = 0$. An alternative strategy can shift the target response away from $m = 0.5$ in order to achieve greater coefficient equality at $m = 0$ and $m = 1$, respectively, and thus minimize the maximum contribution from the corrective signal. This can be achieved by modifying S_1 and S_2 in equation set (12) to

$$S_1 = \frac{m \sin(\sigma z)}{\sigma \sin(mz)} \qquad S_2 = \frac{m \sin(\sigma z/2)}{\sigma \sin(mz/2)}$$

and searching σ to give near equality in a_1 and a_2 at the extremes of m. As an example for $f_s = 176$ kHz and $f_u = 22$ kHz, the best match is achieved at $\sigma = 0.72373$, giving peak values of $a_1 = 0.0292264$, $a_2 = -0.001855974$ at $m = 0$ and $a_1 = -0.0292264$, $a_2 = 0.002353529$ at $m = 1$, respectively, which can be compared against Fig. 7 for $\sigma = 0.5$. Also, to demonstrate that the modified target transfer function represents a valid solution, a spectral plot is shown in Fig. 10 that should be compared with Fig. 9. Although the magnitudes of the correction coefficients are now reduced overall, their values at low signal level (such as $m \approx 0.5$) are actually greater than in the case where $\sigma = 0.5$. To regain symmetry in the processing of signals in the ranges 0 to 0.5 and 0.5 to 1, a differential topology can be used where two modulators, including linearization, now operate with input signals of the form $1 + m(r)/2$ and $1 - m(r)/2$, respectively. The output is then formed by either a difference stage or a

Table 1. Response of linearization algorithm to single impulsive input.

		Coefficients		
Input sample $x(n)$	Output sample $y(n)$	a_0	a_1	a_2
0.5	0.5	1.0	0.0	0.0
0.5	0.5	1.0	0.0	0.0
0.5	0.5	1.0	0.0	0.0
0.5	0.4999995	1.0	0.0	0.0
0.5	0.5	1.0	0.0	0.0
0.5	0.4999995	1.0	0.0	0.0
0.5	0.5000148	1.0	0.0	0.0
0.5	0.4997598	0.9999622	2.167327E-05	−2.816349E-06
0.5	0.5049456	1.000537	−2.889769E-04	2.065322E-05
0.5	0.4555999	0.9954436	2.449079E-03	−1.708585E-04
0.83	0.9943724	1.08405	−4.578839E-02	3.763581E-03
0.5	0.4555992	0.9954542	2.441855E-03	−1.689809E-04
0.5	0.504934	1.000537	−2.889769E-04	2.065322E-05
0.5	0.4997598	0.9999622	2.167327E-05	−2.816349E-06
0.5	0.5000229	1.0	0.0	0.0
0.5	0.4999959	0.9999946	3.612211E-06	−9.387829E-07
0.5	0.5000018	1.0	0.0	0.0
0.5	0.4999995	1.0	0.0	0.0
0.5	0.5	1.0	0.0	0.0
0.5	0.5	1.0	0.0	0.0
0.5	0.5	1.0	0.0	0.0

bridge, as shown in Fig. 11. The technique is applicable for both DAC and power-amplifier applications, where the latter could use the differential output stage. Also symmetry can be regained irrespective of the chosen target alignment via the σ factor, so it is applicable for both $\sigma = 0.5$ and $\sigma = 0.72373$.

3 NOISE SHAPING AND OVERSAMPLING IN PWM

3.1 Single-Stage Noise Shaper

The example presented in Sec. 2 used input data in an essentially nonquantized format, whereby extremely fine resolution is required to position the edges of the PWM transitions. However, if input data are uniformly quantized (such as 16 bit, 44 kHz), then the process should attempt to accommodate quantized data in an efficient way. Since oversampling (typically 4 to 32 times) is commonly used in PWM, there is an opportunity, as previously suggested [11], to use noise shaping to quantize more coarsely each rectangular pulse to below the resolution of the input data. A basic scheme is shown in Fig. 12. It consists of an oversampling filter, a noise shaper possibly with an order of up to 6, and a linearization algorithm.

The noise-shaping process is essentially that presented in an earlier paper [16], where the following subroutine describes the algorithm:

```
REM PWM noise-shaping routine
REM S(T%) input data; X(T%) output data; RN% noise-shaper order; U0% clock
S(T%) = DC + S0*SIN(K0*U0%) + S1*SIN(K1*U0%) + S2*SIN(K2*U0%)
B(0) = S(T%) - DO : DI = 0
```

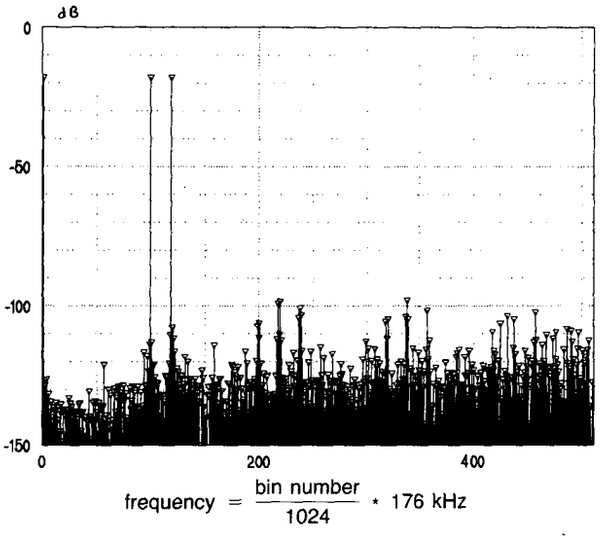


Fig. 9. Corrected PWM spectrum, computed over 1024 samples. Sampling rate 176 kHz; correction band 0–22 kHz; input frequencies 171.875 Hz, 17.1875 kHz, and 20.45313 kHz. (Compare with Fig. 4.)

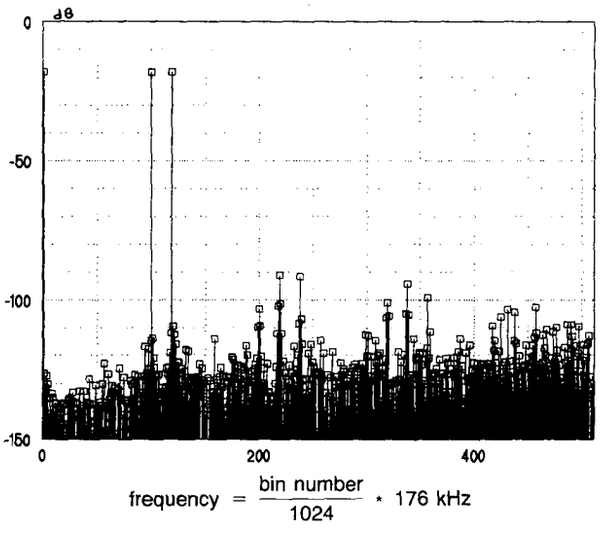


Fig. 10. As Fig. 9, but with optimized σ for a_1 symmetry at $m = 0$ and $m = 1$.

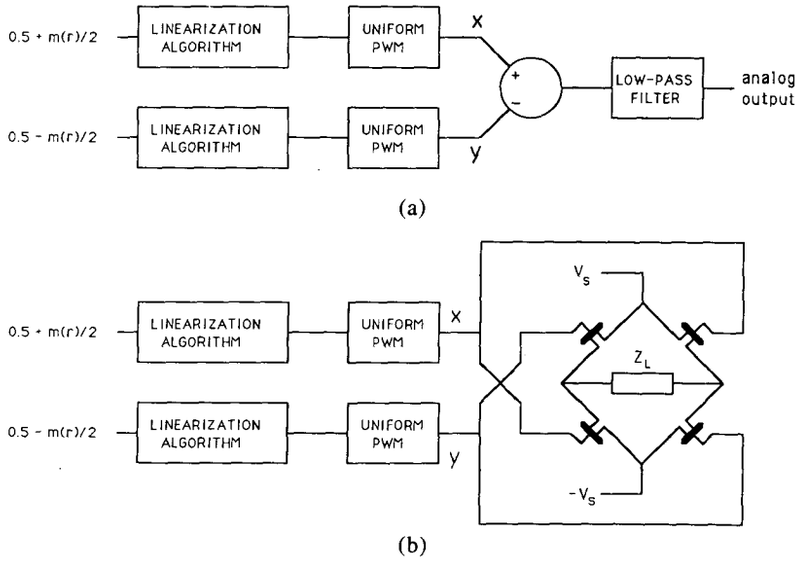


Fig. 11. Differential linearized PWM systems to enhance symmetrical operation for use in high-resolution applications. (a) Configured as DAC. (b) Configured as power amplifier with bridge output stage.

```
FOR L% = 1 TO RN%
B(L%) = B(L%-1) + C(L%) : C(L%) = B(L%)
DI = DI + C(L%)
NEXT
X(T%) = INT(DI) + .5
RETURN
```

The output of the noise shaper, where the word length has been reduced while maintaining an adequate signal-to-noise ratio, can be applied directly to the linearization algorithm described in Sec. 2. However, because the output data of the correction algorithm are quantized more finely than the input data, a requantization process may be used where it is recommended that to reduce additional noise, the output sequence be resolved to either 2 or 3 bit greater than the input sequence, possibly with a further stage of noise shaping, as described in Sec. 3.2.

Alternatively, the PWM signal can be formed using a coarse–fine modulator where the output of the linearization algorithm is initially quantized to the same resolution as the noise-shaper output. The quantization error is then used to fine-tune the edge timing by selecting appropriate taps on a precision delay circuit, as shown in Fig. 13, where for example the clock phase can be controlled to time indirectly the PWM output. In this scheme, if $\Delta\phi$ corresponds to the clock phase determined by the quantization error, then in region A the phase is, say, $\Delta\phi/2$, while in region B it is $-\Delta\phi/2$, thus maintaining symmetry about the sampling instant.

Results of a computer simulation of a fourth-order noise-shaper and linearization system are shown in Figs. 14–17, corresponding to 4, 8, 16, and 32 times oversampling, respectively. In each case the PWM code is quantized to 13 bit and the noise shaper to 10 bit, and the a_1 and a_2 coefficient generators are described by Eq. (12), respectively. As a benchmark, spectral plots are shown for the PWM system, including the noise shaper, both without and with the linearization algorithm, where the noncorrected plots are taken directly from the pulse-width modulator driven by a noise shaper with 10-bit output resolution. The results demonstrate that although noise shaping yields a progressive reduction in intermodulation distortion at higher oversampling ratios, the linearization process is the more effective, although with direct truncation there is a noise penalty.

In computing the Fourier transform of baseband noise-shaped systems over a finite number of samples, the low-level spectral detail at low frequency can be lost. To illustrate this point, the spectral output of a fourth-order, four times oversampled noise-shaped PWM system is shown in Fig. 18, where spectral flattening is evident. The distortion results from the mean level of the finite-analysis sequence being nonzero. Therefore to compensate, one sample is modified so as to force the dc average to zero (or in the special case of our PWM system, to 0.5). In practice, this single sample modifier is small, but it is sufficient to subtract a constant-level spectral distortion, which otherwise corrupts low-level spectral components particularly evident at low frequency with a baseband noise shaper. (Note that a single sample produces a constant spectral response.) The same noise-shaped spectrum, but with a single modified sample to correct the mean value of the sequence, is shown in Fig. 14(a), where the low-level noise-shaped characteristic is now clearly portrayed. This compensation procedure is applied to all the spectral domain plots in this paper, where the justification is that the noise-shaper loop gain is infinite at dc, meaning that an infinite sequence would have a dc value equal to that of the input sequence.

3.2 Two-Stage Noise Shaping

The problem of data-truncation postlinearization can be improved using a second stage of low-order noise shaping to reduce the requantization distortion, where the scheme is shown in Fig. 19. Here the output of the linearization algorithm is requantized using a quantizer enclosed within a first order noise-shaping loop to aid maintenance of resolution prior to driving the PWM.

There is a balance to be made between linearity and additional low-level noise. If the order of the second-stage noise shaper is increased, chaotic behavior [16] makes the output sequence deviate substantially from that of the input sequence, and therefore invalidates the linearization procedure, as the coefficient sets are now no longer accurately matched to the pulse pattern.

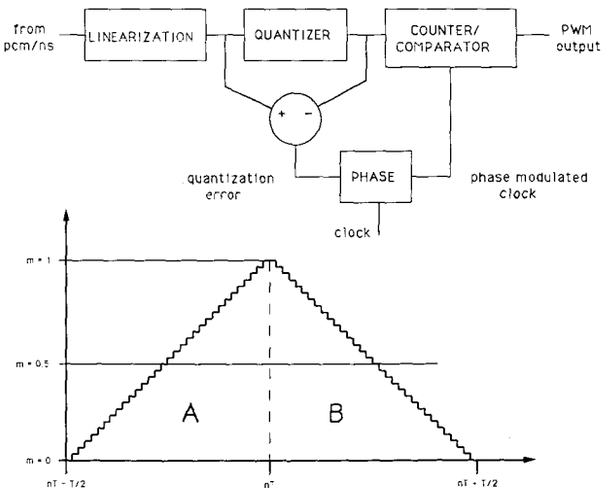
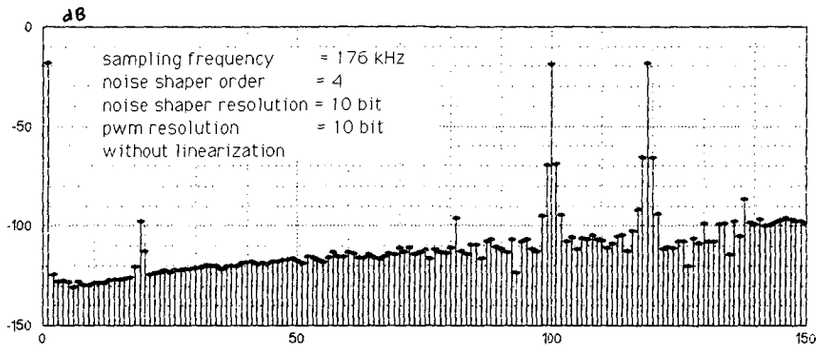


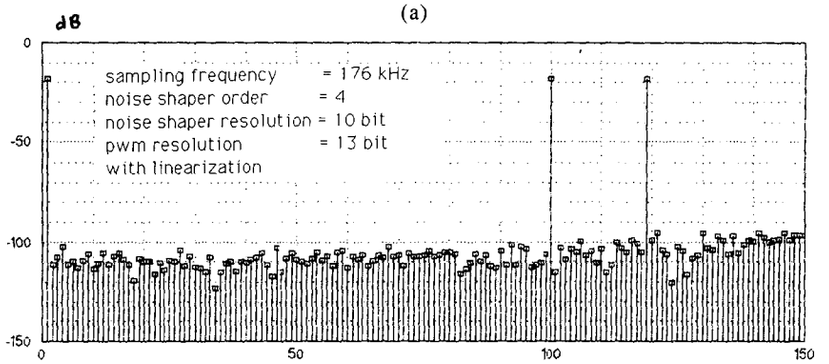
Fig. 13. Double-edge modulation using counter/comparator.

Fig. 12. Uniform sampled PWM with linearization, noise shaping, and data truncation.



$$\text{frequency} = \frac{\text{bin number}}{1024} * 176\ \text{kHz}$$

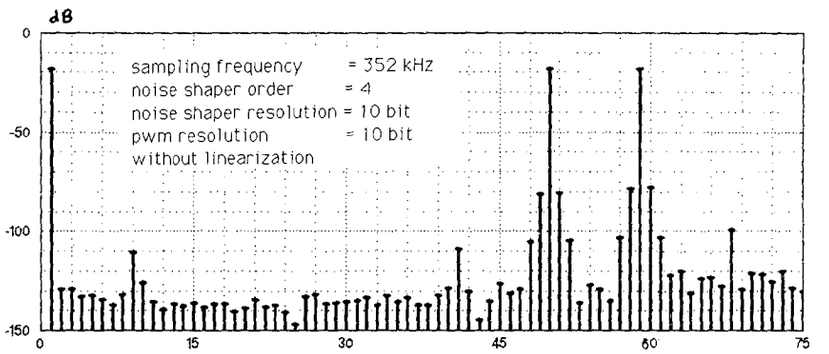
(a)



$$\text{frequency} = \frac{\text{bin number}}{1024} * 176\ \text{kHz}$$

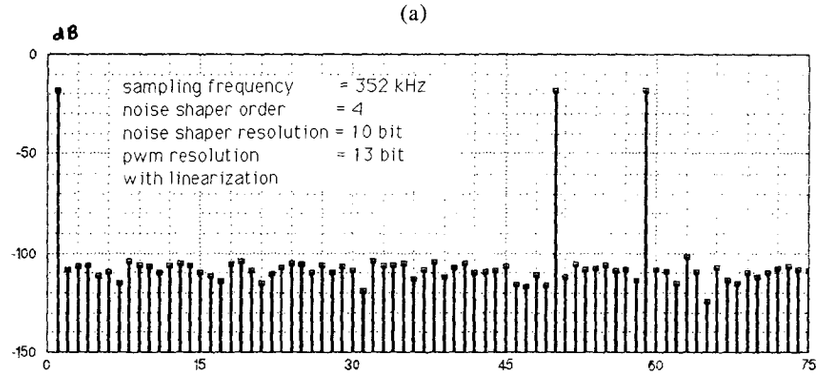
(b)

Fig. 14. Single-stage noise shaping and PWM (a) Without linearization. (b) With linearization and direct truncation.



$$\text{frequency} = \frac{\text{bin number}}{1024} * 352\ \text{kHz}$$

(a)



$$\text{frequency} = \frac{\text{bin number}}{1024} * 352\ \text{kHz}$$

(b)

Fig. 15. Single-stage noise shaping and PWM. (a) Without linearization. (b) With linearization and direct truncation.

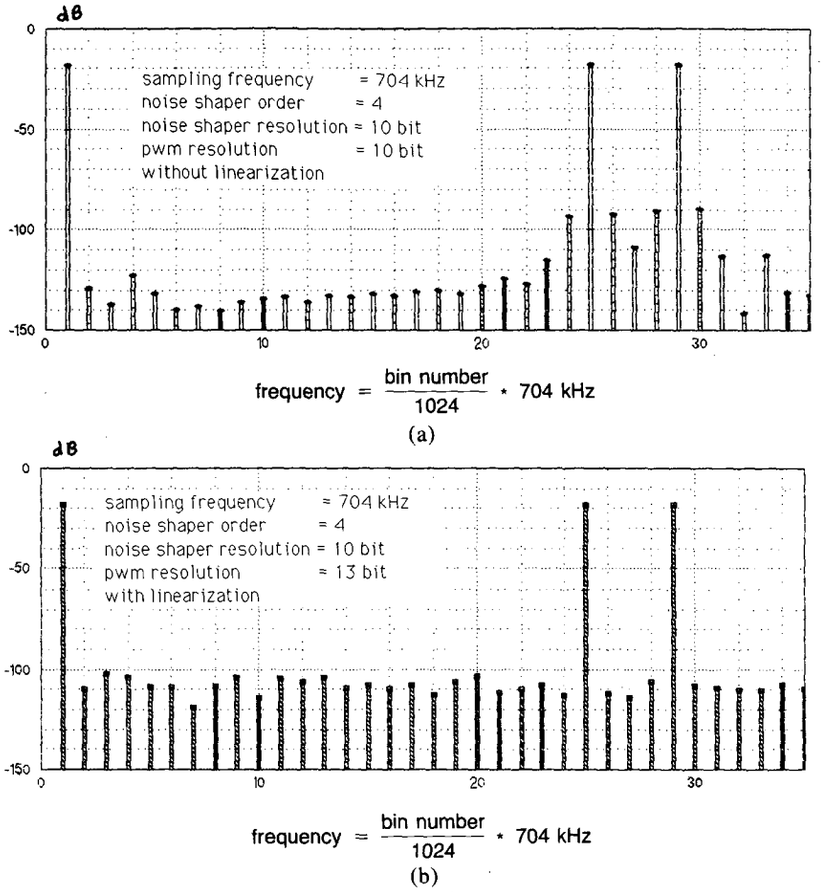


Fig. 16. Single-stage noise shaping and PWM. (a) Without linearization. (b) With linearization and direct truncation.

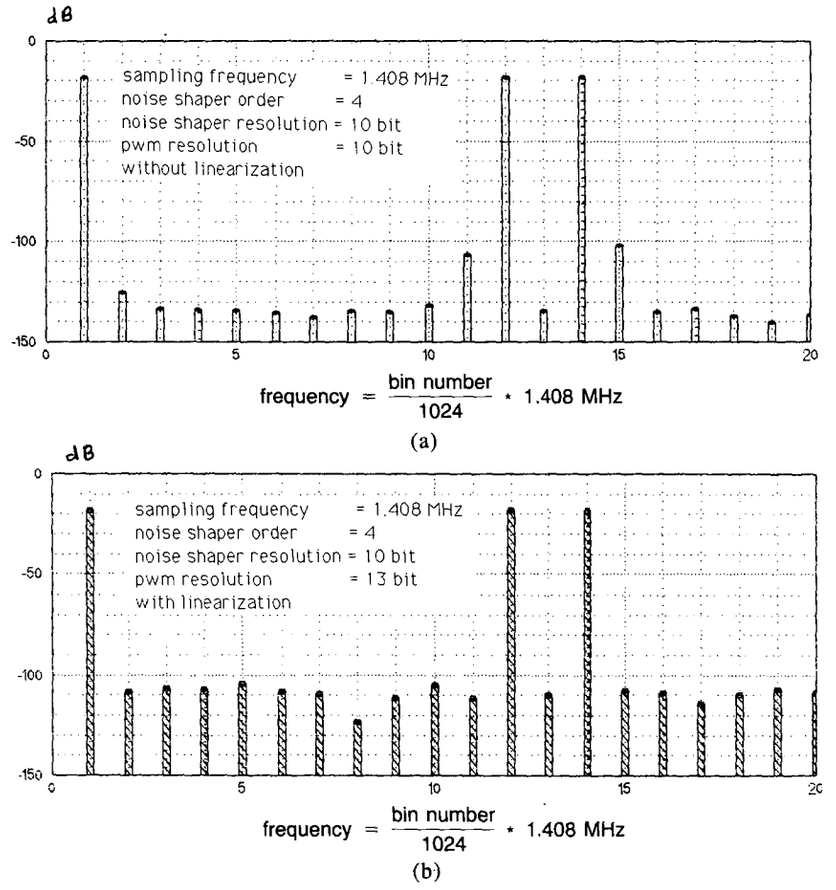


Fig. 17. Single-stage noise shaping and PWM. (a) Without linearization. (b) With linearization and direct truncation.

To compare performance, a set of results of simulations is presented in Figs. 20–23, corresponding to the system data listed in Table 2.

3.3 Pulse Construction Accuracy in PWM Systems

It has been shown that noise shaping can reduce the resolution requirement of the PWM samples, thus easing the problem of pulse construction by requiring coarser steps in pulse width. However, there remains the question of absolute accuracy for each pulse to attain an acceptable overall performance commensurate with, say, a 16-bit 44.1-kHz PCM specification. Such considerations are independent of signal processing prior to PWM modulation and relate to errors introduced after theoretical ideal conversion, where system-dependent error sources can be attributed to the following:

1) Edge jitter and timing errors
2) Slew limiting on pulse edges (meaning pulse area is not proportional to pulse width)
3) Amplitude errors (producing both scale and dynamic error)
4) Processing errors (adding a noise source to the output samples).

These sources contribute to pulse area error which, via dynamic modulation of the spectrum of each pulse including a dc term, produce output noise and distortion. However, in a system where these contributions are small and random we can, to a good approximation, consider an impulse error sequence at the uniform sampling rate f_s Hz, where each impulse weighting equates with an area error in the output pulse. As a method of performance estimation, a system resolution clock f_r Hz is defined, which represents the accuracy to which pulses are required to be timed. Thus if A is the amplitude of a PWM sample, then the pulse area error, hence error impulse weighting, ranges from $-Af_s/2f_r$ to $Af_s/2f_r$, where the PWM sampling rate f_s Hz also represents the error sequence repetition frequency. The error sequence appears as a final system level of requantization distortion where, if treated as a random noise source with a uniform probability distribution function, it contributes an extra noise source of $(Af_s/f_r)^2/12$, distributed over the half-sampling band 0 to $f_s/2$ Hz. Hence, assuming a uniform spectral power density, the noise power N_a within the half-band 0 to $f_n/2$ Hz (where f_n is the Nyquist sampling frequency, $f_n = 44.1$ kHz) is

$$N_a = \frac{A^2}{12}\frac{f_s f_n}{f_r^2} . \tag{16}$$

As a benchmark, consider an N-bit PWM system sampled at the Nyquist rate f_n Hz, where the system clock $f_r = f_n 2^N$. Then the reference in-band noise N_{ref} is

$$N_{ref} = \frac{A^2}{12}\frac{f_n^2}{(f_n 2^N)^2} = \frac{A^2}{12.2^{2N}}$$

whereby

$$\frac{N_a}{N_{ref}} = \frac{f_s f_n}{f_r^2}\, 2^{2N} .$$

To maintain a constant audio-band noise performance

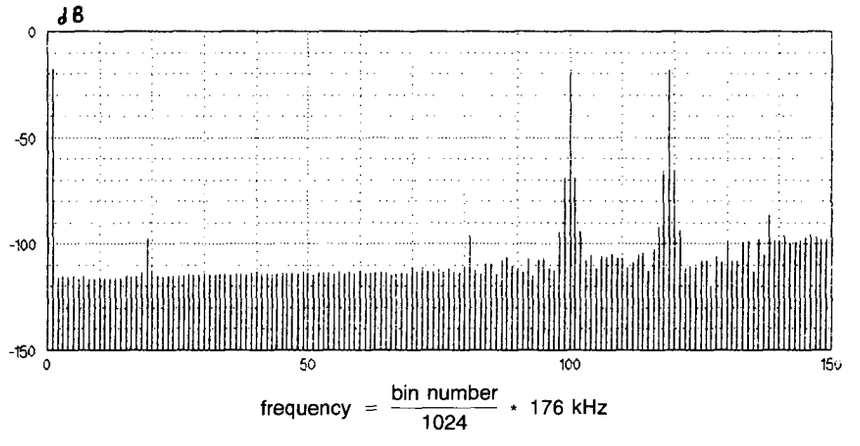


Fig. 18. PWM/noise-shaped spectrum illustrating spectral flattening. [See also Fig. 14(a).]

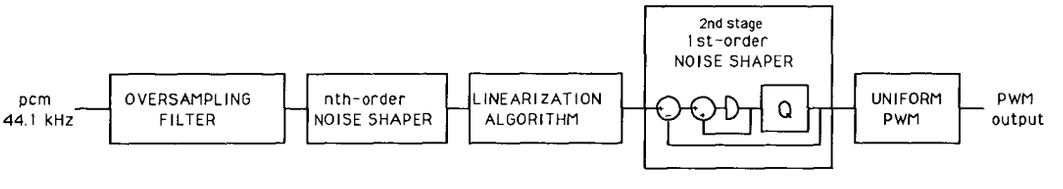
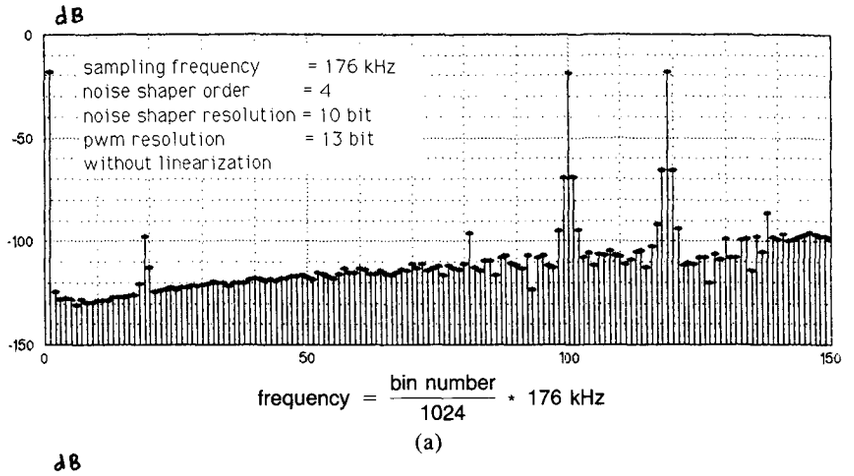
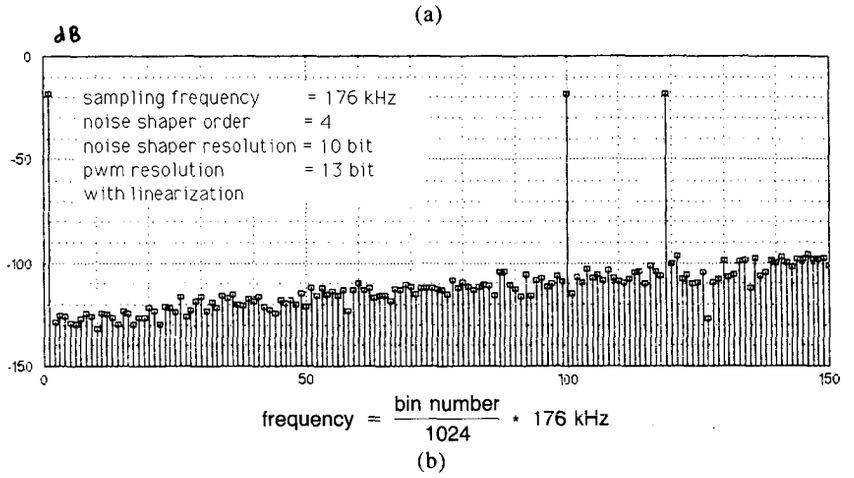


Fig. 19. Uniform sampled PWM with linearization and two stages of noise shaping.

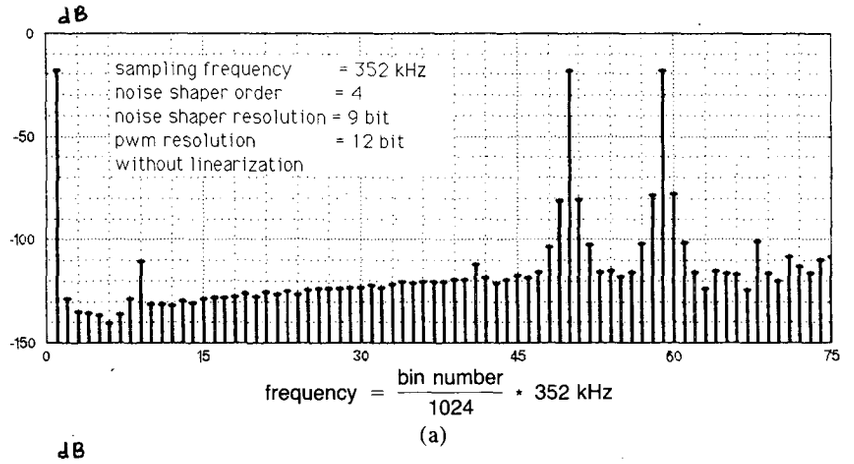


(a)

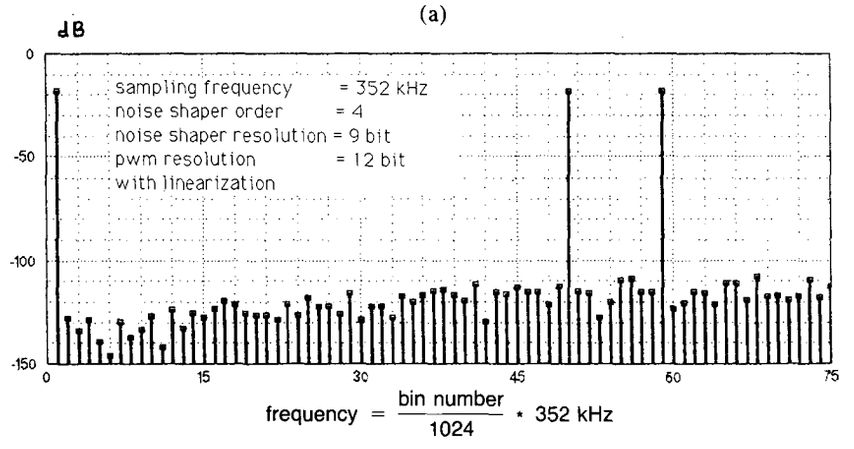


(b)

Fig. 20. Two-stage noise shaping and PWM. (a) Without linearization. (b) With linearization and first-order noise-shaped truncation prior to PWM.



(a)



(b)

Fig. 21. Two-stage noise shaping and PWM. (a) Without linearization. (b) With linearization and first-order noise-shaped truncation prior to PWM.

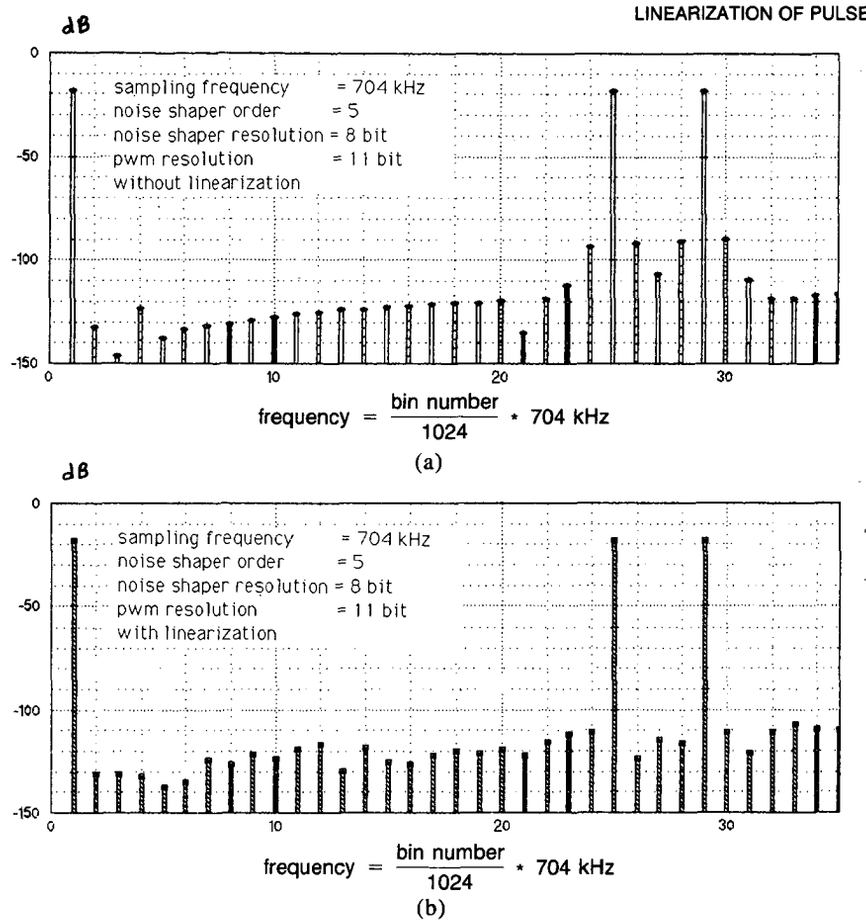


Fig. 22. Two-stage noise shaping and PWM. (a) Without linearization. (b) With linearization and first-order noise-shaped truncation prior to PWM.

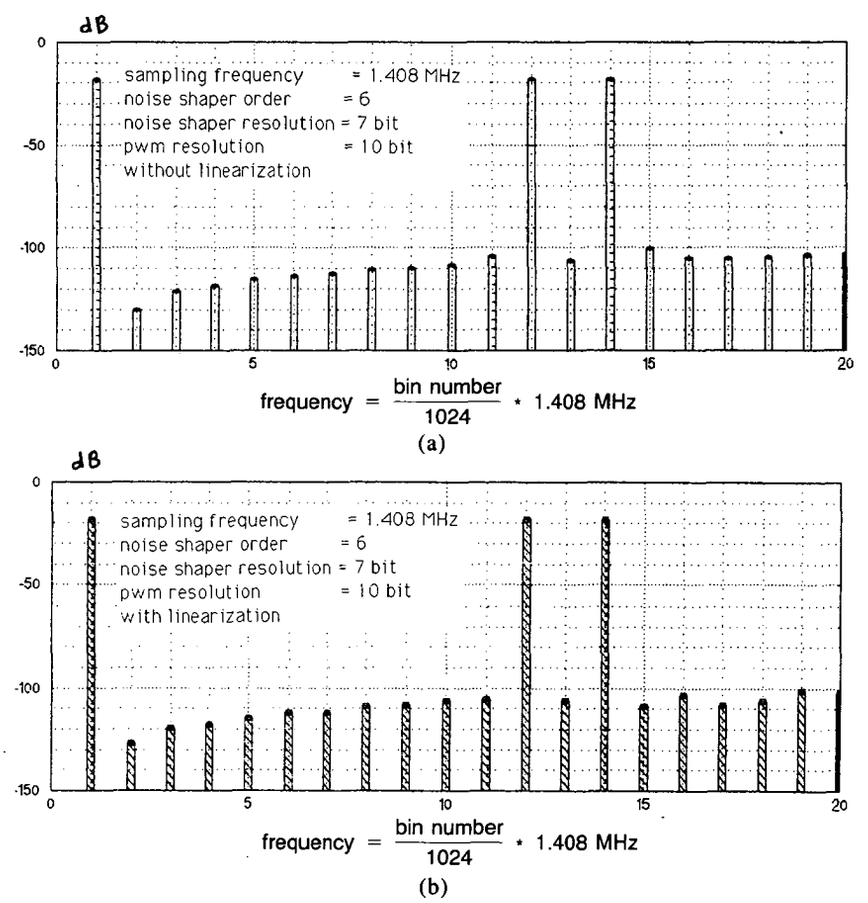


Fig. 23. Two-stage noise shaping and PWM. (a) Without linearization. (b) With linearization and first-order noise-shaped truncation prior to PWM.

with sampling rate, make $N_a/N_{ref} = 1$, that is,

$$f_r = 2^N(f_s f_n)^{1/2} . \quad (17)$$

Eq. (17) suggests that if the system dynamic range is to be maintained, even if noise shaping is used, to reduce pulse resolution, then system accuracy must actually increase in proportion to $(f_s)^{1/2}$. Consequently noise shaping can allow a more coarse quantization of pulse-width resolution, but the absolute accuracy of the timing and amplitude stability must be maintained or, indeed, increased. This is an important system design consideration, especially where the linearization algorithm described in Sec. 2 allows enhanced linearity at lower sampling rates.

4 CONCLUSION

A method of linearizing uniform sampled PWM has been described, where excellent linearity can be achieved at moderate sampling rates. Although numerous results can be computed to show the variation of distortion (for example, % total harmonic distortion) as a function of signal level and frequency, we recognize that for PWM, high-frequency, high-amplitude input signals are a particularly severe test, a fact adequately demonstrated in much of the supporting literature. Consequently the results presented here have been restricted to to a high-level three-tone equiamplitude intermodulation test, where simulations demonstrate the correction procedure to result in almost zero distortion residues within the audio band, including both intermodulation components and harmonics of the fundamental.

The nonlinear, model-based process was then extended to include noise shaping in order to reduce the resolution of the PWM code and therefore facilitate PWM modulator implementation. However, for an effective noise-shaping advantage, the sampling rate needs to be a minimum of four times, with a preference for higher rates. Also the linearization algorithm can lead to requantization, which must be controlled if degradation in dynamic range is to be minimized.

A number of computer simulations were presented that demonstrate the effectiveness of linearization for various system parameters. It was also shown that provided the optimization band is maintained at 0–22 kHz, the magnitude of the dispersive, corrective signal diminishes as the sampling rate is increased where there was greater opportunity for noise shaping to reduce the resolution of the output code, although with higher sampling rates there is some advantage in extending the optimization frequency band. However, an increase in sampling rate also implies more pulse transitions per second, reflecting a reduction in amplifier power efficiency. Also, pulse jitter, slew rate, and amplitude-related distortions, including the effects of nonideal power supplies, will limit the resolution of each pulse and thus degrade the dynamic range, implying that the promise of a very high signal-to-noise ratio with low distortion may prove illusive.

An interesting comparison therefore emerges between a noise-shaped system with a low pulse resolution requirement but higher sampling rate and a non-noise-shaped system using lower sampling rates but with a higher sample resolution requirement. It is important to observe that noise shaping was shown in Sec. 3.3 not to be a method of reducing accuracy, only resolution, where the accuracy requirement at a given sampling rate is similar for both a noise-shaped and a non-noise-shaped system, with actually a more stringent specification required at the higher sampling rates. The linearization algorithm means that uniform sampling can be used, which yields a simpler implementation of the pulse-width modulator. Although a high clock rate is required for pulse timing, the sampling rate remains modest, leading to good power efficiency. In earlier systems much higher sampling rates were reported with the uniform sampling process in order to achieve an adequate high-frequency intermodulation distortion performance.

The results presented in this paper suggest that although noise-shaped systems offer advantage, a direct-conversion PWM system using linearization is also a serious candidate, provided the pulse timing can be accommodated, and should therefore find applications in both power amplification and precision DAC systems. Clearly for high-power PWM amplifier applications there are a number of associated problems to be addressed in terms of EMC and output-stage design, although regular advances are reported and the potential for reduced sampling rates through linearization are of significance when considering output-stage efficiency. However, the output stage itself can introduce distortion due to edge jitter and edge rise time, and these in association with power-supply compliance can still generate distortion, although this is now an instrumental problem rather than a fundamental limitation. The development of highly linear digital power amplifiers is of major significance and will have wide application in digital audio systems, including digital and active loudspeaker technology.

Table 2. System data.

Figure	Loop order	NS resolution	PWM resolution	PWM sampling frequency (kHz)
20	4	10	13	176
21	4	9	12	352
22	5	8	11	704
23	6	7	10	1408

However, it is also suggested that the low-level PWM system is an attractive alternative to other forms of DAC, especially as with correct pulse edge timing it can exhibit perfect low-level linearity (commensurate with a quantized data sequence). Once linearized, PWM can be considered the dual of multilevel, oversampled DAC technology, where multilevel constant width (that is, PCM) is translated to constant-level, multiwidth coding or, indeed, other intermediate combinations of multilevel and multiwidth formats. It may well prove easier to achieve optimum resolution using pulse-edge timing with a two-level signal than the inherent complexities of designing multilevel DACs with their accuracy problems and myriad of interlevel signal combinations and associated slew-rate and pulse-area modulation errors.

The PWM DAC with reduced sampling rate is also a viable alternative to the widely used bit stream converter [17] and earlier reported look-up table techniques [18]. Although edge resolution is now more stringent, the considerable reduction in the number of edge transitions per second, typically a factor of 64, suggests lower jitter noise and associated high-frequency spuriae. There is also the potential for enhanced low-level resolution with no idle-channel artifacts. Furthermore, as a point of circuit detail, linearized PWM converters do not require either transimpedance (current-to-voltage) amplifiers, particularly significant in oversampled systems, or switched-capacitor filters, either of which can represent additional sources of noise and nonlinear distortion. It is therefore proposed that the linearized PWM DAC can represent an optimum conversion strategy that should also prove well matched to signals that include psychoacoustic noise shaping [19] to enhance overall system dynamic range.

Future research will consider the feasibility of greater reductions in the sampling rate to merge more closely the duality between PWM and PCM. Also methods of implementing the iterative linearization procedures using digital processors will be considered in order to enable real-time performance and subjective measures to be achieved. However, the extension to a fully functional power amplifier system is the main application, even though EMC criteria are severe and the need to control output-stage distortion is crucial to realizing the optimum performance of digital audio signals.

5 REFERENCES

[1] A. H. Reeves, French patent 852, 183 (1938).

[2] A. H. Reeves, British patent 535, 860 (1939).

[3] A. H. Reeves, U.S. patent 2,272,070 (1942).

[4] S. Kashiwagi, "A High Frequency Audio Power Amplifier Using a Self-Oscillating Switching Regulator," 0093-9994/85/0700-0906 *IEEE Int. Appliance Ind. Conf.* (1984 May).

[5] J. Hancock, "A Class D Amplifier Using MOSFETs with Reduced Minority Carrier Lifetime," *J. Audio Eng. Soc.*, vol. 39, pp. 650–662 (1991 Sept.).

[6] M. B. Sandler, "Digital Power Amplifier Design," Ph.D. dissertation, University of Essex, Colchester, UK (1983).

[7] M. B. Sandler, "Toward a Digital Power Amplifier," presented at the 76th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 32, p. 1009 (1984 Dec.), preprint 2135.

[8] M. B. Sandler, "Progress toward a Digital Power Amplifier," presented at the 80th Convention o the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 34, p. 378 (1986 May), preprint 2361.

[9] M. B. Sandler, "Techniques for Digital Power Amplification," *Proc. Inst. Acoustics* (Reproduced Sound 3, Windermere, 1987).

[10] J. Goldberg and M. B. Sandler, "Comparison of PWM Modulation Techniques for Digital Power Amplifiers," *Proc. Inst. Acoustics* (Reproduced Sound 6), vol. 12, pt. 8, pp. 57–65 (1990 Nov.).

[11] M. O. J. Hawksford, "*N*th-Order Recursive Sigma-ADC Machinery at the Analog-Digital Gateway," presented at the 78th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 33, pp. 58–87 (1985 July/Aug.), preprint 2248.

[12] R. G. Bowman, R. E. Hiorns, and M. B. Sandler, "Design and Performance of a Noise Shaping, Pulse Width Modulated, Digital to Analogue Converter," presented at the IEE Colloquium on Digital Audio Signal Processing, London (1991 May).

[13] J. M. Goldberg and M. B. Sandler, "Noise Shaping and Pulse-Width Modulation for an All-Digital Audio Power Amplifier," *J. Audio Eng. Soc.*, vol. 39, pp. 449–460 (1991 June).

[14] P. H. Mellor, S. P. Leigh, and B. M. G. Cheetham, "The Implementation and Performance Enhancement of a Completely Digital Power Amplifier," *Proc. Inst. Acoustics* (Reproduced Sound 6), vol. 12, pt. 8, pp. 67–75 (1990 Nov.).

[15] P. H. Mellor, S. P. Leigh, and B. M. G. Cheetham, "Improved Sampling Process for a Digital Pulse-Width Modulated Class D Power Amplifier," presented at the IEE Colloquium on Digiital Audio Signal Processing, London (1991 May).

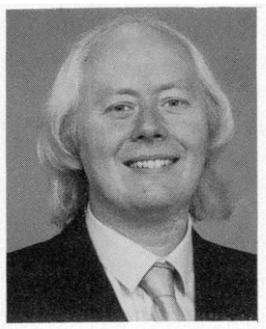
[16] M. O. J. Hawksford, "Chaos, Oversampling, and Noise Shaping in Digital-to-Analog Conversion," *J. Audio Eng. Soc.*, vol. 37, pp. 980–1001 (1989 Dec.).

[17] P. J. A. Naus, E. C. Dijkmans, E. F. Stikvoort, A. J. McKnight, D. J. Holland, and W. Bradinal, "A CMOS Stereo 16 Bit D/A Converter for Digital Audio," *IEEE J. Solid-State Circuits*, vol. SC-22, pp. 390–394 (1987 June).

[18] M. O. J. Hawksford, "Multi-Level to 1 Bit Transformations for Applications in Digital-to-Analogue Converters Using Oversampling and Noise Shaping," *Proc. Inst. Acoustics*, vol. 10, pp. 129–143 (1988 Nov.).

[19] R. A. Wannamaker, "Psycho-Acoustically Optimal Noise Shaping," presented at the 89th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 33, p. 871 (1990 Nov.), preprint 2965.

THE AUTHOR



Malcolm Omar Hawksford is a reader in the Department of Electronic Systems Engineering at the University of Essex, where his principal interests are in the fields of electronic circuit design and audio engineering. Dr. Hawksford studied at the University of Aston in Birmingham and gained both a First Class Honors B.Sc. and Ph.D. The Ph.D. program was supported by a BBC Research Scholarship, where the field of study was the application of delta modulation to color television and the development of a time compression/time multiplex system for combining luminance and chrominance signals. Since his employment at Essex, he has established the Audio Research Group, where research on amplifier studies, digital signal processing, and loudspeaker systems has been undertaken. Since 1982 research into digital crossover systems has begun within the group and, more recently, oversampling and noise shaping investigated as a means of analog-to-digital/digital-to-analog conversion.

Dr. Hawksford has had several AES publications that include topics on error correction in amplifiers and oversampling techniques. His supplementary activities include writing articles for *Hi-Fi News* and designing commercial audio equipment. He is a member of the IEE, a chartered engineer, a fellow of the AES and of the Institute of Acoustics, and a member of the review board of the *AES Journal*. He is also a technical adviser for *HFN* and *RR*.

Linearization of Multilevel, Multiwidth Digital PWM with Applications in Digital-to-Analog Conversion*

MALCOLM OMAR HAWKSFORD, *AES Fellow*

Department of Electronic Systems Engineering, University of Essex, Colchester CO4 3SQ, UK

The process of uniformly sampled pulse-width modulation is investigated as a means of high-performance digital-to-analog conversion where comparisons of duality are made with signals in standard modulated format. Emphasis is placed on using linearization algorithms to suppress nonlinear distortion products, and a new digital-to-analog conversion architecture is proposed that employs digital linearization together with a multiwidth and multiamplitude signal format configured within a current-steering autocalibration system topology.

0 INTRODUCTION

Research [1]–[3] has shown that partial linearization of uniformly sampled pulse-width modulation (PWM) can be achieved by introducing a modified data sequence that takes account of the dynamic distortion inherent within the spectral domain when the width of a pulse is modulated. An oversampling factor of times 4 was used both to relax the design of the dynamic bidirectional recursive compensation and to allow latitude for low-order analog filters to achieve signal reconstruction in digital-to-analog conversion.

In this paper we explore the possibility of using lower orders of oversampling with the aim of implementing a low-distortion PWM converter that can operate at twice the Nyquist rate of a typical digital audio system. The advantage of lower sampling rates is a reduction in the number of pulse transitions per second and a corresponding lowering of the system clock rate in situations that do not use noise shaping. Also, a near-Nyquist sampled PWM system can be considered the dual of the near-Nyquist sampled pulse-code modulated (PCM) data. The former maintains constant amplitude but variable width, whereas the latter has constant width and variable amplitude.

The method of linearizing near-Nyquist sampled PWM is extended to a more general pulse format using both multilevel and multiwidth structures. This both reduces the resolution required of the PWM component and lowers the pulse timing clock rate. The technique of combining multiwidth and multilevel pulses enables a parallel digital-to-analog conversion (DAC) architecture to be implemented, a system we designate a flash DAC (FDAC). It is shown that an FDAC can be synthesized from an array of identical fast current-switching cells, and because of the low number of pulse transitions per Nyquist sample, the system offers low jitter noise and slew-rate distortion. Finally a calibration procedure is introduced for the FDAC, which, in association with a stable low-jitter clock source and fast logic circuits, offers a further candidate for a high-quality DAC for use in digital audio systems.

1 DESIGN OF STATIC COMPENSATION FILTER

It has been shown [3], [4] that nonlinearity in uniformly sampled PWM arises because of variations in the spectral shape within the audio band when the width of the PWM samples changes. Thus instead of the overall spectrum being scaled in proportion to the area under a pulse, modulation of pulse shape means that only the dc term is proportional to the pulse area while other regions of the spectrum undergo varying gain changes. A fundamental requirement of any DAC is that the transfer function of all samples remain in exact proportion and only undergo uniform scaling with the signal level, a condition readily met by rectangular PCM samples of constant width but varying level. This is a theoretical ideal where in practice slew-rate limiting, pulse jitter, and a nonconstant pulse crest can lead to dynamic spectral errors and are factors that contribute to nonideality in the DAC process.

* Presented at the 97th Convention of the Audio engineering Society, San Francisco, CA, 1994 November 10–13.

In Fig. 1 a rectangular pulse is shown centered on a sampling instant where the overall width can vary from 0 to $1/f_{ns}$, f_{ns} being the Nyquist sampling frequency in hertz. Also shown is a normalized family of spectral domain plots for a rectangular pulse of varying pulse width computed over the band of 0 to f_{ns} Hz. This result is computed directly from the Fourier transform of a normalized rectangular pulse $F_p(f)$ of width m/f_{ns}, where m is the modulation depth, $0 < m < 1$,

$$F_p(f) = \frac{\sin(m\pi f/f_{ns})}{m\pi f/f_{ns}} . \quad (1)$$

The dc component of $F_p(f)$ is normalized to unity for all m.

Next we consider the design of a uniformly sampled digital finite impulse response (FIR) filter that can compensate for the spectral error within the band of 0 to $f_{ns}/2$ Hz. However, unlike the examples considered earlier [3], the correction band now extends to approximately one-half of the sampling frequency. Therefore to achieve similar correction accuracy, an increase in the number of taps in the FIR filter is required. Also, to reduce the magnitude of the correction coefficients, a target transfer function $\{\sin(m_t\pi f/f_{ns})/(m_t\pi f/f_{ns})\}$ is used, which, in the first instance, corresponds to a rectangular pulse of duration $1/2f_{ns}$. Here the target modulation index $m_t = 0.5$, although for the multilevel case a value of $m_t = 1.0$ is preferred as this width is dominant at higher signal levels. The equalization transfer function $E(m, f)$ over the band of 0 to $f_{ns}/2$ Hz follows from Eq. (1),

$$E(m,f) = \frac{m}{m_t}\frac{\sin(m_t\pi f/f_{ns})}{\sin(m\pi f/f_{ns})} \quad \text{matched over 0 to } f_{ns}/2 \text{ Hz} . \quad (2)$$

The impulse response of the FIR correction filter can be obtained from the inverse Fourier transform of $E(m, f)$. However, this process is approximate as the number of coefficients in the FIR filter is finite and because spectral replication causes a peak of $f_{ns}/2$ Hz, as shown in Fig. 2. Spectral replication about the sampling frequency and its harmonics means that correction cannot be applied to frequencies $\geq f_{ns}/2$. Also, modulation products are reflected back into the baseband and prevent a total cancellation of distortion.

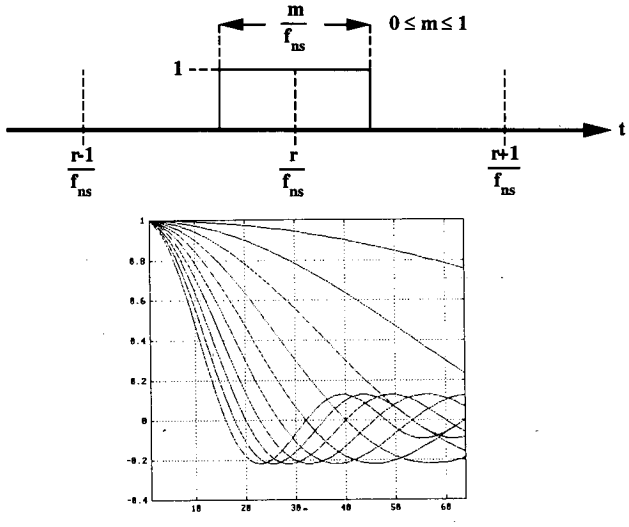


Fig. 1. Pulse of normalized width 1 together with family of corresponding transforms.

1.1 Window Function

The first proposal for the approximation process is to introduce a guard band of width f_g Hz, consisting of a frequency and amplitude scaled, reflected image of the response within the acceptance band to form a smooth overall function with no discontinuities in the first derivative. This strategy reduces the higher order terms of the signal within the transform signal space, which in this case yields less time dispersion in the nonrecursive equalization filter. The construction of the overall function is shown in Fig. 3, where f_0 Hz is the transition frequency defining the boundary between acceptance band and guard band and is defined as

$$f_0 = \frac{f_{ns}}{2} - f_g . \quad (3)$$

The function within the acceptance band of 0 to f_0 Hz is given by Eq. (2). Its value A_0 at $f = f_0$ follows as

$$A_0 = \frac{m}{m_t}\frac{\sin[m_t\pi(0.5 - f_g/f_{ns})]}{\sin[m\pi(0.5 - f_g/f_{ns})]} . \quad (4)$$

To form the scaled and reflected function $A(m, f)$ in the guard band that smoothly interfaces with $E(m, f)$, the function for $E(m, f)$ is scaled by a factor k, then offset and reflected about $f_{ns}/2$ to give

$$A(m,f) = \frac{1}{k}\left\{A_0(1 + k) - \frac{m}{m_t}\frac{\sin[m_t\pi k(0.5 - f/f_{ns})]}{\sin[m\pi k(0.5 - f/f_{ns})]}\right\} . \quad (5)$$

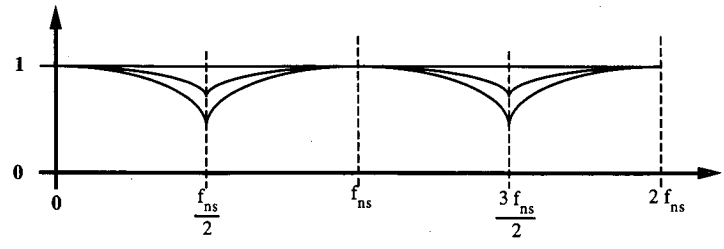


Fig. 2. Transfer function $E(f)$ of digital FIR correction filter for $m = 0$, 0.5, and 1.

The scaling factor k is determined by equating the arguments of the sine functions in Eqs. (2) and (5) at $f = f_0$,

$$k = \frac{f_{ns}}{2f_g} - 1 . \quad (6)$$

Examination of Eq. (5) shows that $A(m, f_0) = A_0$ and that differentiation of $E(m, f)$ and $A(m, f)$ reveals a continuous functional form over the transition frequency f_0.

Consider a symmetrical transversal filter with coefficients $\{a_0, a_1, \ldots, a_n\}$, where the transfer function $F_c(f)$ is given by

$$F_c(f) = a_0 + 2\sum_{r=1}^{n} a_r \cos\left(\frac{2\pi rf}{f_{ns}}\right) . \quad (7)$$

The FIR filter design requires a transfer function that is matched to the composite equalization filter response $E_c(f)$, where

$$E_c(m,f) = E(m,f) + A(m,f) . \quad (8)$$

Here $E(m, f)$ is valid for $0 < f < f_0$, whereas $A(m, f)$ is valid for $f_0 < f < 0.5f_{ns}$.

The coefficient set $\{a\}$ can be determined by forming a set of simultaneous equations to match the windowed Fourier transform of a rectangular pulse corresponding to a modulation depth m (where $0 < m < 1$) with that of the finite data sequence of the FIR filter. Assuming NC independent coefficients in a symmetrical FIR filter, then using a uniform frequency distribution where $f = if_{ns}/2NC$ and defining a matrix element $Y(r, i)$, where

$$Y(r, i) = 2\cos\left(\frac{\pi ri}{NC}\right) \quad (9a)$$

the matrix equation is defined as

$$[E_c] = [Y][a] . \quad (9b)$$

Inversion of Eq. (9b) produces the NC coefficients where, to preserve the dc content of the input data sequence, the central coefficient a_0 is given by

$$a_0 = 1 - 2\sum_{r=1}^{n} a_r . \quad (10)$$

For the special case where the sampling rate is $2f_{ns}$, the guard band f_g can be set to $f_{ns}/2$, which relaxes the filter design. By following the procedure discussed in Section 1.1, the scaling factor $k = 1$ and the target response show odd symmetry about $f_{ns}/4$ Hz. Consequently, even-order terms in the equalization filter are zero, and the filter can be compared with a half-band filter where alternate samples are zero.

1.2 Scaling of Frequency-Domain Sampling

An alternative approach to the estimation of coefficients is to concentrate the frequency domain sampling over a subset of the Nyquist band. Hence if there are NC unknown coefficients, the NC samples are focused into the relevant frequency space, thus improving the approximation in that region.

The scaling is achieved by introducing a factor μ into the arguments of Eqs. (2) and (9a),

$$E(m,f) = \frac{m}{m_t}\frac{\sin(m_t\mu\pi f/f_{ns})}{\sin(m\mu\pi f/f_{ns})}$$

matched over 0 to $\mu f_{ns}/2$ Hz (11)

$$Y(r, i) = 2\cos\left(\frac{\mu\pi ri}{NC}\right) . \quad (12)$$

A similar analysis using Eq. (9b) is then used to determine the transversal filter coefficients $\{a\}$.

Fig. 4 shows example approximation functions for filters using four coefficients for $\mu = 0.5$ and 0.9. Equalization characteristics are computed for discrete m in the range $0 \leq m \leq 1$ in steps of 0.1. The results show that by reducing the upper frequency limit of equalization, improved in-band approximations result as the frequency domain sampling is concentrated into a narrower bandwidth. For $\mu = 0.9$ the in-band approximation is less accurate.

Although coefficient sets can be calculated for each value of m, this is computationally inefficient. Thus two methods are proposed:

1) A look-up table is computed for coefficient values against m and linear interpolation performed for intermediate values. Eq. (10) is always used to ensure that there is no overall loss of pulse area.

2) Polynomial approximations are estimated from the discrete m values following a similar procedure reported in [3].

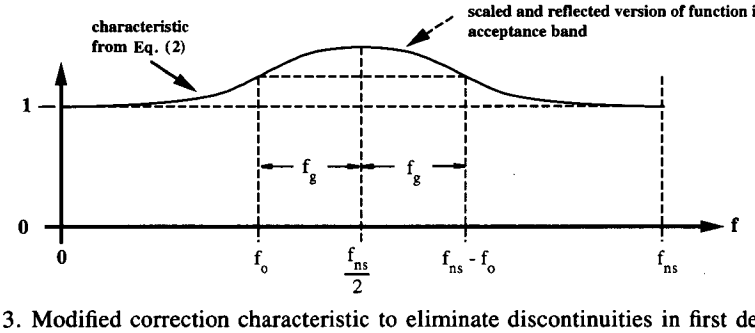


Fig. 3. Modified correction characteristic to eliminate discontinuities in first derivative.

2 DYNAMIC FILTER SIMULATION FOR BINARY PWM

The filter design procedures presented in Sections 1.1 and 1.2 show how a filter coefficient set can be calculated for a particular m value of input data. However, in a general PCM sequence ip(L), each pulse can take any value within the data range and requires individual correction for the PCM-to-PWM mapping. Effectively, a unique FIR filter is required for each input data sequence, where for sample L a coefficient set (of NC independent coefficients) is determined and is represented using the notation $a(x, y)$ as follows:

Coefficient Set for Sample L

Sample L	$a(L, \text{NC}), a(L, \text{NC}-1), \ldots a(L, 1), \ldots a(L, \text{NC}-1), a(L, \text{NC})$

The filter structure is shown in Fig. 5, where the coefficients are not constant but are uniquely associated with each sample value. Thus as the samples are shifted through the FIR filter, the coefficients are also shifted through the individual coefficient sets. In Fig. 6 the filter process is extended to show the computation of three adjacent output samples op(L), op($L + 1$), and op($L + 2$), corresponding to input samples ip(L), ip($L + 1$), and ip($L + 2$).

As reported earlier [3], the calculation of coefficients requires iterative adaptation as the output samples are dependent on both the present sample and contributions from adjacent FIR filters. In this paper the "single-sided, unidirectional shifting algorithm" is modified for use with NC coefficients and is illustrated using matrix notation. For NC = 2,

$$\begin{bmatrix} a(1,3) & 0 & 0 & 0 & 0 & 0 \\ a(1,2) & a(2,3) & 0 & 0 & 0 & 0 \\ a(1,1) & a(2,2) & a(3,3) & 0 & 0 & 0 \\ a(1,2) & a(2,1) & a(3,2) & a(4,3) & 0 & 0 \\ a(1,3) & a(2,2) & a(3,1) & a(4,2) & a(5,3) & 0 \\ 0 & a(2,3) & a(3,2) & a(4,1) & a(5,2) & a(6,3) \\ 0 & 0 & a(3,3) & a(4,2) & a(5,1) & a(6,2) \\ 0 & 0 & 0 & a(4,3) & a(5,2) & a(6,1) \\ 0 & 0 & 0 & 0 & a(5,3) & a(6,2) \\ 0 & 0 & 0 & 0 & 0 & a(6,3) \end{bmatrix} \begin{bmatrix} \text{ip}(1) \\ \text{ip}(2) \\ \text{ip}(3) \\ \text{ip}(4) \\ \text{ip}(5) \\ \text{ip}(6) \\ \text{ip}(7) \\ \text{ip}(8) \\ \text{ip}(9) \\ \text{ip}(10) \end{bmatrix} = \begin{bmatrix} \text{op}(1) \\ \text{op}(2) \\ \text{op}(3) \\ \text{op}(4) \\ \text{op}(5) \\ \text{op}(6) \\ \text{op}(7) \\ \text{op}(8) \\ \text{op}(9) \\ \text{op}(10) \end{bmatrix}.$$

Matrix [a] is initialized such that all coefficients are zero except for elements $a(x, 1) = 1$. Computation starts by calculating op(3) and the corresponding coefficient set $\{a(3, 1), a(3, 2), a(3, 3)\}$, as described previously, where upon the new elements take their places in matrix [a]. The procedure is repeated for the next row and so on throughout the matrix, forming, for this example, an output op(7). The input vector [ip] is then shifted and the [a] matrix diagonally shifted by $a(x + 1, y + 1) = a(x, y)$ and $a(1, 1) = 1$. The procedure then repeats to determine the next output sample. The iterative procedure accommodates the nonlinear interaction between samples due to the PWM process and allows the coefficient values to converge to a stable solution, although in practice more rows in the [a] matrix may be desirable.

To validate the procedure for calculating coefficients for use with binary PWM, a simulation program was run over NS = 2048 samples, where $m_t = 1$ and the input signal is dc biased at $x = 0.5$,

$$x(L) = 0.5 + 0.3\,[\sin(k_0 f_{ns} t) + \sin(k_1 f_{ns} t) + \sin(k_2 f_{ns} t)] + \text{rnd} * 2^{-17}. \quad (13)$$

Here $k_0 = 2\pi/\text{NS}$, $k_1 = 300k_0$, $k_2 = 310k_0$ and rnd has a uniform probability density function from -0.5 to 0.5 and $t = L/f_{ns}$ for $\leq L \leq$ NS. Fig. 7 shows computed output spectra for the PWM process with and without correction for $\mu = 0.5$ and NC = 6.

3 MULTILEVEL DAC WITH PWM INTERPOLATION

Direct DAC using PWM requires a high clock rate to time the pulse transitions. For example, 16-bit resolution at a 44.1-kHz sampling rate corresponds to a PWM clock of 2.89 GHz. Although this rate is high, when the timing and jitter performance of a high-quality DAC are considered, this rate is not unreasonable, and fast-counting circuits suggest that economic realization will be feasible with clocks offering subnanosecond jitter.

However, although techniques now exist using oversampling and noise shaping which can substantially reduce the timing clock rate, alternative systems that combine multilevel and PWM are also attractive. For example, if we assume a 7-bit multilevel DAC (128 levels), then for 20-bit resolution, a PWM edge timing clock rate of 44.1 kHz $* 2^{(20-7)} = 361$ MHz appears feasible.

Two desirable characteristics of an audio DAC are accurate conversion of low-level signals and a monotonic coding characteristic. It is here that the PWM technique offers significant advantage as a progressive change in pulse width is synonymous with monotonic coding of small signals. Of course, once a multilevel format is employed, there is an additional requirement to match

the levels to a high degree of accuracy. However, provided the number of levels is not too great, this can be achieved by a parallel array of identical cells that can be progressively switched in and out of the circuit as required. It is proposed that 7-bit (128-level) multilevel coding can be handled by a parallel architecture and that economic automatic calibration is feasible within a VLSI system.

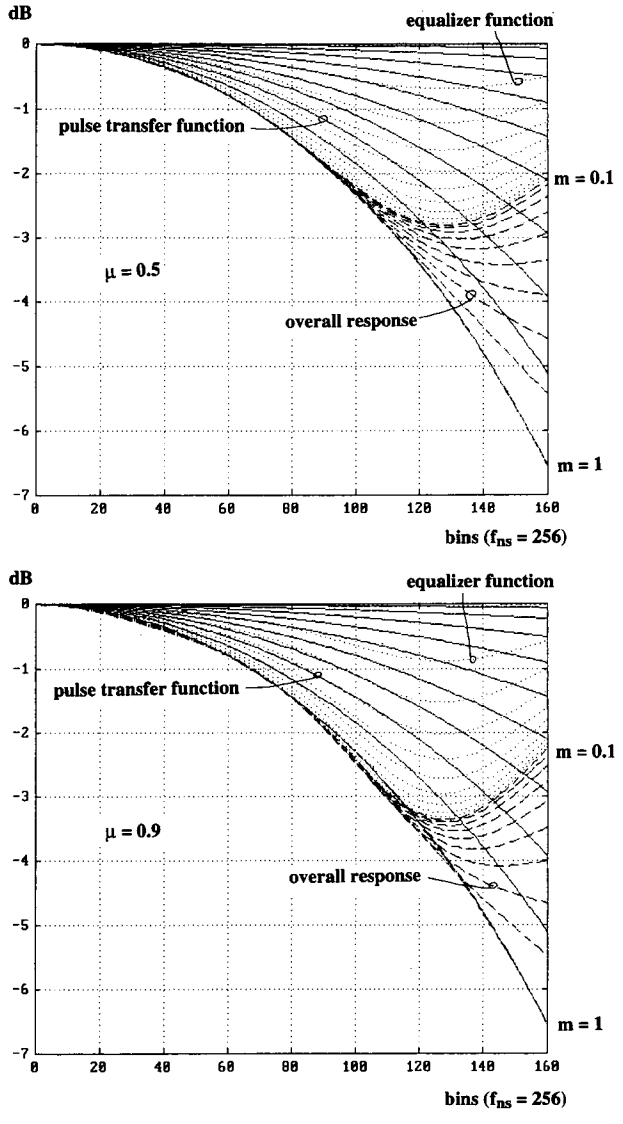


Fig. 4. Approximate equalization characteristics for $\mu = 0.5$ and $\mu = 0.9$.

Before considering PWM interpolation with linearization, a possible FDAC architecture is outlined based on a current-steering topology. The system is shown in Fig. 8 and uses emitter-coupled logic (ECL) bistables to drive differential current switches. The use of ECL allows fast pulse transitions, which are necessary to give precise edge definition that is compatible with low-jitter clock sources. Also, it is imperative that the current sources be matched to a defined reference current, although this matching must also include the effect of current loss, for example, in the bases of the switching transistors. A possible calibration procedure nominates one current source as reference and then compares individually its value against the remaining sources. This can be done by alternate switching and adjusting the current level until no perceived ac component appears on the output. The technique theoretically has a high accuracy because of the following attributes:

1) Observing the error as an ac signal allows a high-gain error amplifier without dc drift problems.

2) Switching rates at the DAC output are kept low, reducing effects of slew distortion and jitter.

3) Fine adjustment current values for each cell can be stored in a local memory and DAC.

4) Current steering is fast and with appropriate capacitive elements on the output, slew-induced distortion is eliminated as there are no rapid dv/dt effects or feedback amplifiers being momentarily operated near open loop [8].

Once the multilevel DAC current sources are calibrated, PWM can be used to interpolate between output levels with guaranteed monotonicity and potentially high

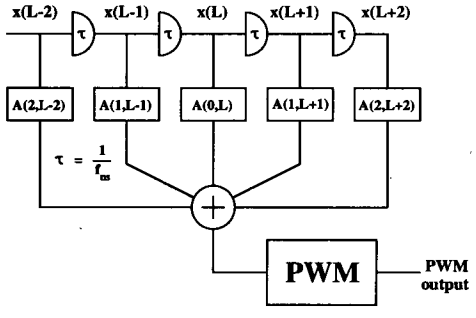


Fig. 5. Five-coefficient FIR filter.

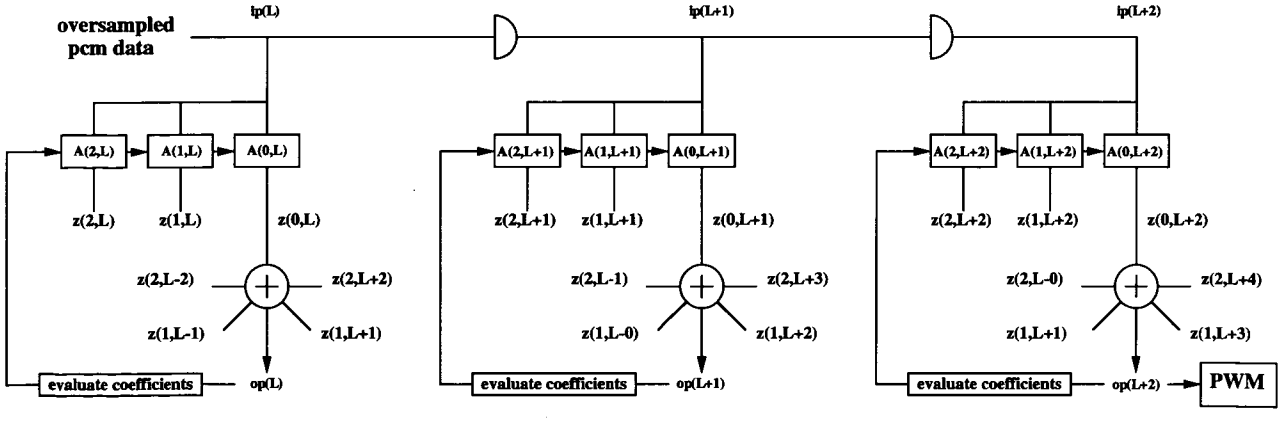


Fig. 6. Interconnected coefficient sets where each set is calculated from knowledge of surrounding samples.

linearity, where the proposed form of PWM interpolation can take a variety of formats. However, to reduce switching rates and to have a monotonic change in pulse area with progressive changes in input data, the two formats illustrated in Fig. 9 have been selected and designated types 1 and 2.

Type 1, as shown in Fig. 9(a), offers reduced switching transitions but has a step change in pulse width at each multilevel boundary, which may add broad-band distortion within the correction process. However, type 2 has symmetry about even quanta and maintains similar pulse widths either side of a multilevel boundary. These differences can be seen by examining the progressive pulse structures in Fig. 9(a) and (b).

In type 1, for an increasing signal level the pulse width always grows from the center switching between two adjacent output DAC levels. When the pulse width equals the sample period, a new pulse emerges at the center switching up to the next quantization level. However, in type 2 the pulse alternates between outward growth from the center and inward growth from the pulse edges. The symmetrical pulse structure guarantees zero dynamic phase distortion, and the PWM format yields a smooth transition between adjacent DAC levels with no repeated codes. In the DAC interpolation process only one cell is switched at a time, and, between adjacent

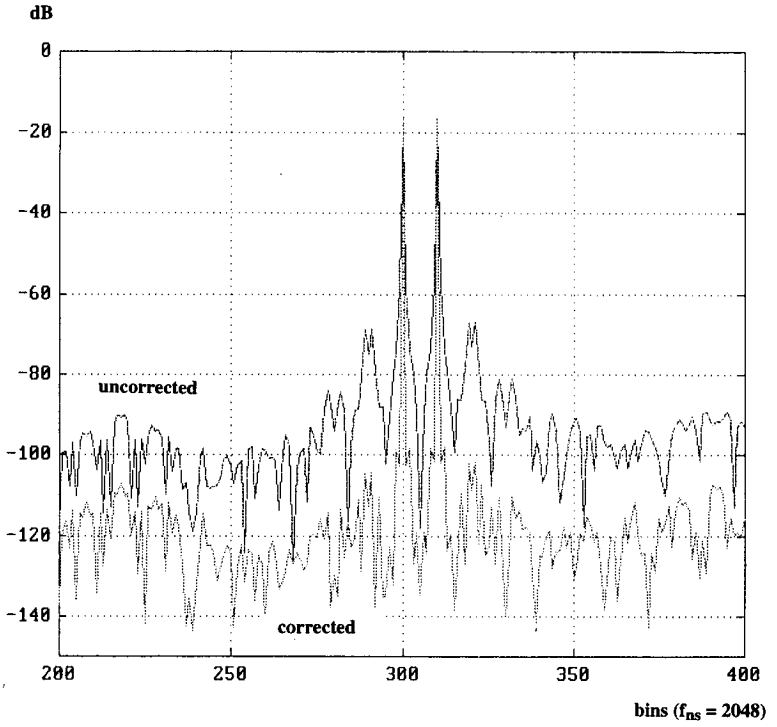


Fig. 7. Output spectra for binary PWM for $\mu = 0.5$ and NC = 6.

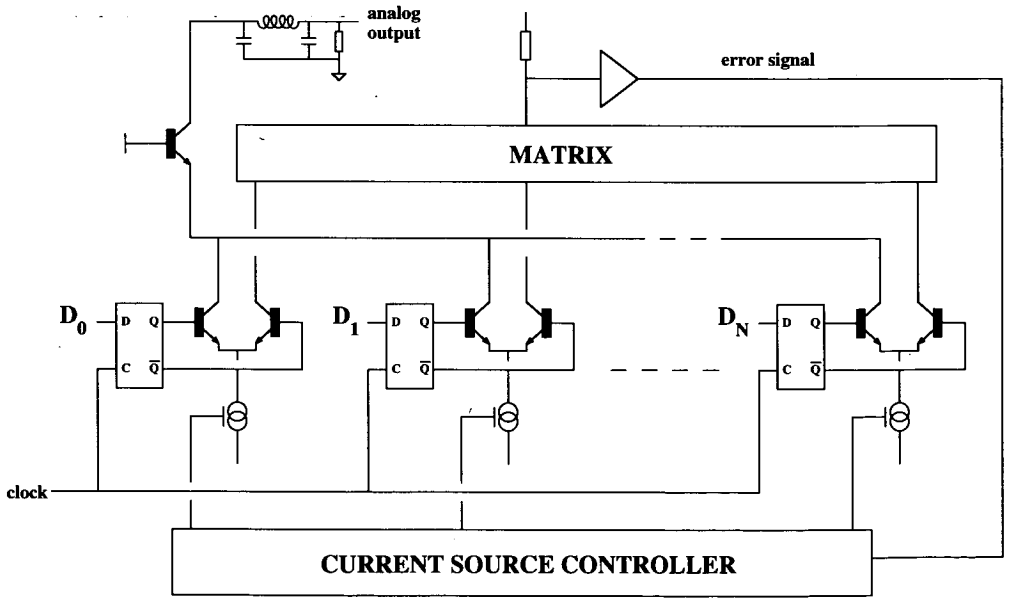


Fig. 8. Basic FDAC system topology.

samples, current-switching cells become active in a controlled and progressive sequence.

For the type 1 sample format, consider a sample $x(L)$ of amplitude $\{N + m\}$, where N is an integer and $0 < m < 1$, that is,

$$x(L) = N + m . \tag{14}$$

N corresponds to the Nth quantization level of the FDAC, which has a width $1/f_{ns}$, whereas m determines the intermediate sample width m/f_{ns} (see example pulse in Fig. 9(a)) of PWM interpolation between levels N and $(N + 1)$. The dc normalized transfer function $F_{np}(f)$ of this composite pulse is

$$F_{np}(f) = \left[\frac{\sin(m\pi f/f_{ns})}{\pi f/f_{ns}} + N\frac{\sin(\pi f/f_{ns})}{\pi f/f_{ns}}\right]\frac{1}{N + m} .$$

If the normalized target transfer function is $\sin(\pi f/f_{ns})/(\pi f/f_{ns})$, corresponding to $m_t = 1$, then the equalizer transfer function $E_m(m, f)$ follows as

$$E_m(m,f) = \frac{1}{F_{np}(f)}\frac{\sin(\pi f/f_{ns})}{\pi f/f_{ns}}$$

whereby

$$E_m(m,f) = \frac{N + m}{N + \sin(m\pi f/f_{ns}/\sin(\pi f/f_{ns})} . \tag{15}$$

For $N = 0$ this reduces to

$$E_m(f, m)\Big|_{N=0} = \frac{m\sin(\pi f/f_{ns})}{\sin(m\pi f/f_{ns})} \tag{16}$$

which is equivalent to Eq. (2) for $m_t = 1$.

The reason for selecting $m_t = 1$ is that for $N >> 1$, $E_m(m, f) \approx 1$, which reduces the equalizer coefficient values, hence dependence on correction, and improves

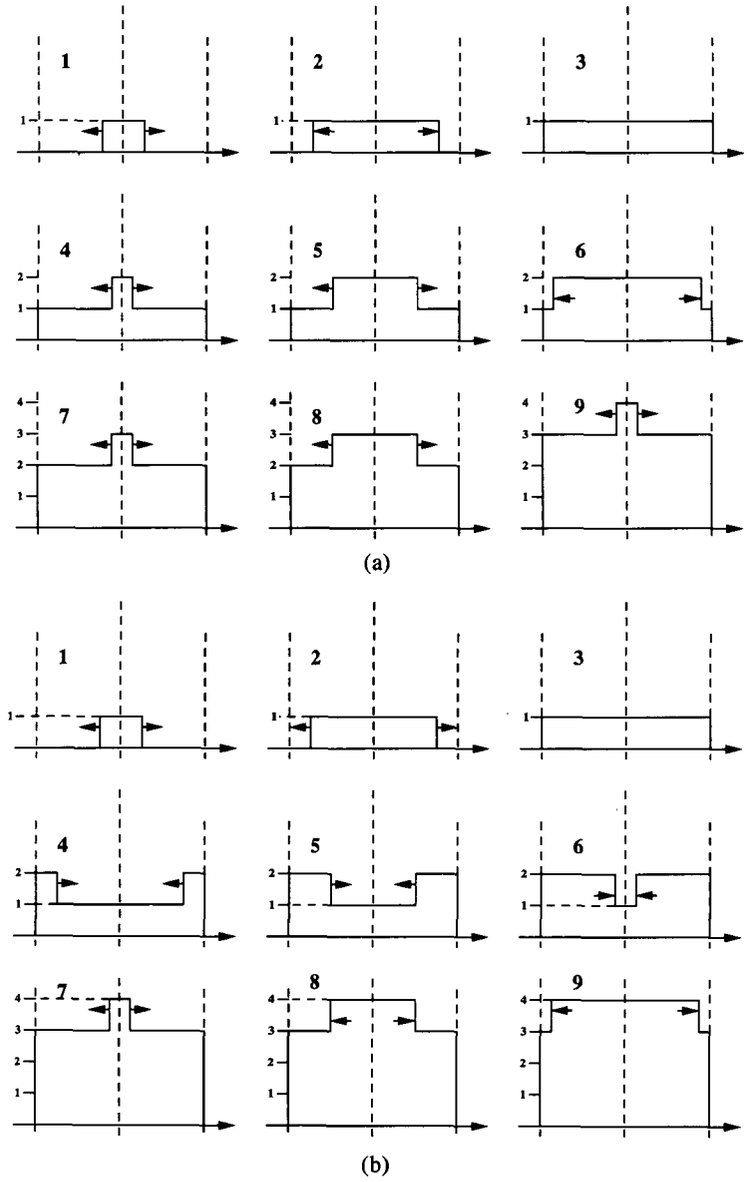


Fig. 9. Simultaneous pulse-amplitude and pulse-width modulation. (a) Type 1. (b) Type 2.

the accuracy of the matching functions.

If a type 2 pulse format is used, the process is similar, except that N in Eqs. (14) and (15) is rounded to even integer values and m takes bipolar values, where $-1 < m < 1$.

Using Eq. (15), the optimum equalization characteristics are plotted in Fig. 10 as functions of m and f for both type 1 and 2 pulse formats, where type 1 reveals more ripples at the integer boundaries.

Because of the PWM format, a dynamic correction filter is required to operate in a mode similar to that of two-level PWM. However, observation of Fig. 10 shows that the form of correction differs in detail such that as the pulse amplitude is increased, there is a corresponding dilution of spectral modulation. In fact spectral modulation is greatest for the lowest levels, where fortunately absolute distortion levels are lowest. The equalization function $E_m(m, f)$ for the multilevel code is the reciprocal of the transfer function of the composite pulse multipled by the target transfer function corresponding to $m_t = 1$.

To determine coefficients for the FIR filter, a procedure similar to that described in Section 1.1 can be followed by introducing a guard band of f_g Hz and forming a function $A_c(m, f)$ which takes a scaled and reflected form of the function in the acceptance band. At the transition band edge f_0 let $E_c(m, f) = A_c$,

$$A_c = \left(1 + \frac{m}{N}\right) \frac{\pi(0.5 - f_g/f_{ns})}{\sin[\pi(0.5 - f_g/f_{ns})]} \left\{1 + \frac{1}{N} \frac{\sin[m\pi(0.5 - f_g/f_{ns})]}{\sin[\pi(0.5 - f_g/f_{ns})]}\right\}^{-1}. \quad (17)$$

The function $A_c(m, f)$ in the guard band is then given by

$$A_c(m, f) = \frac{1}{k}\left\{A_c(1 + k) - \frac{\pi k(0.5 - f/f_{ns})}{\sin[\pi k(0.5 - f/f_{ns})]} \frac{1 + \frac{m}{N}}{1 + \frac{1}{N} \frac{\sin[m\pi k(0.5 - f/f_{ns})]}{\sin[\pi k(0.5 - f/f_{ns})]}}\right\} \quad (18)$$

where the scaling factor k again follows from Eq. (6). Using Fourier analysis, a set of coefficients can then be determined directly. Alternatively, the technique described in Section 1.2 can be used, where the equalization bandwidth is set by selecting the factor μ.

As $x(L)$ increases above unity, the variations in the transfer function become progressively reduced, which relaxes the iterative evaluations of the coefficients. Also, the coefficients can be estimated using interpolation, which simplifies overall computation. In Fig. 11 a coefficient map for NC = 6 is presented for signal levels corresponding to a range of $m = 0$ to ~7, where the general trends can be observed. Further simplification also results as positive and negative signals of equal magnitude can be assigned the same coefficient values.

To demonstrate the performance of the multilevel DAC, results are plotted in Figs. 12 to 15 for type 1 and type 2 pulse formats, with $\mu = 0.5$ and 0.9, respectively. For these examples, the input excitation is defined by

$$x(L) = 50[\sin(k_0 f_{ns} t) + \sin(k_1 f_{ns} t) + \sin(k_2 f_{ns} t)] + \text{rnd} * 2^{-17}. \quad (19)$$

Again, $k_0 = 2\pi/\text{NS}$, $k_1 = 300k_0$, $k_2 = 310k_0$, and rnd has a uniform probability density function from -0.5 to 0.5 and $t = L/f_{ns}$ for $1 \leq L \leq \text{NS}$ where Ns = 2048. Data relating to each computation are presented in Table 1.

The results show that, because of the multilevel structure of the signal, once the higher quantization levels are excited, the distortions become noiselike rather than discrete frequencies. The error-correction process reduces this distortion significantly and also shows that there is little advantage gained in using more than NC = 6.

4 CONCLUSION

This paper has considered the application of linearization to the PWM process approached from the spectral domain, where improved linearization of uniformly sampled PWM was achieved. The work extends earlier results by achieving better spectral matching, which allowed a reduction in the system sampling rate close to the Nyquist frequency.

Although the technique is directly applicable to two-level PWM, a multilevel converter was also explored as a means of enhancing linearity and also for reducing the internal clock rates necessary for timing the PWM edges.

A feature of the converter was the monotonicity of pulse areas as a function of the input sample level together with dynamic linearization to compensate for spectral domain modulation with the signal level. It was demonstrated that low-level signals show less distortion in the corrected FDAC together with a general reduction in PWM-related distortion at higher levels.

An advantage of combining a calibrated FDAC with PWM interpolation is to reduce switching rates at the digital–analog gateway. Fast switching of pulses using current steering without associated high dv/dt is an important attribute in minimizing distortion. Also low logic propagation times and stable clock sources lower jitter-related errors.

In comparison with bit-stream technology, the number of signal transitions at the output of the DAC is reduced greatly, which offers the potential of lower jitter sensitivity. It is also suggested that the problems of RF circuit

design and layout, commonly associated with interconnecting bit-stream integrated circuits, are reduced, even though high clock rates remain necessary for timing the width of the PWM pulses. The technique overcomes the need to use excessive oversampling ratios and orders of noise shaping to achieve an acceptable clock rate. By their nature such processes imply a greater sensitivity to jitter and slew rate artifacts as the number of significant signal transitions is considerably higher.

This paper has presented an alternative DAC for high-performance digital audio. The technique offers a fast and readily calibrated architecture, where known nonlinearities can be reduced to low levels. It does not require excessive oversampling ratios or the use of noise shaping and therefore moves against the trend of many bit-stream DACs. However, for power DAC applications [5], the use of binary PWM remains attractive where alternative error-correction strategies have been reported [6], [7].

5 REFERENCES

[1] P. H. Mellor, S. P. Leigh, and B. M. G. Cheetham, "Improved Sampling Process for a Digital Pulse Width Modulated Class D Power Amplifier," presented at the IEE Colloquium on Digital Audio Signal Processing, London, UK (1991 May).

[2] P. Craven, "Toward the 24-bit DAC: Novel Noise-Shaping Topologies Incorporating Correction for the Nonlinearity in a PWM Output Stage," *J. Audio Eng. Soc.*, vol. 41, pp. 291–313 (1993 May).

[3] M. O. J. Hawksford, "Dynamic Model-Based Linearization of Quantized Pulse-Width Modulation for

Table 1. Computational information for type 1 and 2 multilevel PWM spectra.

Fig. 12: Type 1 pulse format μ = 0.5, NS = 2048, NC = 2, 4, 6, and 8 Frequency range 1 to 1000 bins, amplitude range −150 dB to 40 dB
Fig. 13: Type 1 pulse format μ = 0.9, NS = 2048, NC = 2, 4, 6, and 8 Frequency range 1 to 1000 bins, amplitude range −150 dB to 40 dB
Fig. 14: Type 2 pulse format μ = 0.5, NS = 2048, NC = 2, 4, 6, and 8 Frequency range 1 to 1000 bins, amplitude range −150 dB to 40 dB
Fig. 15: Type 2 pulse format μ = 0.9, NS = 2048, NC = 2, 4, 6, and 8 Frequency range 1 to 1000 bins, amplitude range −150 dB to 40 dB

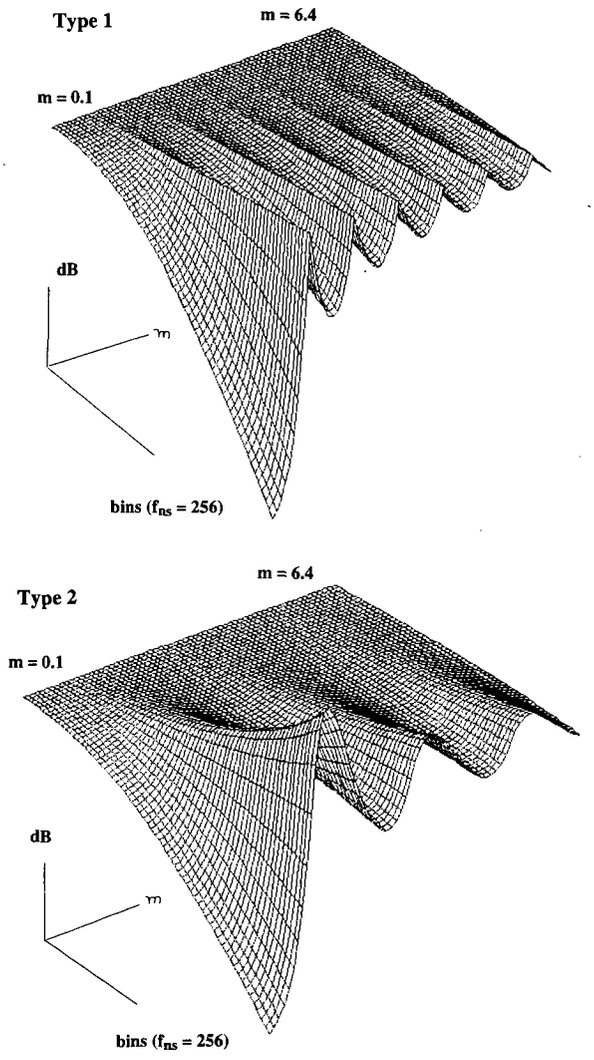


Fig. 10. Optimum equalization $E_m(m, f)$ as a function of amplitude and frequency for type 1 and 2 pulse formats.

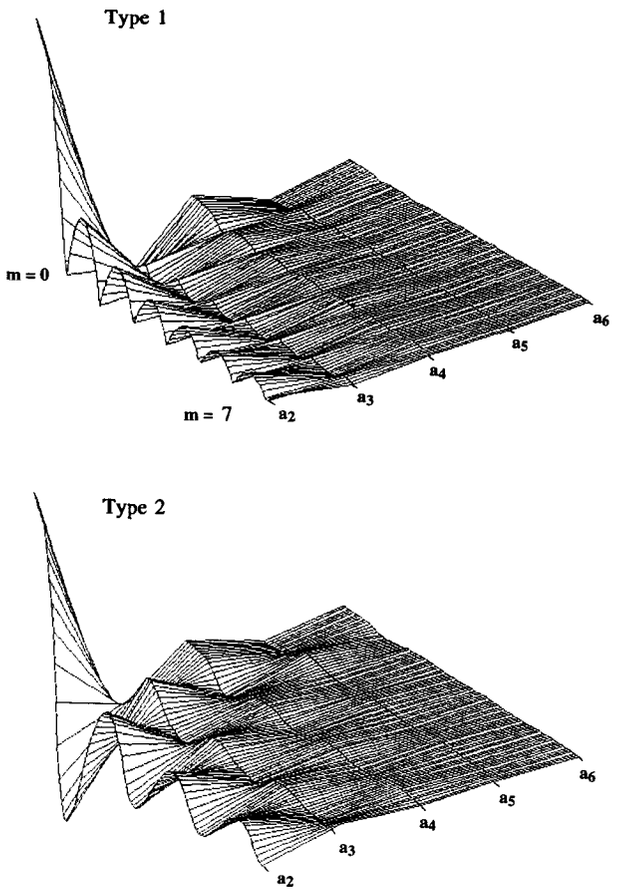


Fig. 11. Coefficient maps for NC = 6 (a_1 not shown) and m = 0 to 7.

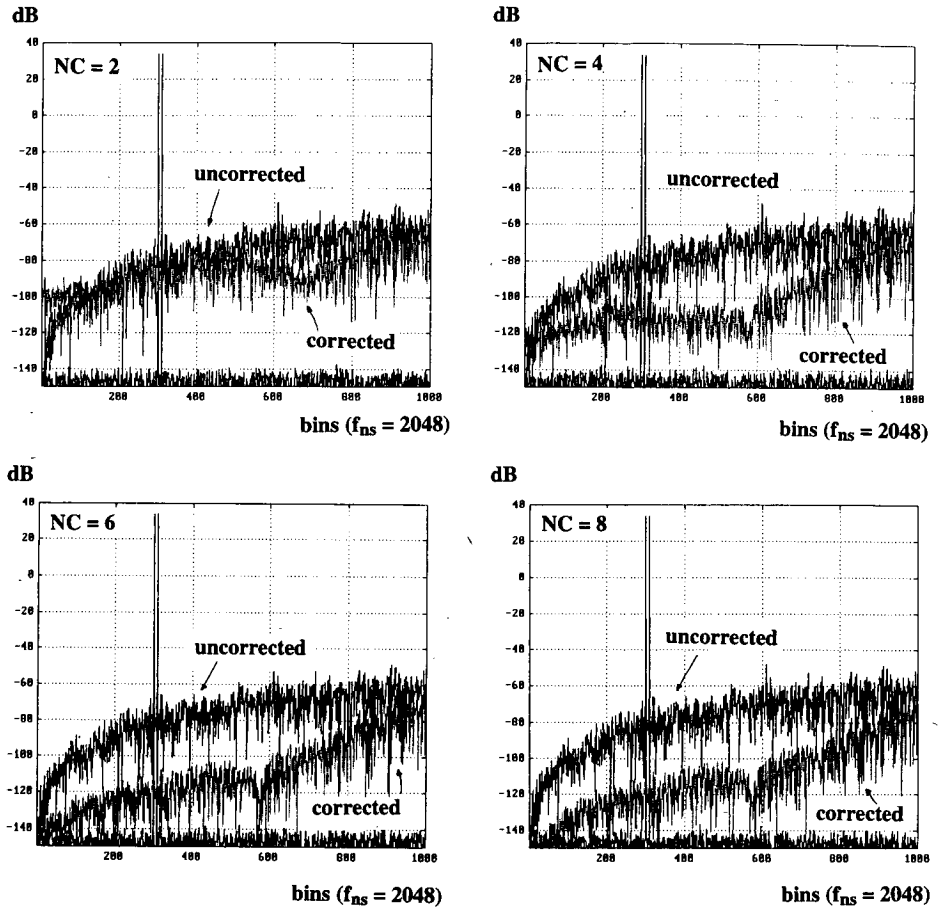


Fig. 12. Output spectra of multilevel DAC. Type 1, μ = 0.5.

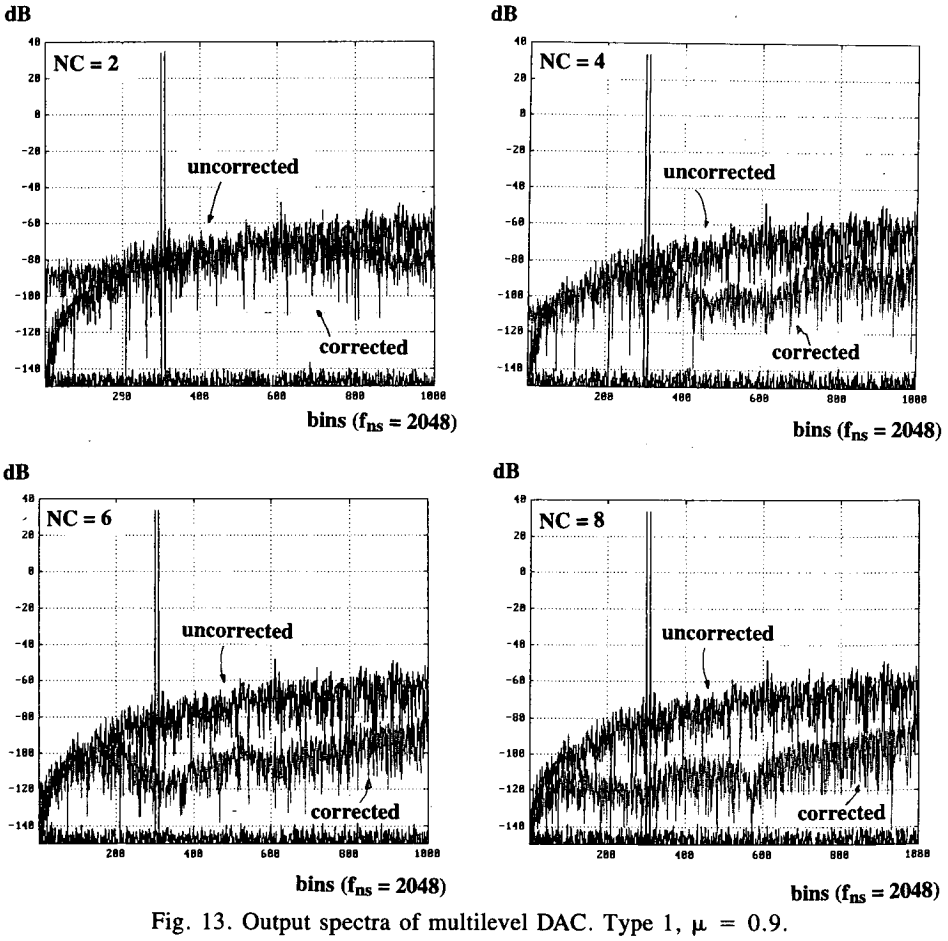


Fig. 13. Output spectra of multilevel DAC. Type 1, μ = 0.9.

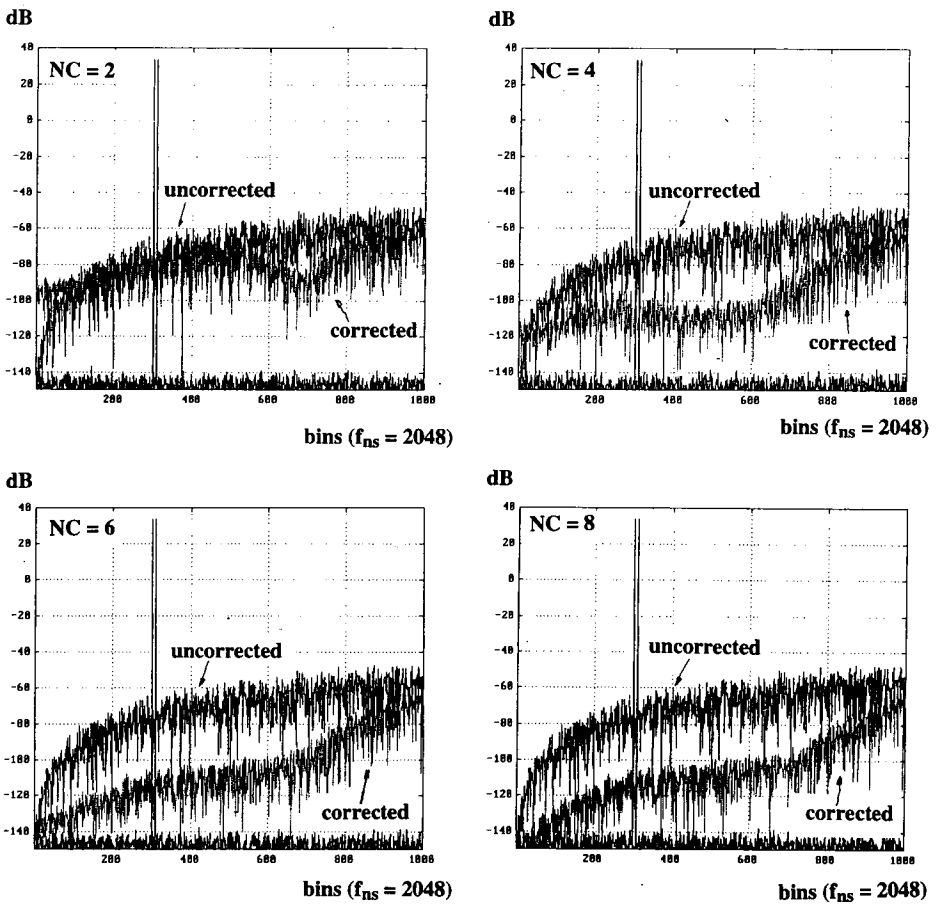


Fig. 14. Output spectra of multilevel DAC. Type 2, μ = 0.5.

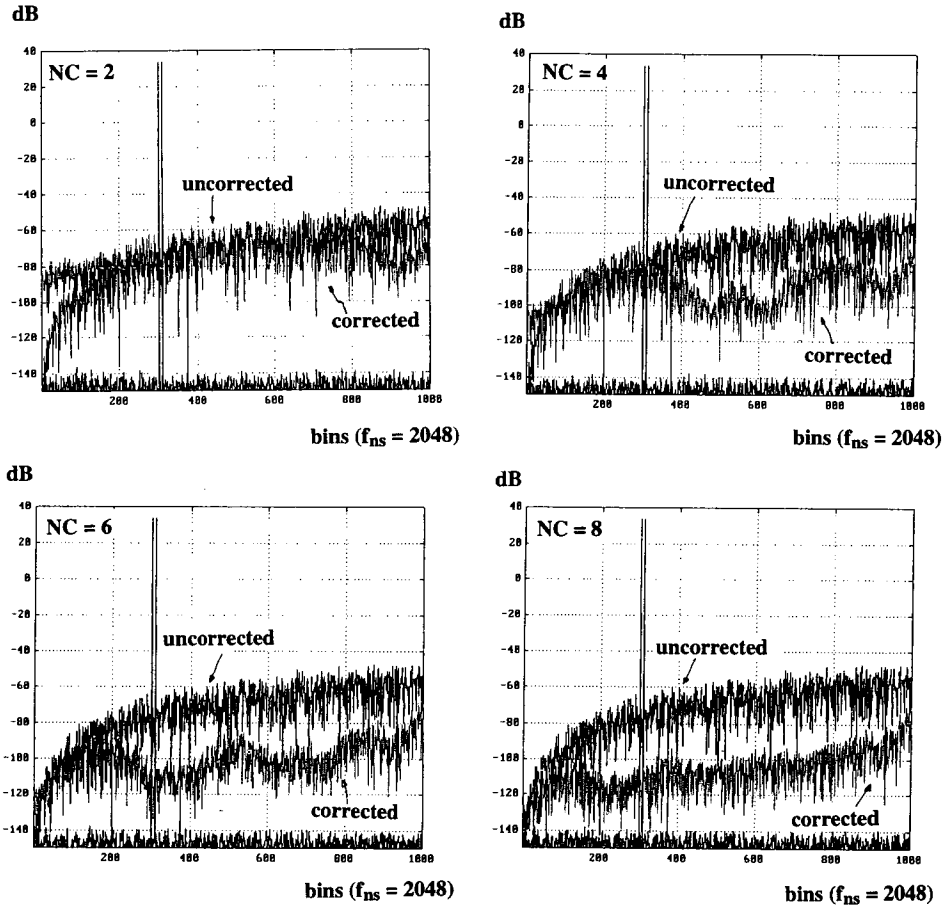


Fig. 15. Output spectra of multilevel DAC. Type 2, μ = 0.9.

Applications in Digital-to-Analog Conversion and Digital Power Amplifier Systems," *J. Audio Eng. Soc.*, vol. 40, pp. 235–252 (1992 Apr.).

[4] M. O. J. Hawksford, "Multi-Level to 1 bit Transformations for Applications in Digital-to-Analogue Converters Using Oversampling and Noise Shaping," *Proc. Inst. Acoust.*, vol. 10, pp. 129–143 (1988 Nov.).

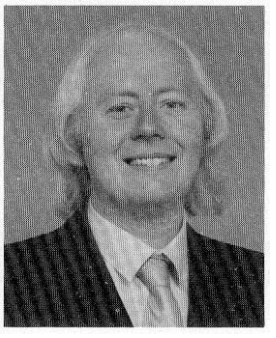
[5] J. M. Goldberg and M. B. Sandler, "Noise Shaping and Pulse-Width Modulation for an All-Digital Audio Power Amplifier," *J. Audio Eng. Soc.*, vol. 39, pp. 449–460 (1991 June).

[6] S. Logan, "Linearization of Class D Output Stages for High Performance Audio Power Amplifiers," M.Sc. dissertation, Essex University, Colchester, UK (1993 Autumn).

[7] M. O. J. Hawksford and S. Logan, "Linearization of Class D Output Stages for High-Performance Audio Power Amplifiers," presented at the 2nd International IEE Conference on Analogue-to-Digital and Digital-to-Analogue Conversion, Cambridge University, Cambridge, UK (1994 July).

[8] M. O. J. Hawksford, "Digital-to-Analog Converter with Low Intersample Transition Distortion and Low Sensitivity to Sample Jitter and Transresistance Amplifier Slew Rate," *J. Audio Eng. Soc.*, vol. 42, pp. 901–917 (1994 Nov.).

THE AUTHOR



Malcolm Omar Hawksford is director of the Centre for Audio Research and Engineering, a professor in the Department of Electronic Systems Engineering at the University of Essex, and Scheme Director for the M.Sc. in Audio Systems Engineering, where his interests encompass audio engineering, electronic circuit design, and signal processing. Professor Hawksford studied electrical engineering at the University of Aston in Birmingham where he gained a First Class Honours BSc and Ph.D. The Ph.D. program, sponsored by the BBC via a Research Scholarship, investigated the application of delta modulation for color television and the development of a time-compression/time-multiplex system for combining luminance and chrominance signals. While at Essex University, he has undertaken research principally in the fields of analog amplifiers, digital signal processing, and loudspeaker systems. Since 1982 research into digital crossover networks and equalization for loudspeakers has been pursued, which has culminated in an advanced digital and active loudspeaker system being designed within the university. Research topics have also encompassed oversampling and noise shaping techniques applied to analog-to-digital and digital-to-analog conversion and the linearization of PWM encoders.

Professor Hawksford has published in the *Journal of the Audio Engineering Society* on topics that include error correction in amplifiers, oversampling techniques, and MLS techniques. His supplementary activities include writing for *Hi-Fi News and Record Review* and designing high-end analog and digital audio equipment. He is a chartered engineer and is a fellow of the AES, the Institution of Electrical Engineers, and the Institute of Acoustics. He is a member of the technical committee of Acoustic Renaissance for Audio (ARA), a group currently promoting a system for storing multichannel, high-definition audio signals on high-capacity DVD optical disks. He is also technical adviser to *HFN and Record Review* and a technical consultant to LFD Audio, UK.

An Oversampled Digital PWM Linearization Technique for Digital-to-Analog Conversion

Jin-Whi Jung and Malcolm J. Hawksford

***Abstract*—An algorithmic-based linearization process for uniformly sampled digital pulsewidth modulation (PWM) is described. It is shown that linearization of the intrinsic distortion resulting in uniformly sampled PWM can be achieved by using a fractional delay digital filter embedded within a noise shaping re-quantizer. A technique termed direct PWM mapping is proposed as a pre-compensation filter scheme for applications in high-resolution digital-to-analog conversion.**

***Index Terms*—Digital-to-analog (D/A) conversion, pulsewidth modulation (PWM), sigma–delta modualtion (SDM).**

I. Introduction

TIME-DISCRETE pulsewidth modulation (PWM) has a natural synergy with digital circuitry used within both digital systems and more general very large-scale integrated (VLSI) devices. Since MASH [6], the amalgamation of low-bit uniformly sampled PWM and noise shaping re-quantization has invited considerable research, not least because it relaxes the problem of quantizer saturation encountered in two-level quantization sigma–delta modulatoion (SDM). However, the intrinsic nonlinearity of uniformly sampled PWM is a well-known problem that can produce significant levels of in-band distortion whereas naturally sampled PWM, as commonly used in analog PWM systems, is linear within the baseband frequency range. Consequently, uniformly sampled PWM is subject to spectral distortion implying the filtered output signal is partially corrupted, see Craven [3], Rowe [11], and Leigh [12] for detailed discussions.

Several research papers that address the subject of linearization in PWM have been reported. Mellor *et al.*, [5] and Leigh [12] introduced a method using time-domain interpolation, while Goldberg [2] and Goldberg and Sandler [4] presented a more refined interpolation technique to approximate uniformly sampled PWM to natural sampling. Hawksford [1] developed a novel uniformly sampled PWM distortion compensation technique for uniformly sampled PWM that was later adapted to embrace multilevel, multiwidth PWM [10]. Craven [3] then proposed a similar compensation technique but where error correction was implemented using negative feedback incorporated within a noise shaping loop.

Manuscript received January 4, 2002; revised March 28, 2004. This paper was recommended by Associate Editor U.-K. Moon.

J.-W. Jung is with the Department of Electronic and Electric Engineering, University College London, London WC1E 6BT, U.K. (e-mail: j.jung@ucl.ac.uk).

M. J. Hawksford is with Department of Electronic Systems Engineering, University of Essex, Colchester CO4 3SQ, U.K. (e-mail: mjh@essex.ac.uk).

Digital Object Identifier 10.1109/TCSI.2004.834487

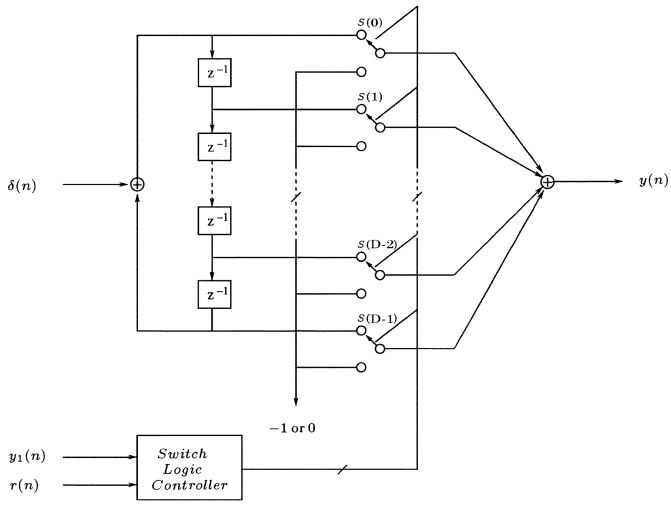


Fig. 1. Implementation of digital PWM in which 2-level quantizers s_i are connected within circular formation and gated by the switch logic controller according to α, where D is the PWM resolution bits, $\delta(n)$ is the unit impulse, $r(n)$ is the PWM reference signal, $y_1(n)$ is the output of the noise shaping re-quantizer, and $y(n)$ PWM output.

This paper presents a fractional sample interpolation scheme to reduce the intrinsic distortion resulting from the uniformly sampled PWM. The technique designated direct PWM mapping (DPM), exploits a Farrow structure fractional delay finite-impulse response (FIR) digital filter employing third-order Lagrange interpolation. The principal feature of the DPM scheme is that the normalized fractional sample grid of the Farrow structure directly generates the digital PWM time base defining the PWM output-wave transitions. It is shown here that a 3-tap Farrow structure FIR filter is sufficient for linearization and as such contributes to the simplicity of the overall PWM system.

II. Digital PWM and Intrinsic Distortion

Fig. 1 shows the proposed digital PWM where the modulating signal α is derived from a logic combination of both the quantized output signal $y_1(n)$ and the PWM reference signal $r(n)$. Let α be scaled to an integer value i so that $i \in (0, D-1)$ where $i \in \mathbb{Z}$ and D is derived from the pulse-repetition period $T_D = DT_s$, i.e., integer multiples D of the sampling period $T_s = 1/f_s$. The signal α addresses a switch state control circuit where s_1 and s_0 denote the on-state and off-state switches, respectively. The PWM time-domain sequence $h(\alpha, n)$ is therefore given as

$$h(\alpha,n) = h(\alpha,n-D) + s_1(0)\cdot\delta(n) + s_1(1)\cdot\delta(n-1) + \cdots + s_1(\alpha-1)\cdot\delta(n-\alpha+1)$$

$$+ s_1(\alpha)\cdot\delta(n-\alpha) + s_0(\alpha+1)\cdot\delta(n-\alpha-1) + \cdots + s_0(D-1)\cdot\delta(n-D+1). \quad (1)$$

Alternatively, in z domain, $h(\alpha, n)$ is expressed as

$$H(\alpha,z) = \frac{1}{1-z^{-D}}\cdot\left(\sum_{i=0}^{\alpha} s_1(i)\cdot z^{-i} + \sum_{i=\alpha+1}^{D-1} s_0(i)\cdot z^{-i}\right). \quad (2)$$

The last term of (2) is required for zero padding in order to represent the logic state 0 within a pulse-repetition period for unipolar leading-edge PWM; also, if s_0 is set to -1 it represents negative magnitude padding in bipolar leading-edge PWM.

Expressions (1) and (2) provide an intuitive understanding of digital PWM.

1) The sampled data z^{-i} is gated by corresponding 2-level quantizers and connected within a circular formation.
2) While the sampled data $z^{-\alpha}$ relates to pulse-position modulation (PPM) where the summation of impulses from 0 to α constitutes a rectangular pulse having pulsewidth α.

In this manner, the denominator $R(z) = 1/(1-z^{-D})$ determines the PWM pulse repetition with D roots.

For the purpose of illustration, if the allocation of α and D are chosen arbitrarily, noting that all the logic states of the switches $s_1(i)$ are asserted 1, then the periodic impulse response sequence will appear as

$$h = \begin{bmatrix} s_1(0) & s_1(1) & \cdots & s_1(\alpha) & 0 & \cdots & 0 \\ s_1(0) & s_1(1) & s_1(2) & \cdots & s_1(\alpha) & \cdots & 0 \\ s_1(0) & \cdots & s_1(\alpha) & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ s_1(0) & s_1(\alpha) & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 & \cdots & 0 \\ 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}. \quad (3)$$

Note also that the transfer function of the digital PWM can be expanded in an infinite geometric series

$$H(\alpha,z) = (1+z^{-1}+\cdots+z^{-\alpha}+0+\cdots+0)\cdot(1+z^{-D}+z^{-2D}+\cdots). \quad (4)$$

The pulse-repetition period D has a base of $1-z^{-D}$, hence there exist D frequencies within the Nyquist interval with D roots of unity. In this notation each PWM bit corresponds to the delays at each sampling instant. Defining $\omega_i = (2\pi f_i/f_s) = (2\pi i/D)$, the roots of $H(\alpha,z)$ are obtained by solving $z^D = 1$ for D solutions, that is D roots of unity in the digital PWM system. The zeros of the denominator $R(z)$ that constitute dips in spectral response form D roots of unity on the unit circle, where if D corresponds to the PWM bit resolution, then, generally, $D = 2^M$ where M is even integer.

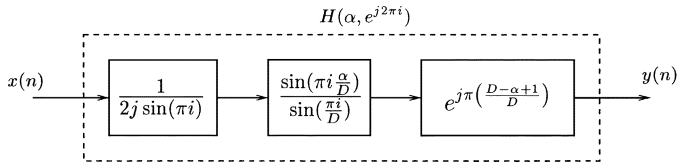


Fig. 2. Uncompensated PWM system represented as a nonlinear magnitude response and a variational delay response, derived from (5).

Hence, from (2), the transfer function $H(\alpha, \exp[j2\pi i])$ may be derived as

$$H(\alpha,e^{j2\pi i}) = \frac{1}{1-e^{-j2\pi i}}\cdot\frac{1-e^{-j2\pi i\frac{\alpha}{D}}}{1-e^{-j\frac{2\pi i}{D}}} = \frac{1}{2j\sin(\pi i)}\cdot\frac{\sin\left(\pi i\frac{\alpha}{D}\right)}{\sin\left(\frac{\pi i}{D}\right)}\cdot e^{j\pi i(\frac{D-\alpha+1}{D})}. \quad (5)$$

The transfer function (5) represents conventional uncompensated digital PWM that is characterized in the frequency domain by a nonlinear magnitude response, see Fig. 2. As such, this process can be compared to a similar result given previously by Hawksford [1], where the linear magnitude component of a PWM pulse is multiplied by a nonlinear transfer function, i.e., $H(\alpha, \exp[j2\pi i]) = \alpha\cdot H(\exp[j2\pi i])$ that results from modulating the pulsewidth, also [10] extended this discussion to include multilevel digital PWM.

For large values of D, the transfer function of $H(\alpha, \exp[j2\pi i])$ in (5) becomes

$$H(\alpha,e^{j2\pi i}) = \frac{1}{2j\sin(\pi i)}\cdot\frac{\sin\left(\pi i\frac{\alpha}{D}\right)}{\frac{\pi i}{D}}\cdot e^{j\pi i(\frac{D-\alpha+1}{D})} \quad (6)$$

where (6) shows that transfer function $H(\alpha, \exp[j2\pi i])$ has an α-dependent nonlinearity of the form $\sin(\alpha x)/x$ that describes deterministically the nonlinear modulation process inherent in uniformly sampled digital PWM. From (6), it follows that for sufficiently small α, the transfer function $H(\alpha, \exp[j2\pi i])$ approximates to

$$H(\alpha,e^{j2\pi i}) = \frac{\alpha}{2j\sin(\pi i)}\cdot e^{j\pi i(\frac{D-\alpha+1}{D})}. \quad (7)$$

The spectral distortion described by (5) can be calculated directly using computer simulation, where over a bandwidth of 192 kHz, Fig. 3 reveals a typical spectrum for uniformly sampled PWM. The intermodulation components that occur at the multiples of sampling frequency of 48 kHz can also be calculated using the mathematical results in [11], they are observed for each harmonic of the pulse-repetition frequency, where their magnitudes become lower and spectra broaden as frequency increases. Also, reflected distortion components can be seen within the baseband close to the fundamental frequency of 750 Hz. Both these classes of distortion are exploited in this study to assess the performance of PWM linearization and are estimated by simulation.

1) *Harmonic distortion:* With an input signal of frequency $f = 6$ kHz sampled initially at 48 kHz and using four times upsampling, PWM for both 6-bit leading-edge sampling Fig. 4(a) and 7-bit double edge sampling Fig. 4(b) are simulated to observe the reflected spectral distortion appearing within baseband. Harmonics $2f$ and $3f$ are the

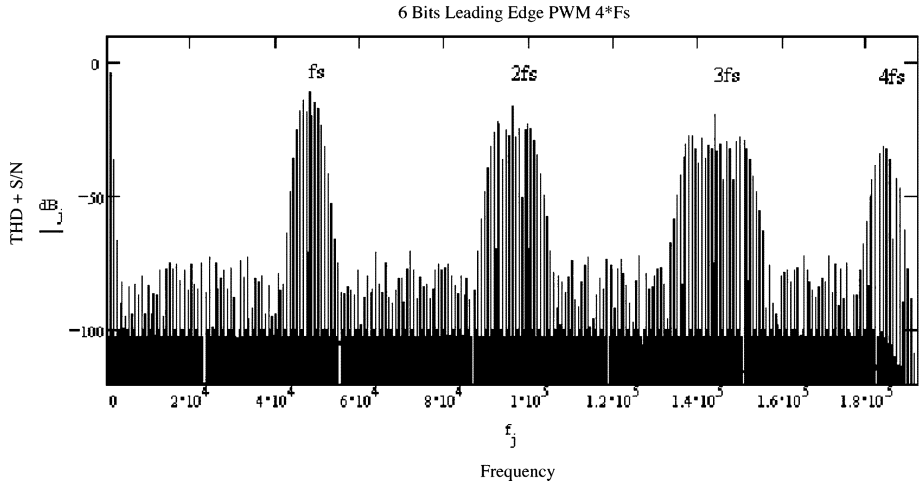


Fig. 3. Typical spectra of uniformly sampled PWM; intermodulation harmonic distortions at each multiples of PWM pulse-repetition frequency 48 kHz and reflected harmonic distortions in baseband, where the input frequency is 750-Hz single tone.

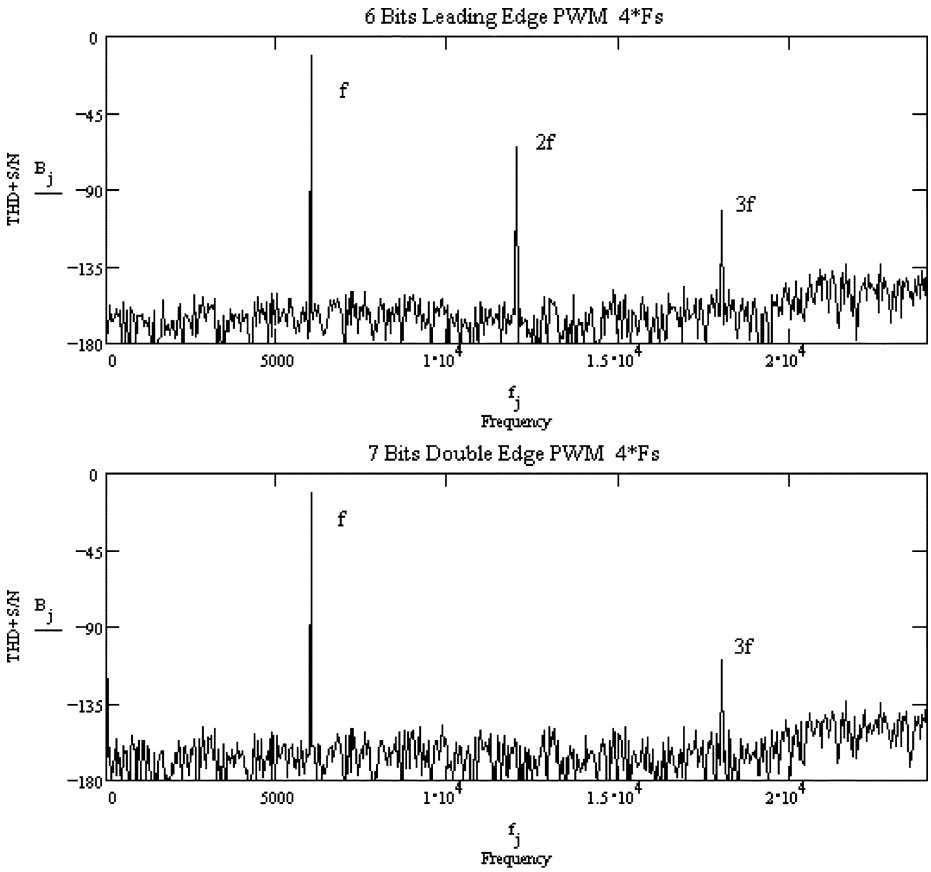


Fig. 4. PWM distortion in audio frequency band. (a) A 6-bit leading-edge sampling. (b) A 7-bit double edge sampling PWM systems at 4 times $f_s = 48$ kHz, input signal is 6 kHz.

reflected distortion components with respect to the 6-kHz input signal. The results confirm that double edge sampling achieves even order harmonic cancellation where for example the component $2f = 12$ kHz is absent, while it remains for leading-edge PWM.

2) *Intermodulation distortion:* The intermodulation results are shown in Fig. 5 where three superimposed input frequencies $f_1 = 0.75$ kHz, $f_2 = 6$ kHz, and $f_3 = 9$ kHz drive a 6-bit $4f_s$ PWM and hence the sampling rate is 192 kHz. The observed intermodulation distortion components are located both sides of the fundamental frequencies f_1, f_2, and f_3 and their multiples.

III. DPM

A. Windowing and Transfer Functions

In the following discussion, only leading-edge sampling digital PWM is presented although DPM can also be implemented for use with double-edge sampling PWM.

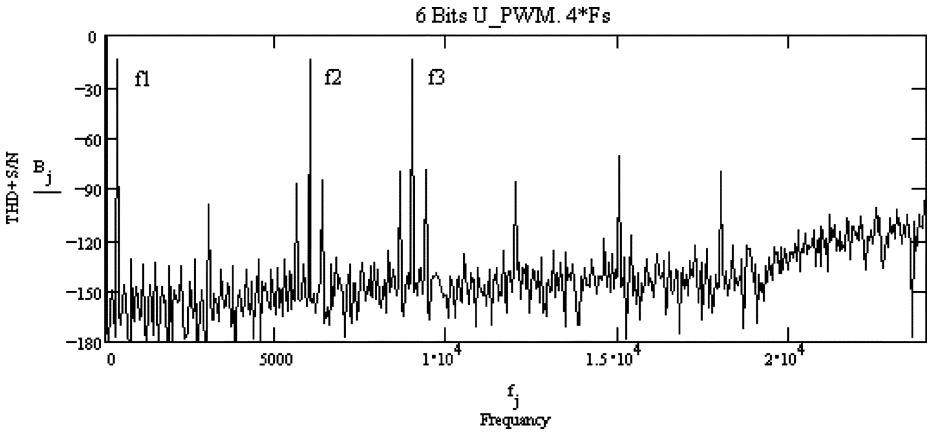


Fig. 5. Intermodulation distortion simulation results, the input frequencies are $f_1 = 0.75$ kHz, $f_2 = 6$ kHz, and $f_3 = 9$ kHz.

DPM is a dynamic switch operation map applied to digital PWM that is composed of a circular structure of 2-level quantizers whose delay length varies with α where, $0 \leq \alpha < D$, consequently, D is the pulse-repetition period. Let a time sequence $\{\cdots, k = D, k+1 = 2D, k+2 = 3D, \cdots\}$ be defined to determine each PWM pulse edge sampling instant where each has the sampled grid $i = 0, 1, 2, \cdots, D-1$ used to upsample in between two consecutive sequences. For notational convenience, in this section, we normalize this time sequence so as D is set at 1, and hence, α is correspondingly fractional.

Based upon this iterative definition, (1) can be rewritten as a difference equation

$$y(\alpha, k+1) = y(\alpha, k) + \sum_{i=0}^{\alpha} \frac{\sin(\pi(k-i))}{\pi(k-i)} + \sum_{i=\alpha+1}^{D-1} \Xi(i) \quad (8)$$

where $\Xi(i) = 0$ or $\Xi(i) = -1$ by the PWM definition in (2). Further, let the impulse function in (8) be defined as

$$h_{\text{pwm}}(\alpha, k) = \sum_{i=0}^{\alpha} \frac{\sin(\pi(k-i))}{\pi(k-i)}. \quad (9)$$

The last term $\sin(\pi(k-\alpha))/\pi(k-\alpha)$ represents the transition edge, where the summation from 0 to α for the pulses defines the method of PWM.

In order to apply a windowing function to (9) by taking $N+1$ consecutive samples in k iterations, samples are denoted as integer $j \in (0, N)$ where they are composed of an odd-N Lagrange interpolation FIR filter on an iterative scale of k. Hence, the property can be applied that impulse responses h_{pwm} for $N - j$ are the same as those for j but placed in reverse order, that is,

$$h_{\text{pwm}}(\alpha, j) = \sum_{i=0}^{\alpha} \frac{\sin(\pi(j-i))}{\pi(j-i)} = (-1)^{N-j} \sum_{i=0}^{\alpha} \frac{\sin(\pi i)}{\pi(j-i)}. \quad (10)$$

In our example where $N = 3$, the sign of $h_{\text{pwm}}(\alpha, j)$ in (10) alternates as $\{-1, 1, -1\}$. Therefore, the PWM pulses that determine the transition edges should shift one repetition period to the right in order to prevent the sign of impulses changing during leading-edge sampling (see the dotted line in Fig. 6).

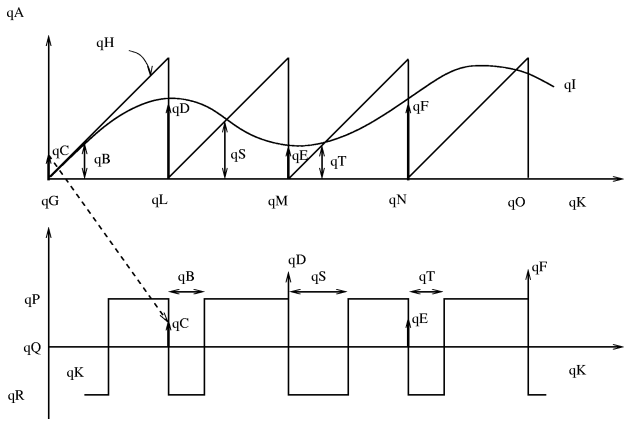


Fig. 6. Mapping is accomplished by an Lagrangian interpolation where the magnitude of modulating signal data α in each frames are linearly converted into time-domain data in leading-edge modulation.

Note that $h(\alpha, j)$ corresponds to the impulse response of DPM defined by the window and the Lagrangian interpolation FIR filter coefficient for Nth order polynomials, which has a classic form

$$h(\alpha, j) = \prod_{k=0,\ k \neq j}^{N} \frac{\alpha - k}{j - k}, \qquad \text{for } j \in \{0, 1, 2, \cdots, N\} \quad (11)$$

where $h(\alpha, j)$ has the Kronecker Delta function property

$$h(\alpha, j) = \begin{cases} 1, & \text{if } j = k \\ 0, & \text{if } j \neq k. \end{cases} \quad (12)$$

Once the Lagrange coefficient polynomial (11) is derived, it is unnecessary to calculate equations simultaneously as the new upsampled data fall on the grid obtained from the summation of products of each coefficient and the output values of the re-quantizer. In order to derive an explicit transfer function expression, take a binomial coefficient for (11) so that it determines the number of ways of choosing sampling time.

For a generic Lagrangian interpolation of (11), one should refer to [7] where several results for the binomial coefficients are introduced. When N is odd, it follows

$$h(\alpha,j)=(-1)^{N-j}\binom{\alpha}{j}\binom{\alpha-j-1}{N-j}$$
$$=(-1)^{N-j}\binom{\alpha}{L}\binom{N}{j}\frac{L}{\alpha-j} \tag{13}$$

since the following relations are satisfied for $N, j \in \mathbb{Z}$ where

$$\binom{\alpha-j-1}{N-j}=\binom{\alpha-j}{N-j}\frac{1}{\alpha-j} \tag{14}$$

and

$$\binom{\alpha}{j}\binom{\alpha-j}{N-j}=\binom{\alpha}{N}\binom{N}{j}. \tag{15}$$

During the upsampling operation that generates the digital PWM clocks, the *sinc* function under the window of length $L=N+1$ for the DPM Lagrangian FIR filter is

$$h(\alpha,j)=(-1)^{N-j}\sum_{i=0}^{\alpha}\left\{\frac{\sin(\pi i)}{\pi(j-i)}\binom{\alpha}{L}\binom{N}{j}\frac{\pi L}{\sin(\pi i)}\right\} \tag{16}$$

in which the two terms, $W_b(j)$ and $C_b(\alpha)$ are defined as

$$W_b(j)=\binom{N}{j} \tag{17}$$

$$C_b(\alpha)=\sum_{i=0}^{\alpha}\left\{\binom{\alpha}{L}\frac{\pi L}{\sin(\pi i)}\right\} \tag{18}$$

is the scaling coefficient and forms an intrinsic equalization function with regard to the digital PWM, which eliminates the intrinsic spectral distortions. This equalization process can be viewed as a pre-processing stage formed by a FIR digital filter whose impulse response compensates for the frequency response of h_{ppm}. Fig. 6(b) depicts frequency-domain magnitude weighting, which can be interpreted as a pre-compensation filter to the digital PWM.

B. Mapping and Implementation

The implementation of (16) for DPM is accomplished using a third-order Lagrangian interpolator. A limit needs to be imposed on the interpolation filter whose normalized gain must not exceed 1 to prevent additional distortion by over modulation resulting from an excessive modulation index. The Farrow structure fractional delay FIR filter (see [8] and [14] for details) is selected for the Lagrangian interpolation filters for DPM in which the coefficients are dynamically adjusted by the reference signal $r(i)$ stored in ROM. Fig. 7 shows the direct mapping principle in which the PWM reference signal $r(i)$ consists of unit-tread staircase $y_1(i)$ is the re-quantized signal at the upsampling instants of the Lagrangian Farrow structure interpolator for which there must be only a single solution to detect the position of the sampling instant for α.

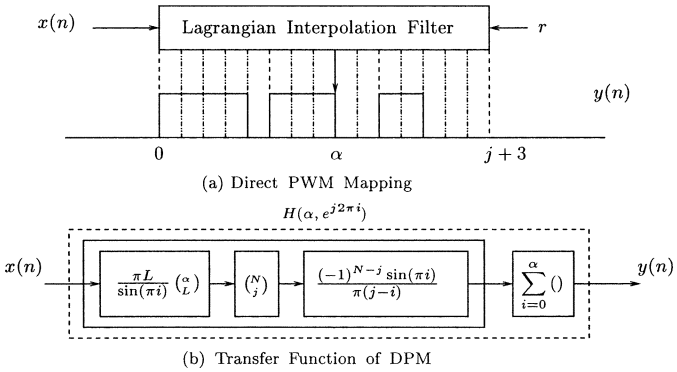


Fig. 7. Direct PWM mapping system represented as a pre-compensation filter and PWM. (a) The Lagrangian interpolation FIR filter consists of the PWM frame where each PPM signal is summed over α. (b) Depicts the transfer function diagram derived from (16).

TABLE I
SAMPLED EXAMPLE OF PWM REFERENCE SIGNAL

t_k	t_{k+1}	t_{k+2}	t_{k+3}
0.4627151	0.6008112	0.7565712	0.8775633

TABLE II
4-BIT DIRECT PWM MAP EXAMPLE BY TABLE I

PWM bit	Interpolation	N.S output	PWM output
16	0.7565712	0.6875	1
15	0.7473108	0.6875	1
14	0.7379273	0.8125	1
13	0.7284337	0.75	1
12	0.7188426	0.8125	1
11	0.7091669	0.625	1
10	0.6994194	0.625	1
9	0.6896129	0.8125	-1, 0
8	0.6797602	0.6875	-1, 0
7	0.6698741	0.5625	-1, 0
6	0.6599674	0.625	-1, 0
5	0.6500529	0.6875	-1, 0
4	0.6401433	0.6875	-1, 0
3	0.6302516	0.6875	-1, 0
2	0.6203905	0.5625	-1, 0
1	0.6105727	0.625	-1, 0

For example, the points at each sample periods before interpolation are shown in Table I, which is taken from a simulation result of the schematic in Fig. 8. α is uniformly sampled and each value provides information on the pulse transition instant of the PWM output signal after one sample delay of the pulse-repetition period. A leading-edge example is illustrated in Fig. 7 where the rising pulse sampling instant is passed to the DPM output signal. Both Tables I and II present examples of 4-bit DPM where t_k, t_{k+1}, t_{k+2}, and t_{k+3} are internal state values of 3-tapped delay line for the Lagrangian Farrow structure FIR filter. In this sequence, the interpolated samples correspond to each digitized 2^4 PWM values. Looking at the period between t_{k+1} and t_{k+2}, the upsampled 16 values can be obtained which are displayed in the Interpolation column in Table II.

The complete digital-to-analog (D/A) conversion system for an audio application is shown in Fig. 8. The pre-compensation FIR filter has a 3-tap delay line and is synchronously controlled

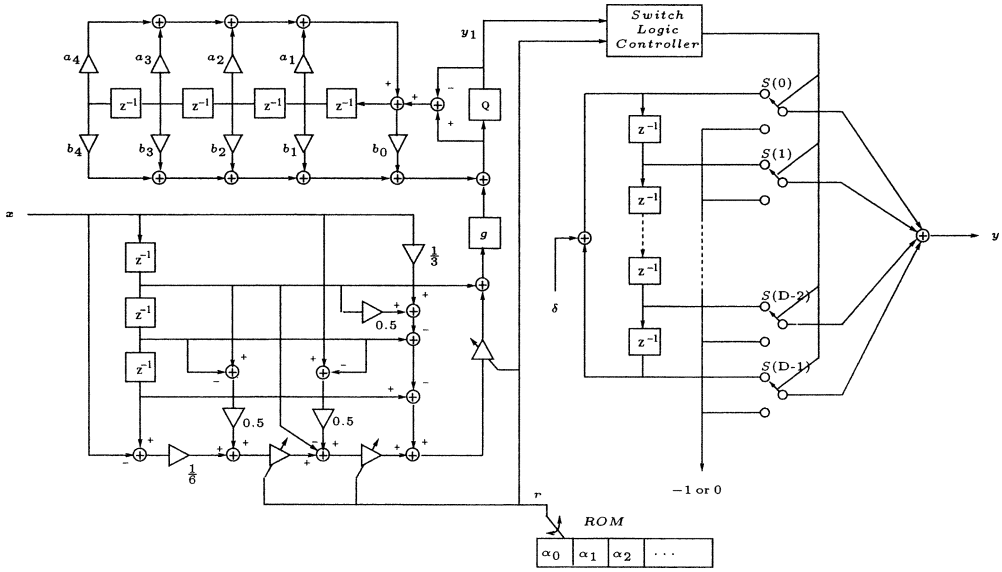


Fig. 8. Direct PWM mapping schematic for $4 \cdot f_s$ input 4-bit PWM D/A converter example.

by r which is stored in the ROM. The output of the Lagrange interpolation filter is scaled by a gain g in order to match the input range of the noise shaping re-quantizer and its output y_1 drives the digital PWM. This re-quantization from the PCM signal, using the *floor* rounding function that returns the greatest integer from the quantizer, has a D-bit range and is given by

$$Q(x) = \frac{\text{floor}(D \cdot x + 0.5)}{D}. \tag{19}$$

The coefficients a_n and b_n in the noise-shaping re-quantization block are set according to the noise shaping system design requirements; a_n should have negative values while b_n positive, etc. Due to the fourth-order noise shaping loop used in this example, which goes unstable unless the overall loop gain is kept to less than 0.5, the coefficient calculation of the noise shaper must be carefully chosen. Although a fourth-order noise shaping re-quantizer is used here, the DPM can take noise-shaping schemes of any type and order. The main idea of using noise-shaping re-quantization is to provide compression of the input PCM signal for the digital PWM. However, the performance of the noise shaping re-quantization has a direct effect on the output signal-to-noise (S/N) ratio.

TABLE III
COMPARISON OF ERROR CORRECTION METHODS OF FOUR PWM DACS

	Linearization Methods	Implementation Methods
PA	Interpolation to approximate analogue PWM - Time Domain	Interpolation and digital signal processing for Newton-Rapson algorithm
MFIR	Frequency moment matching - Frequency Domain	5 tap FIR filter in form of Toeplitz matrix
PNS	Frequency moment matching -Frequency Domain	3 tap FIR filter(ROM) in feedback loop and look-ahead scheme
DPM	Interpolation and windowing -Time and Frequency both	Farrow structure 3 tap FIR filter and digital PWM mapping

TABLE IV
COMPARISON OF SPECIFICATIONS OF FOUR PWM DACS

	PA	MFIR	PNS	DPM	SDM
Sampling Frequency	48kHz	48kHz	48kHz	48kHz	48kHz
PWM Bits	8	10	4	6	1
Oversampling	$8{\cdot}f_s$	$4{\cdot}f_s$	$64{\cdot}f_s$	$4{\cdot}f_s$	$256{\cdot}f_s$
Clock Rate (MHz)	98.304	196.608	49.152	12.288	12.288
THD+S/N	-125dB	-110dB	-160dB	-120dB	-112dB
NS order	5th	(2+2)th	5th	4th	2nd

C. Features and Comparison

The DPM operation can be summarized as D-bit PWM whose 2^D times upsampled values are located within the pulse-repetition period after filtering the input signal by the 3-tap Farrow structure FIR filter. The output of the Farrow structure FIR filter is fed to a noise shaping re-quantizer in order for the input signal to be compressed to D-bit PWM. DPM may be compared with earlier PWM linearization techniques that have been developed previously using time and frequency-domain methods. Time-domain PWM linearization methods have appeared in several studies [2], [4], [5], [12]. These papers demonstrated the typical error found in uniformly sampled PWM and introduced an improved sampling process for digital PWM power amplifiers [5], [12]; to conclude, possible remedies were suggested using upsampling introduced between two consecutive samples to locate the digital PWM edges. However, for finer accuracy, [2], [4] raised the concept of using the polynomial approximation method. On the other hand, the frequency-domain methods are studied in [1], [3], [10] and are based on the concept of moment matching.

In Tables III and IV, outline principles and example results are summarized that correspond to the respective papers, [1]–[3]. However, it should be noted that the clock rates given in Table IV are not related directly to the error correction method and corresponding performance but to the implementation differences. For brevity, the following terms are used; moment matching FIR

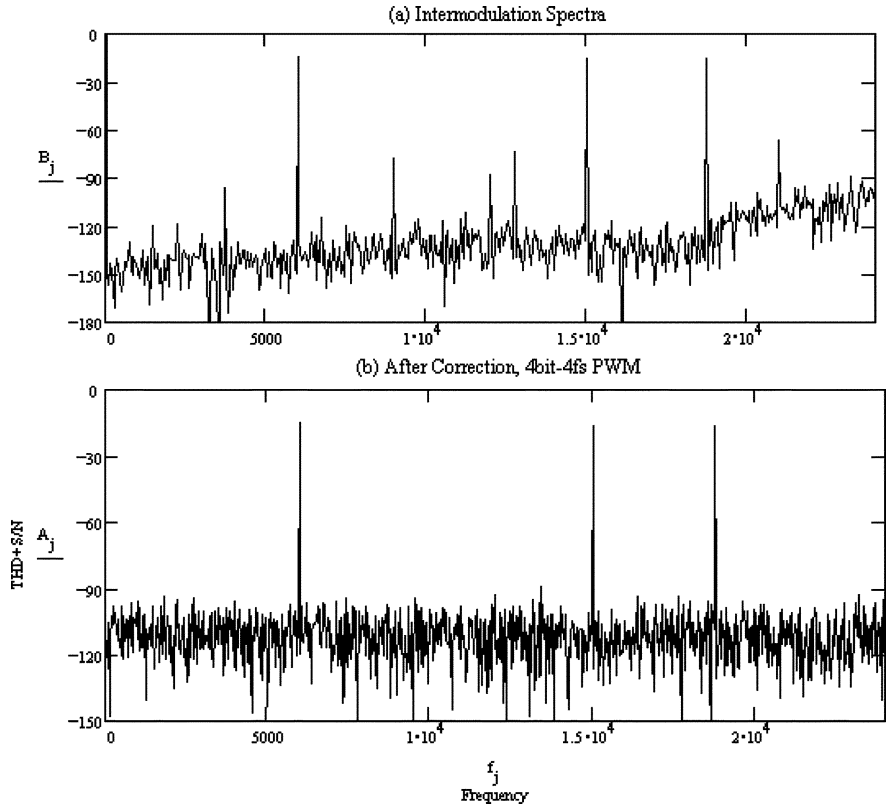


Fig. 9. Simulation results of 4-bit $4f_s$ PWM system (a) uncorrected (b) after correction.

filtering (MFIR) for the paper [1], polynomial approximation (PA) [2], procrastinate noise shaping (PNS) [3], and SDM.

The principal features of the DPM method are as follows.

1) In the time domain, DPM has better computational efficiency since it does not require iterative calculation steps for cross points or roots finding of two polynomials which is used in PA method.
2) In the frequency domain, its implementation is simpler because a 1-tap or 3-tap FIR filter is sufficient for the cascaded pre-compensation FIR filter, compared with MFIR and PNS methods.
3) Unlike SDM adopting 2-level quantization, PWM-based D/A converters (DACs) are known to have a distinctive feature in that the quantization overload problem can be relaxed due to higher bit quantization, [3], this is important in real-time applications although recent developments in Trellis SDM should be observed [17], [18].

The concept behind DPM can possibly be adapted to all forms of PWM's. For example, for double-edge modulation, it should use an even-order (2∼4 tap) Farrow structure Lagrangian interpolation FIR filter. A series of simulations were performed to test the error correction performance of DPM. Taking into account typical DACs; input signals are upsampled by 4 to 8 times.

IV. Simulation Results

In initial investigations, PWM with 4∼8 bit resolution were tested using the processes shown in Fig. 8. Fig. 9(a) shows simulation result of 4-bit, $4f_s$ PWM system with a four times oversampled input signal, where $f_s = 48$ kHz. This uncorrected PWM has an overall system clock rate 3.072 MHz, which is significantly lower than that of the SDM in Table IV. To observe the harmonic and intermodulation distortion in the base-band three superimposed input signals of 6, 15, and 18.75 kHz were used. After DPM operation, Fig. 9(b) reveals that signal purity is preserved more accurately. Following a series of investigations, it was concluded that 3.072 MHz is the minimum system clock frequency required to facilitate appropriate correction for the system; 4-bit PWM fed by $4f_s$ pre-oversampled input signal where $f_s = 48$ kHz.

Next, Fig. 10 presents results where the overall system bit rates are increased to 6.144 and 12.288 MHz, respectively, as the PWM bit resolution is increased from 4 to 5 and 6 bits. With the same superimposed input signal as the earlier example, the PWM intrinsic error is corrected more completely. In Fig. 10 (a), the residual noise and distortion is marked around -120 dB, hence this fact leads us to conclude that the DPM operated at 5-bit $4f_s$ can satisfy the required S/N ratio for high-quality audio applications. The 6-bit $4f_s$ DPM DAC shown in the subfigure (b) has the same system clock rate 12.288 MHz as the commercialized SDM DAC, which is the system proposed in this paper.

Finally, broad-band spectra are considered. Fig. 11 shows two PWM modes 6-bit $4f_s$ and 8-bit $4f_s$, where $f_s = 48$ kHz. The shaped quantization noise and distortion in the high frequency region are chosen so as not to be emphasized significantly where even the highest noise level is lowered to around -100 dB, see the subfigure (b). This is due to one of the distinctive features of the SDM+PWM structure compared to SDM only and can be a desirable characteristic for high resolution DAC applications.

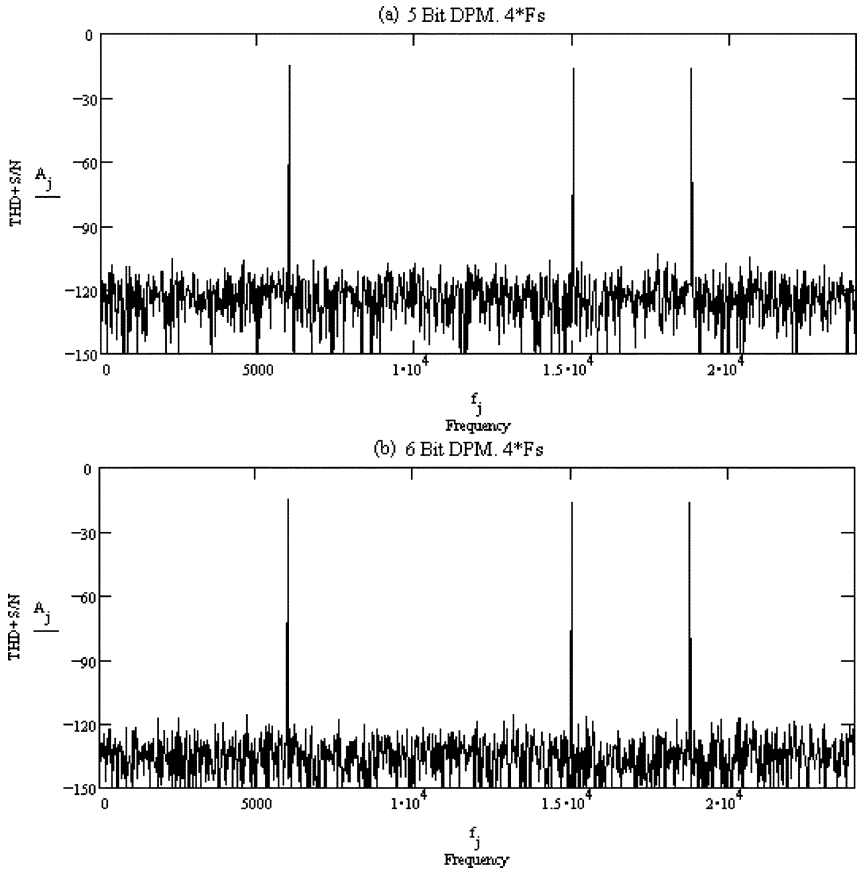


Fig. 10. Simulation results of (a) 5-bit $4f_s$ and (b) 6-bit $4f_s$ PWM systems after error correction by the DPM.

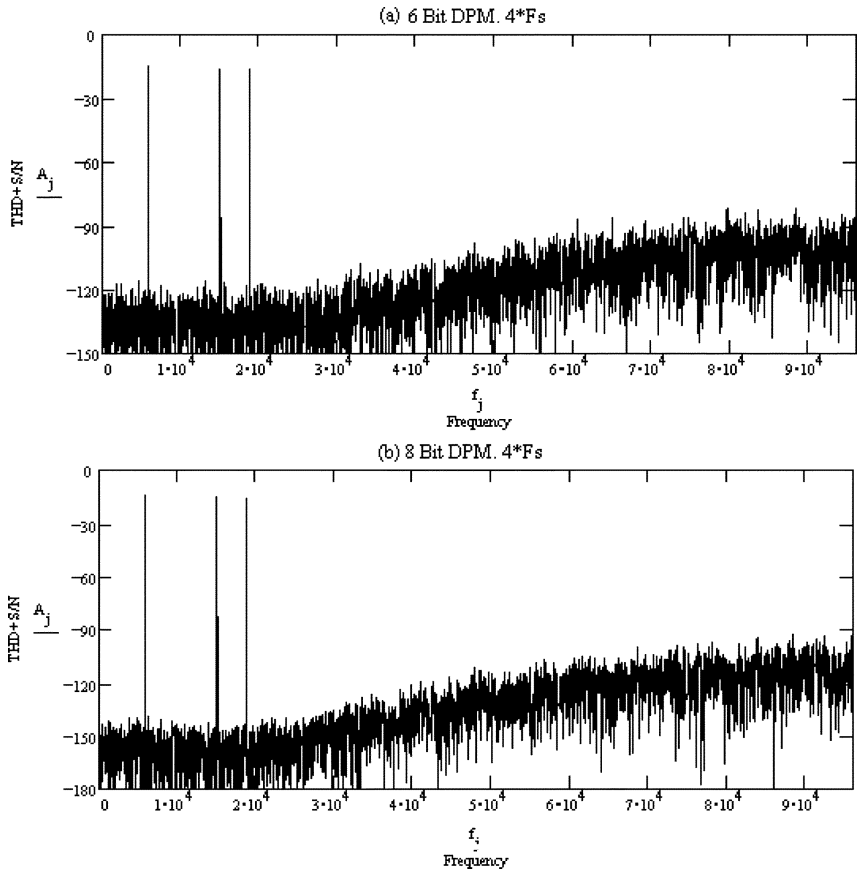


Fig. 11. Simulation results of broad-band spectra of (a) 6-bit $4f_s$ and (b) 8-bit $4f_s$ PWM systems after error correction by the DPM.

V. Summary

A novel error correction scheme for digital PWM has been investigated for use in high-resolution D/A conversion. The main discussion focused on the intrinsic error cancellation achieved using a window function with Lagrangian interpolation in a third-order Farrow structure fractional FIR filter synchronously combined with a digital PWM configured in circular formation. This method is shown to offer advantages against other previously proposed PWM linearization methods.

Although performance estimates are derived from a series of computer simulations, it is shown that relatively high-bit resolution digital PWM with a proper pre-compensation algorithm can outperform SDM; assuming similar system operating environments. Higher resolution D/A conversion can be achieved due not only to the relaxation of the 2-level quantization overload problem but also the significant reduction of noise and distortion generation in high frequency region. In addition, no high-order upsampling digital filter (typically 128 to 256 times) is required compared to SDM, since the third-order Farrow structure fractional FIR filter performs both pre-compensation and upsampling functions in one structure.

References

[1] M. J. Hawksford, "Dynamic model-based linearization of quantized pulsewidth modulation for applications in digital-to-analog conversion and digital power amplifier systems," *J. Audio Eng. Soc.*, vol. 40, no. 4, pp. 235–252, 1992.

[2] J. M. Goldberg, "Signal processing for high resolution pulsewidth modulation based digital-to-analog conversion," Dept. Elect. Eng., Ph.D. dissertation, King's College London, London, U.K, 1992.

[3] P. G. Craven, "Toward the 24 bit DAC: Novel noise-shaping topologies incorporating correction for the nonlinearity in a PWM output stage," *J. Audio Eng. Soc.*, vol. 41, no. 5, pp. 291–313, 1993.

[4] J. Goldberg and M. B. Sandler, "Comparison of PWM modulation techniques for digital amplifiers," in *Proc. Inst. Acoust.*, vol. 12, 1990, pp. 57–65.

[5] P. H. Mellor, S. P. Leigh, and B. M. G. Cheetham, "Improved sampling process for a digital pulsewidth modulated Class D power amplifier," in *Proc. IEE Colloq. Digital Audio Signal Processing*, London, U.K., 1991.

[6] K. Uchimura *et al.*, "VLSI A to D and D to A converter with multistage noise shaping," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP'86)*, Apr. 1986, pp. 1545–1548.

[7] P. J. Kootsookos and R. C. Williamson, "FIR approximation of fractional sample delay systems," *IEEE Trans. Circuits Syst. II*, vol. 43, pp. 269–271, Mar. 1996.

[8] T. I. Laakso, V. Valimaki, M. Karjalinen, and U. K. Laine, "Splitting the unit delay," *IEEE Signal Processing Mag.*, vol. 13, pp. 30–60, Jan. 1996.

[9] A. C. Paul, "A cathedral-2 implementation of a pre-compensation algorithm for use in a PWM DAC," King's College London, London, U.K., Internal Rep., 1993.

[10] M. J. Hawksford, "Linearization of multilevel, multiwidth digital PWM with applications in digital-to-analog conversion," *J. Audio Eng. Soc.*, vol. 43, no. 10, pp. 787–798, 1995.

[11] H. E. Rowe, *Signals and Noise in Communication Systems*. London, U.K.: D. Van Nostrand, 1965.

[12] S. P. Leigh, "Pulsewidth modulation sampling process for digital Class D amplification," Ph.D. thesis, University of Liverpool, Liverpool, U.K., 1991.

[13] V. Valimaki, "A new filter implementation strategy for Lagrange interpolation," in *Proc. IEEE Int. Symp. Circuits Systems(ISCAS'95)*, vol. 1, 1995, pp. 362–364.

[14] C. W. Farrow, "A continuously variable digital delay element," in *Proc. IEEE Int. Symp. Circuits Systems (ISCAS'88)*, vol. 3, June 1988, pp. 2641–2645.

[15] G. D. Cain, N. P. Murphy, and A. Tarczynski, "Evaluation of several FIR fractional—sample delay filters," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP'94)*, vol. 3, Apr. 1994, pp. 621–624.

[16] T. I. Laakso, V. Valmaki, and J. Henrikson, "Tunable downsampling using fractional delay filters with applications to digital TV transmission," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP'95)*, vol. 2, May 1995, pp. 1304–1307.

[17] H. Kato, "Trellis noise-shaping converters and 1-bit digital audio," in *Proc. 112th AES Convention*, Munich, Germany, Mar. 10–13, 2002, paper 5615.

[18] E. Janssen and D. Reefman, "Advances in Trellis based SDM structures," in *Proc. 115th AES Convention*, New York, Oct. 10–13, 2003.

Jin-Whi Jung received the B.Sc. degree in electronics engineering from Dong-A University, Busan, South Korea, in 1985, the M.Sc. degree (with dissertation) in electronic systems engineering from the University of Essex, Colchester, U.K., in 1997. He is currently working toward the Ph.D. degree in nonlinear dynamics at the University College London, London, U.K.

His research interests include digital telecommunication circuits and systems, audio engineering, and application-specified integrated circuit designs.

Malcolm J. Hawksford received the B.Sc. degree with first class honors and the Ph.D. degree from the University of Aston, Birmingham, U.K., in 1968 and 1972, respectively.

He is the Director of the Centre for Audio Research and Engineering, and a Professor in the Department of Electronic Systems Engineering, Essex University, Colchester, U.K., where his research and teaching interests encompass audio engineering, electronic circuit design, and signal processing. While studying for his doctoral degree, he worked on delta modulation and sigma–delta modulation (SDM) for color television applications, and invented a digital time-compression/time-multiplex technique for combining luminance and chrominance signals, a forerunner of the MAC/DMAC video system. His research covers both analog and digital systems with a strong emphasis on audio systems including signal processing and loudspeaker technology. Since 1982, his research into digital crossover networks and equalization for loudspeakers has resulted in an advanced digital and active loudspeaker system being designed at Essex University. A "first" in 1986 was for a prototype system to be demonstrated at the Canon Research Centre, Tokyo, Japan, research that had been sponsored by a research contract from Canon. Much of this work has appeared in the *Audio Engineering Society (AES) Journal* together with a substantial number of contributions at AES conventions. His research has encompassed oversampling and noise-shaping techniques applied to analog-to-digital and digital-to-analog conversion with special emphasis on SDM and its application to super audio compact disc (SACD) technology. In addition, his research has included the linearization of pulsewidth-modulated encoders, diffuse loudspeaker technology, array loudspeaker systems and three-dimensional spatial audio and telepresence including scalable multichannel sound reproduction. He is a Technical Consultant for NXT, U.K., and for LFD Audio, U.K.

Dr. Hawksford was a recipient of the BBC Research Scholarship for his Ph.D. studies, is a recipient of the publication award of the AES for his paper entitled "Digital Signal Processing Tools for Loudspeaker Evaluation and Discrete-Time Crossover Design," for the best contribution by an author of any age to volumes 45 and 46 of the *AES Journal*. He is currently chairman of the AES Technical Committee on High-Resolution Audio and is a Founder Member of the Acoustic Renaissance for Audio (ARA), and a Technical Adviser for *Hi-Fi News and Record Review*. He is a Chartered Engineer, a Fellow of the AES, a Fellow of the Institution of Electrical Engineers, U.K., and a Fellow of the Institute of Acoustics (IOA).

Audio Engineering Society

Convention Paper

Presented at the 117th Convention
2004 October 28–31 San Francisco, CA, USA

Digital audio power amplifier for DSD data streams

F.M. Prime[1] and M.O.J. Hawksford[2]

[1] KEF Audio, Maidstone, Kent, ME15 6QP, UK
Francis.Prime@kef.com

[2] University of Essex, Colchester, Essex, CO4 3SQ, UK
mjh@essex.ac.uk

ABSTRACT

A digital power amplifier topology is proposed optimized specifically for use with DSD-type data streams. The configuration enables direct interfacing of DSD data with no requirement for intermediate signal processing or analogue-to-digital conversion. The output architecture exploits a classic H-bridge configuration and uses a novel form of ac data coupling to simplify internal interface circuitry. Wide range gain control is enabled through modulation of the output-stage power supply voltage that also improves power efficiency at low gain settings. Consideration is given to finite pulse rise time and a modified DSD data format is investigated.

1. INTRODUCTION

This paper presents a class-D audio power amplifier topology that is designed specifically to be digitally interfaced to a 1-bit DSD data stream [1]. As such the power amplifier output stage constitutes directly a digital-to-analogue converter where because the 1-bit data stream is unmodified, the high-resolution SACD attributes including wide bandwidth and wide dynamic range can be fully exploited. Other than the implementation of gain control and certain aspects of the power supply and output filter, all analogue processing and digital processing in the signal path is omitted. The principle is followed that any processing especially that performed on the DSD data stream increases system complexity with the potential for introducing distortion and signal coloration. The intention therefore is to maintain a high degree of signal integrity and to remain faithful to the SACD philosophy by keeping the digital data unmodified within the 1-bit domain and throughout the amplifier chain.

In this scheme an open-loop MOSFET H-bridge [2] is driven directly by the 1-bit DSD data where the high

level output is bandlimited subsequently by a low-order low-pass reconstruction filter [3]. Gain control is implemented by using pulse amplitude modulation (PAM) where by integrating this feature into the output-stage power switch and drive circuitry, the conceptual simplicity of the overall amplifier can be maintained. Although the use of DSD data to drive the output stage reduces processing complexity, it also introduces a number of practical issues that need to be addressed. Most notably the high sampling rate of DSD (2.8224 MHz) [1] poses problems with respect to switching large voltages at high frequency [4]. This occurs because a power MOSFET performs less like an ideal voltage-controlled current switch, especially when operated at higher switching rates. Consequently switching speed is impeded, power efficiency sacrificed and distortion introduced due to imperfections in the output waveform under finite transient switching conditions. The focus of this work is the design of a topology that maintains low distortion, high efficiency and incorporates wide range gain control. The circuit topology is presented in Figure 1 and the important aspects of the design are discussed. Both measured and simulation results are presented based upon an early prototype to demonstrate the operation and efficacy of this particular design approach.

Finally, further discussion and simulation of alternative topologies exploiting zero-voltage switching and resonant-mode power supplies are presented that include features designed to reduce transient power dissipation and EMC factors caused by high frequency signal components.

2. DSD POWER AMPLIFIER CIRCUIT TOPOLOGY

The overall structure of this design is shown in the general schematic diagram in Figure 1. The various sections in the design are discussed in this paper with particular emphasis placed upon the power-switch driver circuitry that allows efficient switching under a wide range of gain control. Amplifier gain control is introduced by the expedience of modulating the output-stage power supply voltage that is used to determine directly the amplitude of the output square-wave signal by employing PAM. Lowering the power supply voltage also has the desirable effect of reducing output-stage power dissipation when the gain setting is low as the voltage swing across output transistors is proportional to the power supply voltage.

The complete amplifier topology designated here as Type 1, is shown in Figure 2. Considerable effort was directed at realizing a simple yet efficient amplifier structure where a key feature required of the driver circuitry is to allow the output devices to switch completely even when the output-stage power supply approaches zero volts. It was also important for the driver circuitry to have a fast response time while retaining low internal power dissipation.

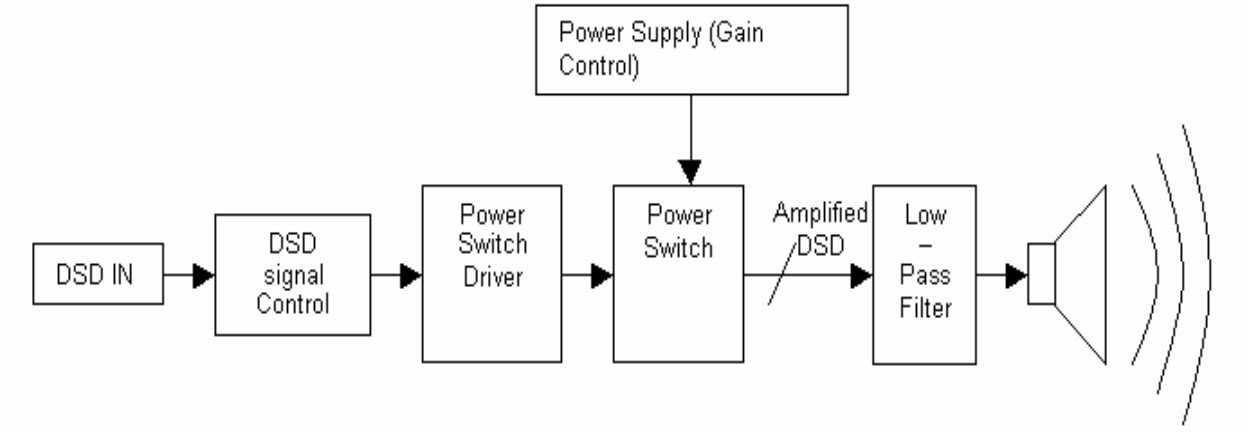


Figure 1 General digital amplifier system overview.

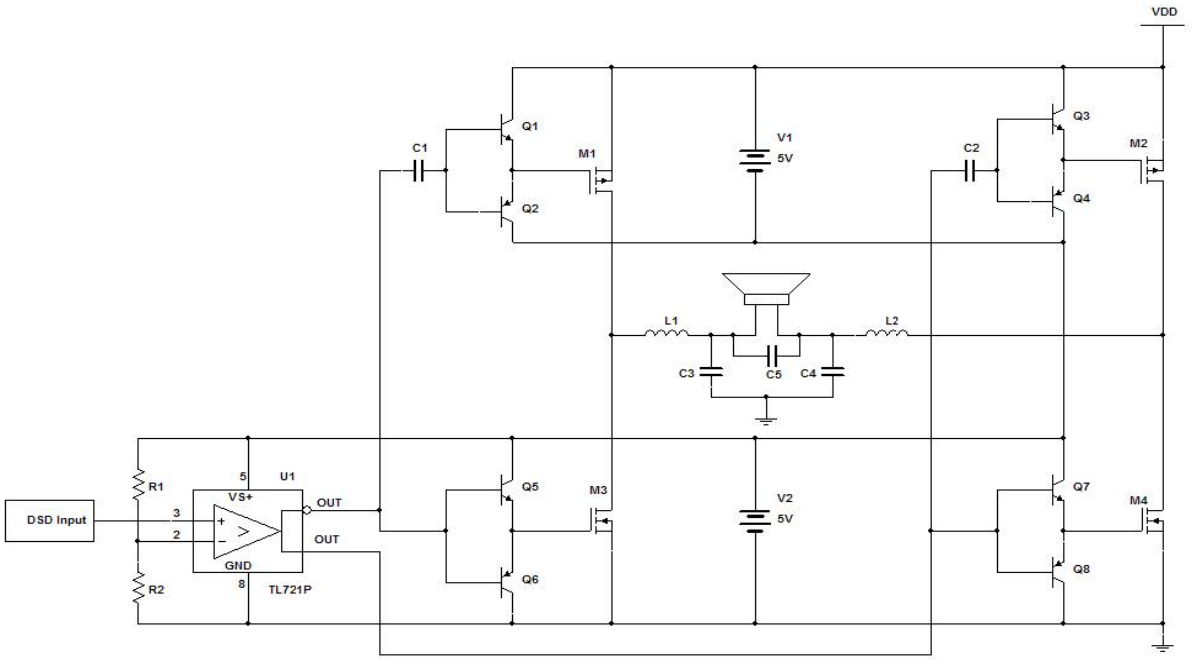


Figure 2 Type 1 DSD power DAC circuit topology generating complementary square-wave output voltages.

2.1 Output-stage H-bridge power switch

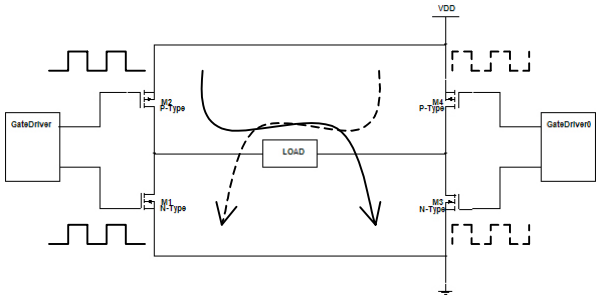


Figure 3 H-bridge MOSFET power switch.

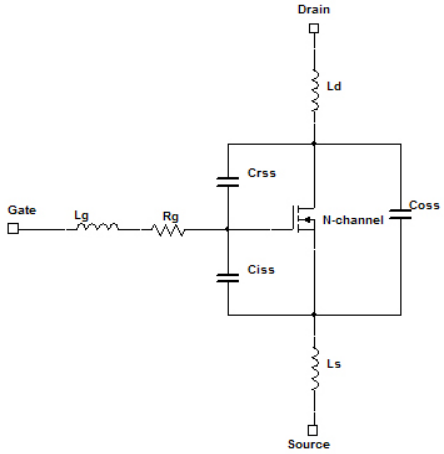
The output-stage power switch shown in Figure 3 is based upon the classic H-bridge topology and employs two complementary pairs of N-channel and P-channel MOSFETs. This configuration provides a differential output signal in order to maximize output voltage swing for a given power supply voltage. It also offers rejection of common-mode noise and distortion.

Also, a principal advantage of the H-bridge is that it allows a single positive constant voltage supply rail that when varied forms PAM of the output signal to achieve gain control. In Figure 3 the arrows indicate the direction of current flow for alternate conduction of each pair of devices. Note also with this configuration, that when the average output voltage of each half of the bridge is zero, then because of common-mode rejection the gain control signal, that is the power supply voltage, does not appear across the load.

There are some key attributes of power MOSFETs that determine the ultimate switching speed of the power switch and power efficiency of the overall amplifier. It is therefore expedient to analyze the operation of the switches and to include their relevant parasitic elements.

2.1.1 MOSFET analysis

The MOSFETs within the H-bridge are required to operate as voltage-controlled current switches. Although an insulated gate suggests no current is taken by the gate terminal either when the device is in its ON or OFF state, in practice this is not the case since under high speed dynamic switching, intrinsic parasitic elements cause gate current I_G to flow during the switching transitions.



- $C_{ISS} = C_{GS} + C_{GD}$, with C_{DS} shorted
- $C_{RSS} = C_{GD}$
- $C_{DSS} = C_{DS} + C_{GD}$
- L_D, drain lead inductance
- L_S, source lead inductance
- L_G, gate lead inductance
- R_G, gate lead resistance

Figure 4 MOSFET parasitic model.

During switching the key parameter is the total gate charge Q_G that consists of two main components Q_{GS} due to C_{ISS} and Q_{GD} due to C_{RSS}. Also, the gate charge is temperature independent where the lower its value the faster the possible switching speed and the lower the switching losses.

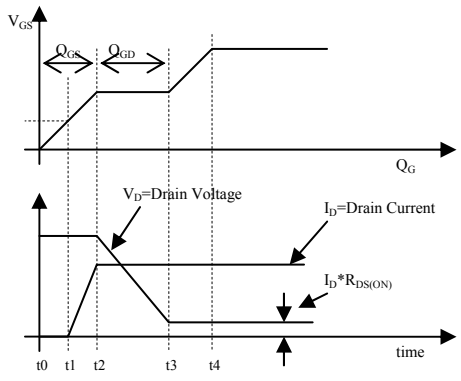


Figure 5 Gate-source voltage versus gate charge (top graph); drain voltage and drain current characteristics at turn-ON (bottom graph).

Using the appropriate V_{GS} versus Q_G graph obtained from MOSFET data, a corresponding gate charge can be read off for a given V_{GS}. The constant gate current I_G required to switch the MOSFET during a given transition time t_r can then follows as,

$$I_G = \frac{Q_G}{t_r} \quad \ldots 1$$

As well a calculating gate drive current versus transition time for a given gate charge, it is also useful to estimate the average gate input power dissipation P_{DRIVE} when driving the MOSFET, where if f_{DSD} Hz is the DSD related device switching frequency, then

$$P_{DRIVE} = Q_G V_{GS} f_{DSD} \quad \ldots 2$$

The conduction loss P_{COND} arising from the low resistance between drain and source whilst the device is in saturation also affects output-stage power efficiency. As revealed in Figure 5, V_D never quite reaches zero even while the MOSFET is fully ON. The small voltage drop $I_D R_{DS(ON)}$ represents the main conduction loss in the MOSFET where,

$$P_{COND} = P_{OUT}\left(\frac{R_{DS(ON)}}{R_L}\right) \quad \ldots 3$$

The on-resistance $R_{DS(ON)}$ is reduced during manufacture as much as possible and represents the key factor in determining power efficiency. Consequently, the efficiency and switching operation of a class-D amplifier is critically dependent on the characteristics of the power MOSFETs.

2.2 MOSFET H-bridge driver

A core feature of DSD power amplifier is the circuitry that provides the drive signals to the H-bridge MOSFET output stage. Analysis in Section 2.1 has shown that MOSFETs only draw significant current from the gate during the switching transitions. To provide the required gate current over a given transition time without sacrificing efficiency requires driver stage output and bias currents to flow only when required so as not to dissipate unnecessary power. As shown in Figure 2, each MOSFET gate terminal is driven by a class-C, complementary emitter-follower circuit. The follower circuit provides both high current gain and

current drive capability to couple the low voltage, low current output of the high-speed DSD control stage to the MOSFET gate. This circuit has incorporated an emitter-follower circuit consisting of wide band bi-polar transistors; however MOSFET devices could also be used although this option was not explored.

In the next sub-section the drive circuitry for the N-channel devices is considered to explore the operation of the driver circuitry.

2.2.1 N-channel MOSFET drive circuit

The drive circuit shown in Figure 6 for an N-channel MOSFET is investigated assuming the following parameters,

- transition time, t_r = 10 ns
- V_{GS} = 5 V
- $V_{DS(max)}$ = 20 V

A typical gate charge Q_G = 15 nC for a N-channel MOSFET was taken from manufacturer's data describing the Q_G versus V_{GS} graph. The gate drive current I_G then follows from Eqn 1 as I_G = 1.5 A. This current is only drawn during the transition time of 10 ns and therefore as long as the current is sourced only for this time, then power efficiency can be kept high. Assuming f_{DSD} = 2.8224 MHz then the average gate power dissipation for this single device is calculated from Eqn 2 as P_{DRIVE} = 168 mW. Note that in practice the gate charge for a P-channel device can differ and depends upon device matching.

Alternatively, the gate current I_G can be considered by calculating the effective gate resistance R_G. R_{GT} represents a low impedance state when the device is in transition and is due to gate parasitic capacitances, while R_{GH} represents a high impedance state that occurs when the device is in saturation.

R_{GH} can be calculated from the I_{GSS} gate-to-source leakage current specified by manufacturer's data where I_{GSS} is given in terms of a V_{GS} typically as,

$$R_{GH} = \frac{V_{GS}}{I_{GSS}} = \frac{20V}{100nA} = 200M\Omega \qquad \ldots 4$$

R_{GT} can be determined using the applied V_{GS} and the gate current calculated above, where

$$R_{GT} = \frac{V_{GS}}{I_G} = \frac{5V}{1.5A} = 3.33\Omega \qquad \ldots 5$$

In Figure 7, straight line approximations of the current and voltage waveforms associated with the circuit in Figure 6 are shown assuming that ideal bi-polar transistors are used.

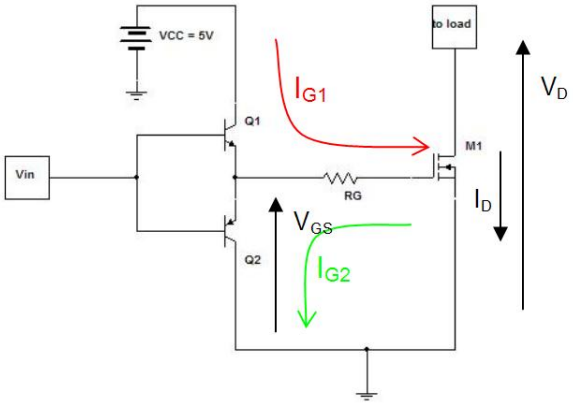


Figure 6 DC-coupled, complementary emitter-follower gate drive circuit for N-channel MOSFET.

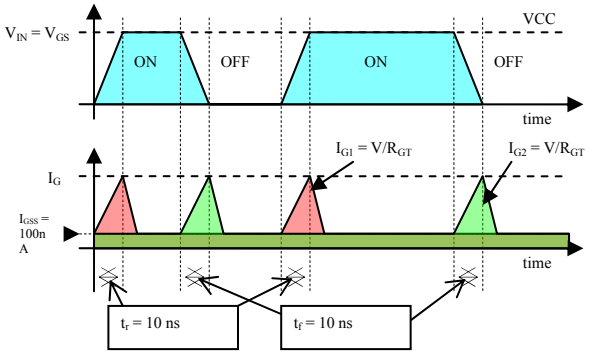


Figure 7 Key waveforms in gate drive circuit.

Transistor Q1 conducts during positive-slope input excursions while Q2 conducts during the negative-slope excursion. These complementary transistors require wide bandwidth to accommodate the DSD sampling rate, where the signal derived from the first-stage differential comparator (see Figure 2) are sufficient to drive the high impedance input of the emitter follower circuits. The complementary emitter-follower cross over region of $2*V_{BE}$ is of little consequence here as switching between conduction of one transistor to the other occurs in less than a nanosecond.

2.2.2 P-channel MOSFET drive circuit

A significant aspect of this amplifier design is the implementation of gain control by using PAM. This approach can cause a significant complication to the drive circuit of the P-channel MOSFETs used in the H-bridge output stage since unlike the N-channel drive circuit shown in Figure 6; the gate drive must now float over a wide range of supply voltage ranging from zero to tens of volt. This is required to accommodate the variable output-stage power supply voltage.

However, although the power supply can vary, the gate-to-source switching voltage used to switch the MOSFET from off state to saturation state must remain the same. Figure 2 shows that for the P-channel MOSFET, the associated complementary emitter-follower buffer amplifier must float with the power supply and therefore requires a local supply voltage that is both constant and referenced to the positive supply rail as shown in Figure 8.

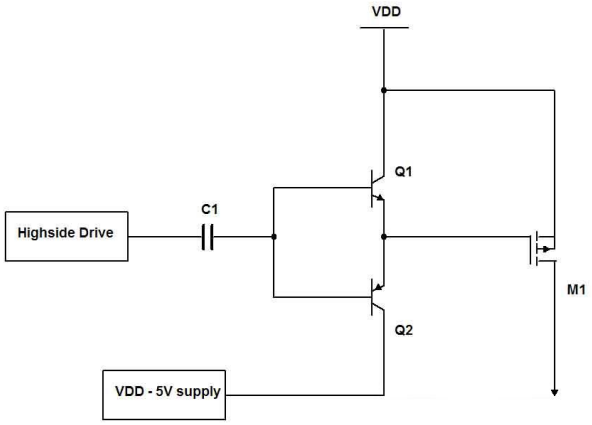


Figure 8 AC-coupled drive circuit for P-channel MOSFET.

To solve this problem and eliminate the requirement for intermediate level shifting buffer stages the simple expedience of incorporating AC-coupling was used as shown in Figure 8. This means that with changes in output-stage power supply voltage the coupling capacitor simply accommodates the voltage change while coupling the gate switching voltage without modification.

However, it is mandatory that the gate-source switching signal is correctly aligned to the output device switching range where this is achieved by diode clamping. The collector-base P-N junctions within the transistors Q1 and Q2 form effective clamping diodes that constrain the base voltage so it maintains mid-point symmetry, thus preventing the capacitor voltage from drifting. For example, if the output-stage power supply voltage is lowered then there is a tendency for the drive voltage on the bases of Q1 and Q2 to swing too positive. However, if this occurs then the base-collector diode of Q1 conducts and clamps the base voltage thus forcing the dc voltage across the coupling capacitor to change. Similarly, if the voltage drift were negative or the power supply voltage increased then a similar clamping action occurs with the base-collector diode Q2. Either way, the voltage on the right hand side of the coupling capacitor is steered to maintain the correct switching voltage range. This process also accommodates any differences between the collector-base current gains of Q1 and Q2 which otherwise would create partial rectification and again cause the drive voltage to drift out of range. The circuit therefore achieves self-centering of the gate drive voltage for the P-channel MOSFET, whereas the N-channel drive circuit being DC coupled, does not exhibit drift.

2.3 Input DSD receiver stage

In order to drive the two pairs of emitter-follower gate drive stages described in Section 2.2.1 and 2.2.2, time coherent complementary DSD data streams are required where this is realized using the comparator shown in Figure 9. Fast pulse transitions (< 10 ns) must be maintained and tight delay matching of the complementary pulse streams achieved, preferably < 1 ns. This strategy helps to reduce transient "shoot through" current between the complementary output devices located on each side of the H-bridge that result in additional current demand on the power supply together with output-stage heating and reduced power efficiency. For digital power amplifiers using pulse-width modulation (PWM), introducing dead time during pulse transitions can be used to eliminate "shoot through" currents [5], together with overall feedback to lower distortion as there is tradeoff between dead time and increased distortion. However, in the present DSD design an open-loop configuration is used, consequently dead time processing has not been implemented where performance relies on fast pulse transitions and accurate timing between transitions to mitigate effects of "shoot through".

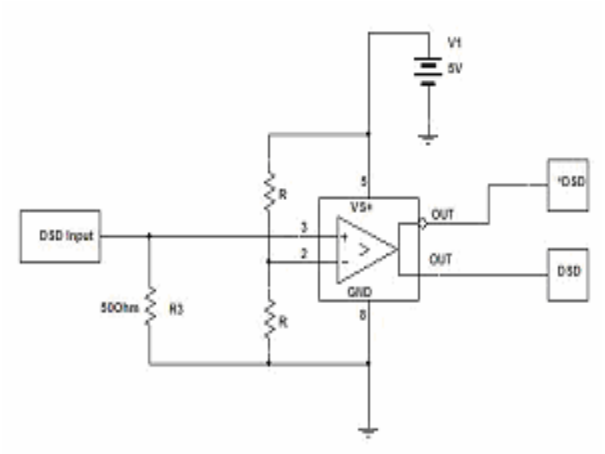


Figure 9 Symmetrical DSD signal generator circuit.

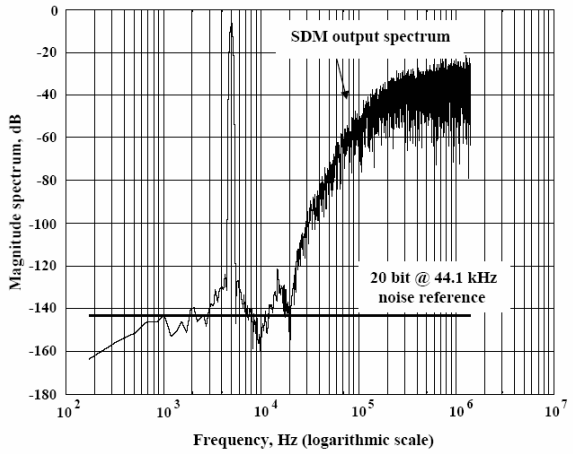


Figure 10 Output spectrum for Sony FF SDM [6].

The comparator shown in Figures 2 and 9, provides the required drive signals for the emitter-follower buffers that in turn switch the power MOSFETs, generating the high-power differential DSD output of the H-Bridge.

3. OUTPUT FILTER DESIGN

As with most class-D PWM amplifiers the final process is a passive low-order, low-pass filter that reconstitutes the analogue signal [3]. The output filter is designed to eliminate high-frequency switching artifacts including DSD code related noise and to reconstruct the audio prior to driving the speaker load. It is important to note however, that in a direct DSD power amplifier there is a significant level of high frequency noise that is inherent to the DSD code where this must be attenuated by the output filter to prevent high frequency signals from entering the loudspeaker and to reduce radio frequency (RF) radiation. Consequently, the H-bridge output stage should be in close physical proximity and directly coupled with short wires to the output filter. However, for a commercial design based upon the present topology further work is necessary to meet EMC regulations.

To illustrate the high frequency noise inherent to DSD code a typical DSD output signal spectrum derived from a Sony FF class sigma-delta modulator (SDM) encoder [6] is shown in Figure 10. Also, later in Section 5 some extensions to the present design are proposed that should help reduce RF interference as well as lead to improved output-stage power efficiency.

A low-pass Butterworth output filter was designed for use with the open-loop DSD power amplifier, where the filter is required to attenuate only the output signal above the audio band while retaining low-loss within the audio band. A passive second-order L-C filter was implemented employing low resistance air coil inductors selected to minimize power dissipation and to avoid high-frequency non-linearity and saturation that is inherent to ferrite inductors. A higher order filter is not required due to the high DSD switching frequency and the relatively relaxed noise shaping of the DSD code as depicted in Figure 10.

The basic filter specification is summarized:

- nominal low-pass Butterworth response assuming resistive load
- second-order LC filter
- fc = 20 kHz
- loudspeaker load (resistive) 8 Ω

The filter circuit is symmetrical as shown in Figure 11 in order to match the balanced output of the H-bridge stage where the transfer function *H(s)* is given as,

$$H(s) = \frac{1}{1 + s\left(\frac{2L}{R}\right) + s^2 3LC} \qquad \ldots 6$$

where *R* is the loudspeaker load resistance. Eqn 6 matches the standard second-order transfer function given by,

$$G(s) = \frac{1}{1 + 2\xi\frac{s}{\omega_n} + \left(\frac{s}{\omega_n}\right)^2} \quad \ldots 7$$

and,

$$\omega_n = \frac{1}{\sqrt{3LC}} rads^{-1} \text{ and } \xi = \frac{1}{R}\sqrt{\frac{L}{3C}}$$

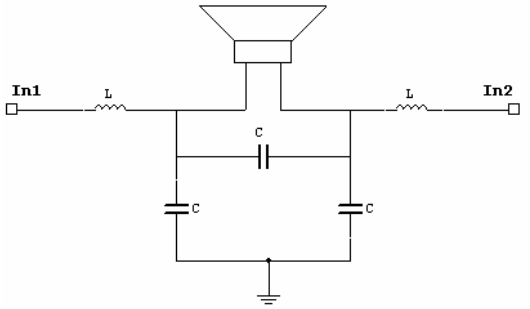


Figure 11 Symmetrical output filter topology.

4. SYSTEM MEASUREMENTS

To test the prototype amplifier DSD code was generated using a dCS972 digital-to-digital converter [7] where a 16-bit @ 44.1 kHz linear pulse-code modulation (LPCM) source file derived from compact disc (CD) was converted into DSD 1-bit @ 2.8224 MHz. A variable voltage source was used to power the H-bridge power switch circuit such that with a supply voltage of 20 V, DSD output pulses of 40 $V_{pk\text{-}pk}$ were generated. The 0 dB output DSD level is defined to be -6 dB such that an output sine wave for example, spans half the peak-to-peak amplitude of the DSD pulses, implying a SDM modulation factor of 0.5. This is a common requirement in DSD design as higher modulation factors can lead to instability and increased distortion.

4.1. Harmonic distortion

The harmonic distortion performance of the prototype amplifier was determined using a 0 dB DSD encoded sinusoid at 1 kHz applied to the amplifier and filter with an 8 ohm resistive load. The output signal level corresponded to a power dissipation of 6.25 watt. The normalized spectrum measured at the output of the filter is displayed on a spectrum analyzer and replicated in Figure 12 from which harmonic levels are determined.

Higher harmonics are not shown as they are lost within the system noise. The second harmonic appears at -75 dB demonstrating a reasonable distortion performance. Also, by summing harmonics to order four, a total harmonic distortion (THD) figure of 0.017% is achieved.

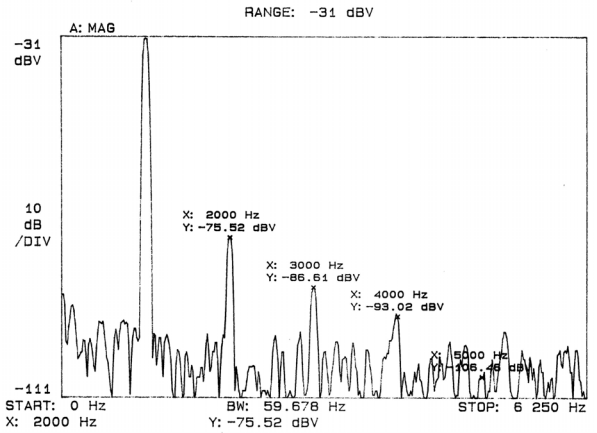


Figure 12 Harmonic distortion: output signal 6.25 watt into 8 ohm (spectral span 0 to 6 kHz).

4.2. 2-tone Intermodulation distortion

A CD was produced with a track containing summed 19 kHz and 20 kHz pure sinusoid tones. These tones are played via a CD player through the dCS972 enabling an intermodulation distortion test, where the results are shown in Figure 13. As shown, the suppression of intermodulation distortion is only ≈ -14 dB.

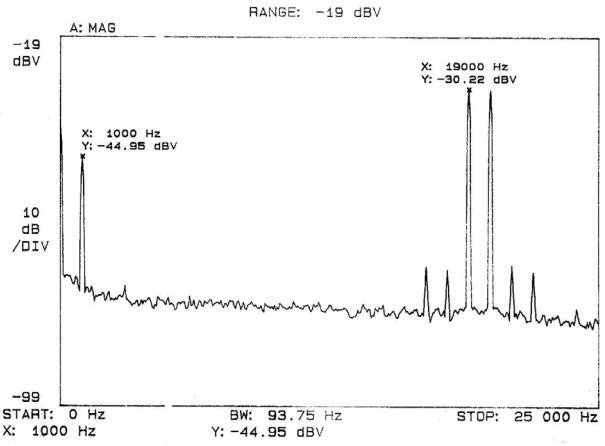


Figure 13 Intermodulation distortion: output signal 6.25 watt into 8 ohm (spectral span 0 – 25 kHz).

5. ENHANCED DSD TOPOLOGIES WITH LOW SWITCHING LOSSES

The design presented in this paper has shown how to configure a power H-bridge switch that can be used directly with a DSD data stream, where relatively straightforward drive circuitry is employed through the use of a self-clamping AC-coupled buffer. However, as with all class-D stages that are predicated on the generation of a square-wave output voltage, there are performance consequences in terms of power dissipation during the transient switching period of the output devices, related slew-induced distortion [8] and high frequency EMC problems again arising from rapid voltage changes.

In this Section a modified output stage is proposed with two variants designated Type 1 and Type 2 that aim to address these limitations and to offer improved performance in a number of critical areas. The viability of the amplifiers is investigated using *Matlab* simulation again incorporating DSD code generated from a Sony FF style SDM [6] algorithm. Both the two amplifier variants are designed to be used in association with a resonant-mode power supply. The resonant-mode power supply is assumed here to produce a pure sinusoidal output that can be combined with a constant voltage component such that the power supply voltage presented to the H-bridge stage appears as raised-cosine repetitive waveform of frequency f_{DSD} Hz given by,

$$V_S = gain\left(\frac{1 + \cos\left(2\pi f_{DSD} t\right)}{2}\right) \quad \ldots 8$$

where the control parameter *gain* defines the peak amplitude of the supply voltage. Eqn 8 describes a power supply voltage that swings between zero and a voltage *gain* volt that is in synchronism with the DSD sample clock, where being sinusoidal there is no harmonic content. Also, the voltage waveform V_S is phase locked so that its minima (i.e. 0 V instants) are aligned precisely to the DSD sampling instants. As with the prototype Type 1 amplifier shown in Figure 2, the programmable nature of the power supply voltage specified in terms of the control parameter *gain* is used to determine the voltage gain of the power amplifier though the use of PAM. This variable output voltage would in practice be achieved by appropriate design of resonant-mode power supply that would require negative feedback control to produce a stable and well-defined supply voltage.

5.1 Type 2 DSD power amplifier output stage

The conceptual model of the Type 2 DSD power amplifier proposal is shown in Figure 14 together with an illustration of the raised-cosine periodic supply voltage V_S. Effectively, the constant voltage power supply has been replaced by the output of the resonant-mode power supply whose voltage is defined by V_S in Eqn 8. This process replaces the rectangular pulse of the Type 1 amplifier with raised-cosine pulses. Consequently, the output pulse stream is pre-filtered and has lower spectral content above the DSD sampling rate. However, of greater significance is that H-bridge switching only occurs when the output devices experience virtually zero drain-source voltage thus dramatically reducing switch-induced power loss as switching between devices can be achieved when all drain-source voltage are small.

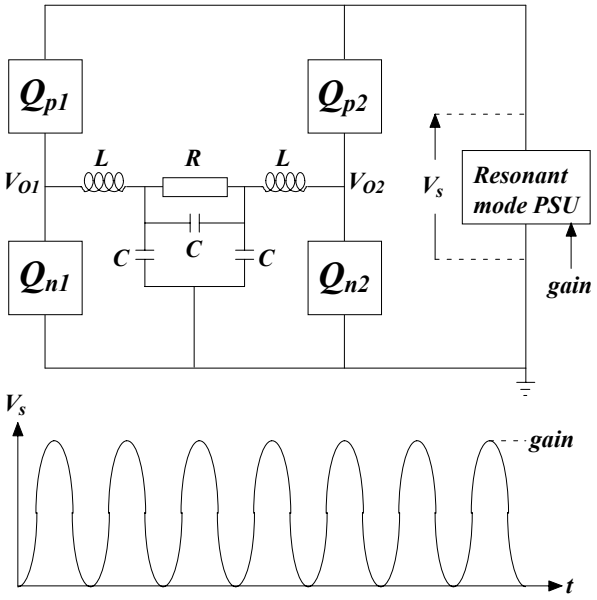


Figure 14 Type 2 output stage powered by variable-output, resonant-mode power supply.

On first encounter it may appear that modifying the pulse waveform will introduce non-linear distortion. However, this does not occur providing the modified symbol shape is the same for all pulses in the data stream and where any further waveform changes resulting for example, from low-pass filtering affect the

sample ensemble as a whole and do not give rise to inter-sample differences or pulse-sequence dependent memory effects.

To confirm this observation, simulations were performed for both a Type 1 and a Type 2 amplifier configurations where normalized time domain output waveforms are illustrated in Figure 15(a,b) with corresponding output spectra shown in Figure 16(a,b). Using oversampling techniques, non-rectangular output pulses can be accommodated and output spectra calculated in excess of the DSD sampling rate to show how using raised-cosine pulses reduces high frequency content. Similar spectra to that presented in Figure 10 can be observed with no evidence of additional in-band distortion resulting from waveform modification. However, it is evident that significant attenuation of high-frequency components produced by the Type 2 amplifier output spectrum has been achieved.

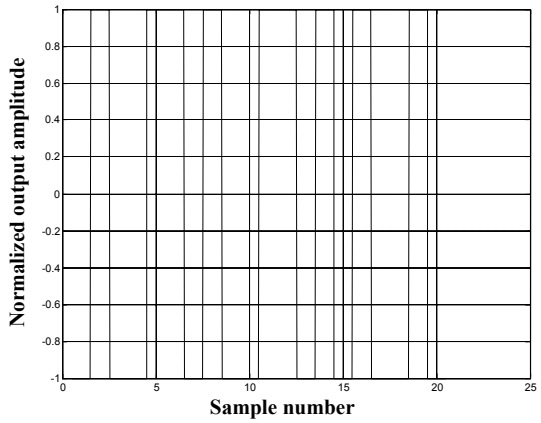


Figure 15a Type 1 time-domain output.

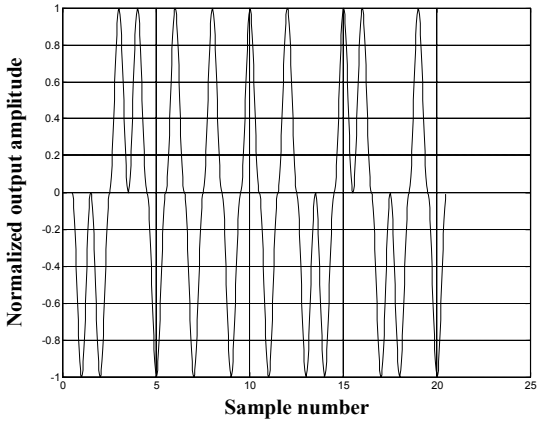


Figure 15b Type 2 time-domain output.

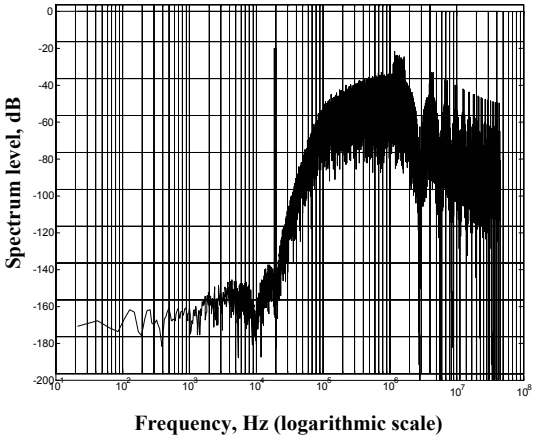


Figure 16a Type 1 frequency-domain output.

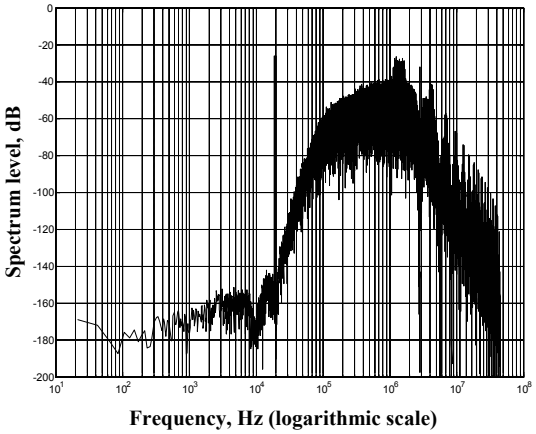


Figure 16b Type 2 frequency-domain output.

In the simulations presented in this Section a computation vector length of 2^17 was chosen and an additional oversampling factor of 32 applied to the DSD output code in order to accommodate the raised-cosine pulse shape. Figure 15b illustrates a short segment of the time domain output of the Type 2 amplifier where the differential output voltages reveals identical 1-pulse and 0-pulse wave shapes even though the single power supply voltage waveform that is applied to the H-bridge is asymmetric.

5.2 Type 3 DSD power amplifier output stage

Although the Type 2 amplifier topology addresses the problem of switching loss and to some extent alleviates

the problem of EMC, it has the disadvantage in that the signal amplitude to peak DSD output voltage ratio is relatively poor compared to a PWM amplifier. This is exacerbated by the fact that even when a sequence of all-1 or all-0 pulses in generated the differential output signal always returns to zero between samples as shown in Figure 15b thus lowering the short-term average of the waveform.

To overcome this deficiency a further modification to the amplifier is made and is shown conceptually in Figure 17. A similar raised-cosine power supply voltage is used but an additional constant amplitude voltage source is introduced with its amplitude set precisely to that of the peak value of the raised-cosine waveform. As with the Type 1 and 2 amplifiers these waveforms are controlled by the parameter *gain* as the PAM method is retained for gain control. Also in addition, is a switch that can select either the constant voltage supply or the dynamic power supply; the positions of this switch are defined in Figure 17 as *position 1* and *position 2* respectively. Figure 17 includes an illustration to show the relative levels of the power supplies output waveforms with respect to control parameter *gain*.

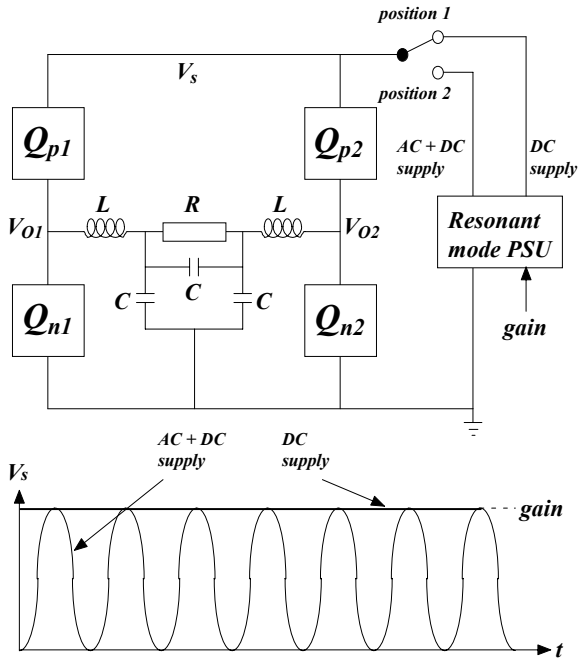


Figure 17 Type 3 output stage powered by variable-output, resonant-mode power supply with AC and DC supply commutation.

The operation of the Type 3 amplifier is as follows: Normally when a change from 1 to 0 or 0 to 1 occurs then switch *position 2* is selected and the amplifier operates as a Type 2 amplifier. However, if a burst of all-1 or all-0 pulses occurs in the DSD data stream then the switch changes to position 1 and the amplifier reverts effectively to a Type 1 configuration. This operation then forces the differential output signal of the H-bridge to remain constant throughout the period of the burst where an example output sequence is shown in Figure 18.

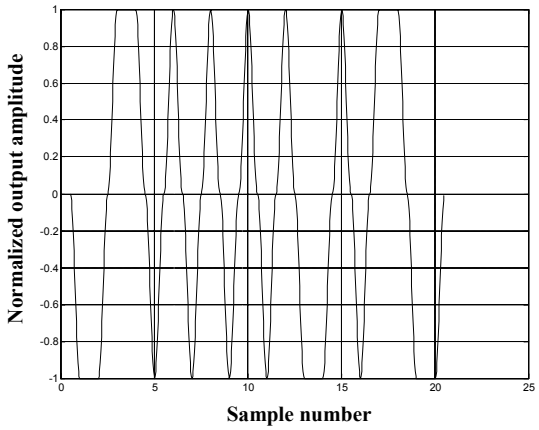


Figure 18 Type 3 time-domain output.

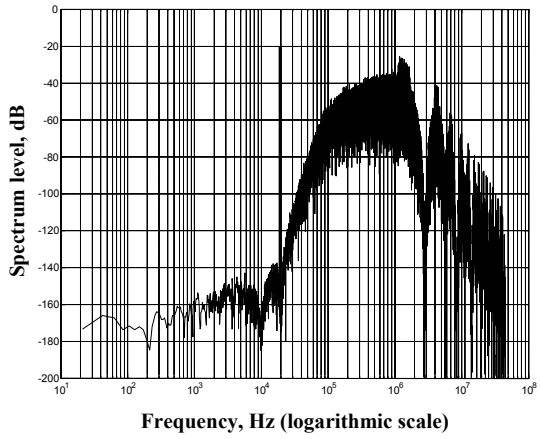


Figure 19 Type 3 frequency-domain output.

The effect of this process, illustrated in Figure 18, is to retain constant amplitude output pulses during a burst of liked-valued DSD data but where a data transition occurs then the instead of a rectangular output the

waveform follows a rounded form determined by the raised-cosine power supply. Consequently, Type 3 is a hybrid of Type 1 and 2 amplifiers. Finally, to confirm that the approach taken in the Type 3 amplifier does not introduce additional in-band distortion, a simulation was performed where the output spectrum is shown in Figure 19.

6. CONCLUSIONS

This paper has explored possible solutions to the design of a digital DSD power amplifier. A key objective was to remain faithful to the DSD source code and not to introduce signal processing in order to modify the data format to say LPCM or PWM. As such the topology is relatively simple although significant development was given to internal buffering in order to eliminate as much circuitry as possible in order to retain a fast response with low jitter and low power dissipation especially in the intermediate stages. Techniques that do not require circuitry with infinite switching speeds to give optimum performance are clearly preferable and in this sense can be compared against earlier work describing digital-to-analog conversion [8].

The intermediate buffer stage employed direct coupling to the N-channel MOSFETS but AC-coupling to the P-channel MOSFETS. This expedient enabled significant simplification compared to other forms of high speed, transistor intensive, level shifting circuitry. As the output stage performs open loop then the design of power supply is critical as is the fast switching and low on-resistance of the H-bridge. However, to enable gain control, modulation of the power supply voltage was employed where this together with a self-clamping AC-coupling circuit allowed the gain to be reduced virtually to zero while retaining proper switching of the MOSFETS over the full control range. A practical circuit was presented to demonstrate overall viability although it is recognized that significant improvements are required where proper circuit layout and the use of RF MOSFETS should allow substantial improvement especially in terms of distortion. Nevertheless, the prototype functioned and returned useful performance data especially with respect to the mode of gain control.

To progress the design process two variants on the prototype amplifier (Type 1) were presented and their conceptual viability explored through simulation. A novel feature for all these DSD amplifiers was to incorporate a dynamic power supply to modulate the H-bridge supply voltage. This expedient was shown to reduce RF interference thus improving EMC performance and significantly, to switch the H-bridge only when the drain-to-source voltages of all MOSFETS were zero. A further variant (Type 3) introduced an effective hold function so as to improve the signal amplitude to peak output voltage ratio. This requires a more sophisticated control system although the cost penalty is not considered severe.

7. ACKNOWLEDGEMENTS

This work was undertaken at University of Essex with technical support from Sony Professional Oxford, dCS, Jensen Capacitors and KEF Audio. We thank all parties for their help in this project.

8 REFERENCES

[1] Verbakel, J. Kerkhof, L.V.D. Maeda, M. Inazawa, Y., Super Audio CD Format, *104th AES Convention, Amsterdam,* 1998, preprint 4705

[2] International Rectifier 'Class-D Audio Amplifier Design', http://www.irf.com

[3] Nielsen, K., High-Fidelity PWM-Based Amplifier Concept for Active Loudspeaker Systems with Very Low Energy Consumption, *JAES,* vol. 45, no. 7/8, pp 554-570, July/August 1997

[4] Macgrath, A.J., and Sandler, M.B., Digital Power Amplification Using Sigma-Delta Modulation and Bit Flipping, *JAES*, vol. 45, no. 6, pp 476-487, 1997

[5] Poon, N.K. and Lau, W.H., A Novel Crossover Delay PWM Amplifier, *98th AES Convention, Paris*, February 1995, preprint no. 3969

[6] Takahashi, H. and Nishio, A., Investigation of Practical 1-bit Delta-Sigma Conversion for Professional Audio Applications, *110th AES Convention*, Amsterdam, May 2001, paper 5392

[7] dCS, dCS972 digital-to-digital converter, http://www.dcsltd.co.uk

[8] Hawksford, M.J., Digital-to-analog converter with low intersample transition distortion and low sensitivity to sample jitter and transresistance amplifier slew rate, *JAES*, vol. 42, no. 11, pp 901-917, 1994

Modulation and System Techniques in PWM and SDM Switching Amplifiers*

MALCOLM HAWKSFORD, *AES Fellow*
(mjh@essex.ac.uk)

University of Essex, Department of Electronic Systems Engineering, Colchester, Essex, CO4 3SQ, UK

Continuous- and discrete-time switching audio power amplifiers are studied both with and without feedback. Pulse-width modulation (PWM) and sigma–delta modulation (SDM) amplifier configurations are simulated and their interrelationship is described using linear phase modulation (LPM) and linear frequency modulation (LFM). Distortion generation encountered when applying negative feedback to PWM is demonstrated and strategies to improve linearity are presented. Recent innovations in SDM coding and output-stage topologies using pulse-shaping techniques are reviewed with emphasis on stable, low-distortion operation, especially under high-level signal excitation. A simplified low-latency variant of predictive SDM with step back is introduced, which together with dynamic compression of the state variables extends stable operation to a modulation depth of unity, thus allowing SDM to compete with PWM power amplifiers in terms of peak signal capability.

0 INTRODUCTION

There is a growing awareness of the role of high-efficiency topologies in power amplifier applications and especially the strategic significance in their application to a wide range of audio products [1], [2]. This is already evident in the digital theater market, where smaller size yet high performance products continue to emerge. There is also the opportunity to lower overall power dissipation, which when mapped into volume production is a most critical issue in terms of environmental factors. Key advantages stem from reduced size and heat loss. However, there are also more fundamental philosophical reasons for adopting class-D switching in the amplification process. (Hereafter switching amplifier infers "class-D" switching amplifier.) Switching amplifiers configured specifically for use with digital signals take on the mantle of a power digital-to-analog converter. As such there is reduced analog processing as the digital signals are, in a figurative sense, brought into closer proximity with the loudspeaker. Consequently there is opportunity to maintain better signal integrity and to achieve a more transparent overall performance, commensurate with appropriate design and physical implementation. The reduction in analog-related artifacts such as dynamic modulation of the closed-loop transfer function through device nonlinearity, including active device transconductance and internal capacitance modulation, implies less amplifier-dependent signal coloration and should lead to a more neutral and consistent sound quality. However, switching output stages offer potentially better control of a loudspeaker, as the source impedance of the amplifier remains low and almost resistive even under overload. This is contrary to most analog amplifiers, which are highly dependent on negative feedback to lower distortion and output impedance. The output impedance of a switching amplifier is determined principally by the actual on resistance of the output switching devices, moderated only by negative feedback, when used, and by the passive low-pass filter necessary to limit the extreme high-frequency signal components resulting as a consequence of the type of modulation scheme selected. Also depending upon both topology and whether or not feedback is used, the characteristics of the power supply can be critical.

Until recently most switching power amplifiers for audio were based only on a paradigm of PWM, which offers a number of attractive features. Classical PWM generates a binary output signal "switched" between two voltage levels, where the width of each pulse is modulated in proportion to the input signal such that the short-term average of the square-wave output follows the instantaneous amplitude of the input signal. Since the output power devices switch rapidly between two states, they only dissi-

*Presented at the 118th Convention of the *Audio Engineering Society,* Barcelona, Spain, 2005 May 28–31, under the title "SDM versus PWM Power Digital-to-Analog Converters (PDACs) in High-Resolution Digital Audio Applications." Manuscript received 2005 May 11; revised 2005 November 11 and 2006 January 11.

pate significant power in the pulse transition regions, enabling the output stage to achieve high efficiency. However, with the application of SDM used in both analog-to-digital converters (ADC) and digtal-to-analog converters (DAC) and also in Super Audio CD (SACD) [3], SDM has become a viable alternative binary modulation method for switching power amplifier systems. Unlike PWM, where the pulse width is modulated, SDM forms a sequence of equal-area binary pulses such that the relative densities of positive and negative pulses are complementary and track the instantaneous amplitude of the input signal. Traditionally SDM forms a digital signal where the pulse instants are assigned to discrete time slots. But just as PWM has both continuous-time (analog) and discrete-time (digital) variants, so can SDM. However, where the source signal is digital, it is prudent to maintain signals within the digital domain and not to impose additional cascaded stages of ADCs and DACs. Also because uniformly sampled PWM associated with digital source signals is inherently nonlinear together with the generation of requantization noise due to the output pulses being constrained in time, additional signal processing is required both to linearize the modulator and to shape the requantization noise spectrally. Nevertheless, even for a digital switching amplifier, when the output stage and the power-supply rejection requirements are considered together with the role of negative feedback, the principal design factors revert back to those of an analog system. Consequently analog feedback techniques can apply either to analog or to digitally (see Section 6) derived PWM.

Digital PWM requires two clocks that ultimately bound its performance. First there is the sampling frequency that determines the repetition rate of pulse transitions, and second there is the much higher frequency clock, which defines the discrete-time locations of each pulse transition. However, with the advent of the direct-stream digital[1] (DSD) format based on SDM [3], the effective sampling rate of the system has been elevated to 2.8224 MHz where, unlike in PWM, the sampling rate and the pulse repetition rates are the same. Also SDM can readily be implemented so that within the audio band the modulator is virtually linear and does not require an additional linearization processor to achieve acceptable levels of distortion. The increase in sampling rate also facilitates a simpler low-pass filter with the potential for reduced signal losses because such output filters have to handle the full output current of the amplifier. However, offsetting this advantage, the high number of pulse transitions per second imply potentially lower power efficiency, thus demanding the use of high-speed switching transistors and possibly zero-voltage switching, as discussed in Section 7.

This paper commences in Section 1 with a study of both PWM and SDM directed at switching amplifiers with an emphasis on modulation and the exploitation of negative feedback. The approach taken considers the relationship between PWM and SDM using modulation models based on linear phase modulation (LPM) and linear frequency modulation (LFM). Analytical modeling establishes the native linearity of each modulator class and enables the results to benchmark the performance of amplifier variants that incorporate combinations of negative feedback and linearization strategies. It is desirable to use negative feedback in a switching amplifier in order to reduce distortion dependence on both output-stage nonidealities and power-supply variations. However, it is shown in Section 3 that for PWM the introduction of output-voltage-derived feedback, although achieving anticipated performance gains, can introduce additional distortion products that are not observed in optimized open-loop PWM. Consequently corrective means are required to reduce these undesirable artifacts where a number of amplifier topologies are presented. This aspect of the study exploits a high precision Matlab[2] simulator of a PWM negative-feedback amplifier, where a spectral resolution in excess of 200 dB reveals all significant distortion products. The simulations show a noise-shaping advantage for induced jitter, the generation of intermodulation distortion for a multitone excitation, and the resulting effects on both noise and distortion when the loop gain is changed.

The study concludes with Section 7 by describing an SDM power amplifier topology [4] designed both to attenuate switching components above the SDM pulse repetition rate and to lower switching losses by forcing the output transistors to commutate only when switching voltages are zero. To complement this amplifier a digital SDM encoder algorithm is presented, which combines predictive look-ahead with dynamic compression of state variables that together enable stable coding up to a modulation index of unity, an important factor in achieving the full output signal capability of an SDM-based switching amplifier.

1 SDM–PWM ANALYTICAL COMPARISON

It is well known that naturally sampled, non-time-domain quantized PWM implemented using a symmetrical sawtooth waveform and a binary comparator offers low intrinsic distortion provided the bandwidth of the input signal is suitably constrained to prevent reflected components about the sampling frequency from falling within the audio band [5]. On the first encounter it may appear unusual for a signal that is processed by a two-level amplitude quantizer to have low intrinsic distortion. However, in this section it is shown that the modulation process has a fundamental affinity with LPM and is therefore intimately related to the LFM model proposed for SDM [6]–[9]. Consequently to establish a proper mathematical framework, models for both SDM and PWM are developed and their interrelationship is established.

1.1 SDM Modeling

Fig. 1 shows an SDM model using two differentially driven (or complementary) LFMs to produce a non-time-domain quantized pulse train of density-modulated posi-

[1]DSD was proposed by Sony to describe sigma–delta modulation for high-resolution audio.

[2]Matlab is a trade name of The MathWorks, Inc.

tive and negative pulses. The principal features of this structure have been reported in previous papers [6]–[7], [9]. LFM is a technique where a sinusoidal carrier is modulated such that its instantaneous carrier frequency is proportional to the amplitude of the input signal. In the SDM model based upon LFM, common reference points are designated on each cycle of the carrier where in this study the positive-slope zero crossings (PSZCs) are chosen for their convenience of definition. Constant-area (such as Dirac) impulses are then located at each PSZC to form the output pulse sequence for non-time-domain quantized SDM. However, because SDM generates two time-interleaved sequences of positive and negative pulses, as illustrated in Fig. 1, then carriers $s_1(t)$ and $s_2(t)$ with complementary LFM are used to form pulse density modulated sequences $P_1(t_{1r})$ and $P_2(t_{2r})$, where $\{t_{1r}, t_{2r}\}$ are the respective pulse time coordinates. Both carrier center frequencies (that is, the frequency when the input signal is zero) are set to f_{lfm}. Thus because of modulation symmetry, the combined pulse repetition rate of positive and negative output pulses always remains constant at $2f_{lfm}$ pulses per second, where the interleaved SDM output stream is $P_1(t_{1r}) - P_2(t_{2r})$. Observing this process it is evident that, say for a positive input signal, the frequency of positive pulses increases while simultaneously the frequency of negative pulses falls by an equal amount. Hence the short-term average of the composite output tracks the input signal where, if time-domain quantization is ignored, this process shows similar behavior to SDM. In the following analysis each LFM is realized as a cascade of signal integration and LPM that yields low intrinsic distortion [7], [8] provided the output pulse time coordinates are unconstrained so as to adopt their natural sampling instants. This modulation process is summarized in the following to facilitate comparisons between SDM and PWM.

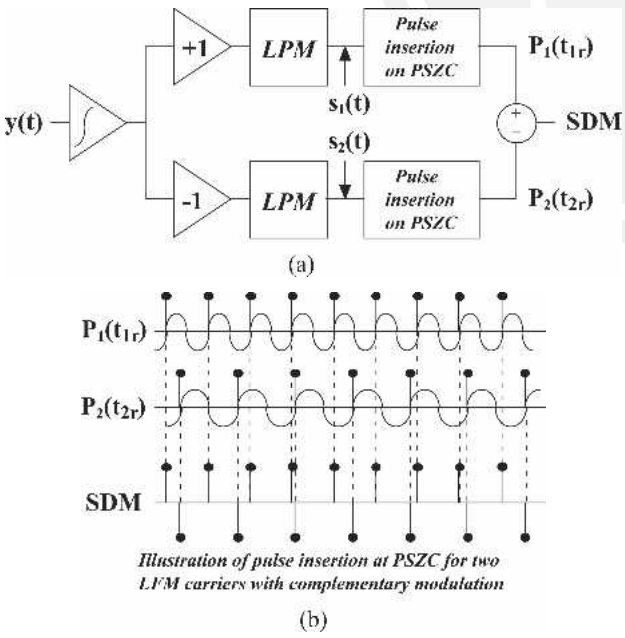


Fig. 1. (a) Non-time-quantized SDM using LFM. (b) Pulse insertion at PSZC for two LFM carriers with complementary modulation.

Assume the input signal $y(t)$, where t is time, is normalized by $\hat{Y}$ to limit the respective instantaneous frequencies $f_1(t)$ and $f_2(t)$ of the two complementary LFM carriers to a range of 0 Hz $< f_{1,2}(t) < 2f_{lfm}$ Hz, which are defined in terms of $y(t)$, $\hat{Y}$, and f_{lfm} as

$$f_1(t) = f_{lfm}\left[1 + \frac{y(t)}{\hat{Y}}\right] \quad \text{and} \quad f_2(t) = f_{lfm}\left[1 - \frac{y(t)}{\hat{Y}}\right]. \tag{1}$$

Instantaneous LFM carrier phases $\theta_1(t)$, $\theta_2(t)$ expressed as functions of $f_1(t)$, $f_2(t)$ are

$$\theta_1(t) = 2\pi \int_{u=0}^{t} f_1(u)\, du \quad \text{and}$$

$$\theta_2(t) = 2\pi \int_{u=0}^{t} f_2(u)\, du. \tag{2}$$

Hence the two carriers $s_1(t)$ and $s_2(t)$ with complementary LFM become

$$s_1(t) = A\cos[\theta_1(t)] = A\cos\left[2\pi f_{lfm}t + \frac{\pi}{2} + \frac{2\pi f_{lfm}}{\hat{Y}}\int_{u=0}^{t} y(u)\, du\right] \tag{3}$$

$$s_2(t) = A\cos[\theta_2(t)] + A\cos\left[2\pi f_{lfm}t - \frac{\pi}{2} - \frac{2\pi f_{lfm}}{\hat{Y}}\int_{u=0}^{t} y(u)\, du\right]. \tag{4}$$

Note in Eqs. (3) and (4) that complementary phase shifts of 0.5π and -0.5π are included, so that under quiescent conditions when $y(t) = 0$, positive and negative output pulses idle with an alternating symmetry.

To derive non-time-domain quantized SDM using the LFM model, Fig. 1 shows positive pulses located at the PSZC of $s_1(t)$ (that is, at times t_{1r}) and negative pulses located at the PSZC of $s_2(t)$ (that is, times t_{2r}), where solutions for t_{xr} occur when $\theta(t_{xr}) = 2\pi r$ for integer r, that is,

$$t_{1r} + \frac{1}{\hat{Y}}\int_{t=0}^{t_r} y(t)\, dt = \frac{r - 0.25}{f_{lfm}} \tag{5}$$

$$t_{2r} - \frac{1}{\hat{Y}}\int_{t=0}^{t_r} y(t)\, dt = \frac{r + 0.25}{f_{lfm}}. \tag{6}$$

To solve Eqs. (5) and (6) in terms of t_{xr} a four-stage methodology is adopted [9].

1) LFM signals $s_1(t)$, $s_2(t)$ are sampled at uniformly spaced time instants at the rate of $\cdot(f_{lfm})$ Hz, where the oversampling factor of is greater than 1, in order to give a finer time resolution (for example, of = 16).

2) Sampled carriers are scanned to identify negative to positive polarity transitions to identify all PSZCs.

3) For each detected PSZC, linear interpolation between adjacent samples improves on $\{t_{1r}, t_{2r}\}$ estimation.

4) An iterative error-driven procedure then seeks optimum solutions for $\{t_{1r}, t_{2r}\}$.

The SDM output is formed by subtracting the two complementary LFM-derived pulse streams. The corre-

sponding spectrum $\text{out}_{\text{sdm}}(f)$ is then calculated [9] by a summation over N, with $r = N$ being the Nth rotation of 2π in $\theta(t_{xr}) = 2\pi r$, as

$$\text{out}_{\text{sdm}}(f) = \sum_{r=1}^{N} (e^{-j2\pi f t_{1r}} - e^{-j2\pi f t_{2r}}). \quad (7)$$

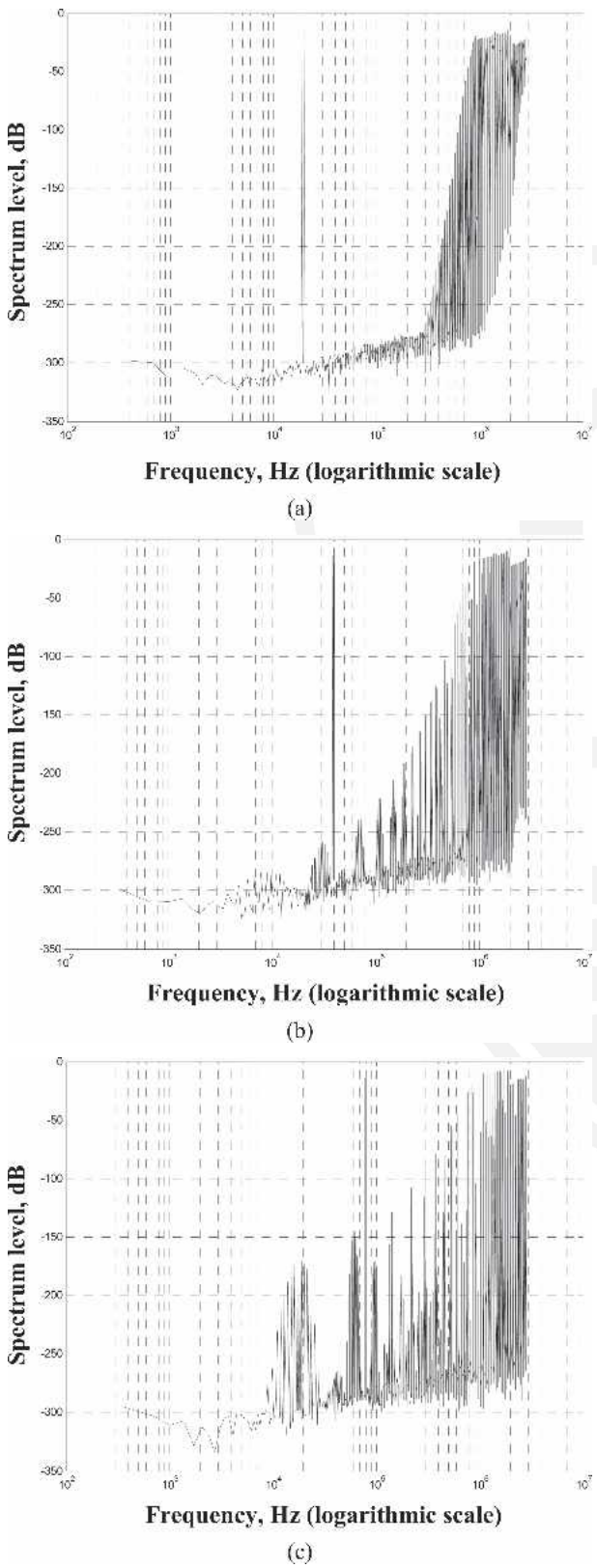


Fig. 2. SDM output spectra. (a) Example 1. (b) Example 2. (c) Example 3.

By way of illustration, Fig. 2 shows three examples of SDM output spectra using two-tone input signals with the following data. All SDM examples have $f_{\text{sdm}} = 64(44.1)$ kHz.

Example 1:
[A1 = 0.2, f1 = 19 kHz]
[A2 = 0.2, f2 = 20 kHz]
Example 2:
[A1 = 0.2, f1 = 39 kHz]
[A2 = 0.2, f2 = 40 kHz]
Example 3:
[A1 = 0.2, f1 = 79 kHz]
[A2 = 0.2, f2 = 80 kHz].

All computations show extremely low in-band distortion, which confirms a high degree of linearity. Distortion only becomes problematic when both input signal frequency and modulation depth are sufficiently high for the sidebands centered about the carrier frequency to migrate toward the audio band.

1.2 PWM Modeling

Having examined SDM, a similar model is now constructed for non-time-domain quantized PWM. Fig. 3 shows open-loop, naturally sampled PWM based on binary amplitude quantization with a symmetrical triangular wave added to the input signal. On first encounter it ap-

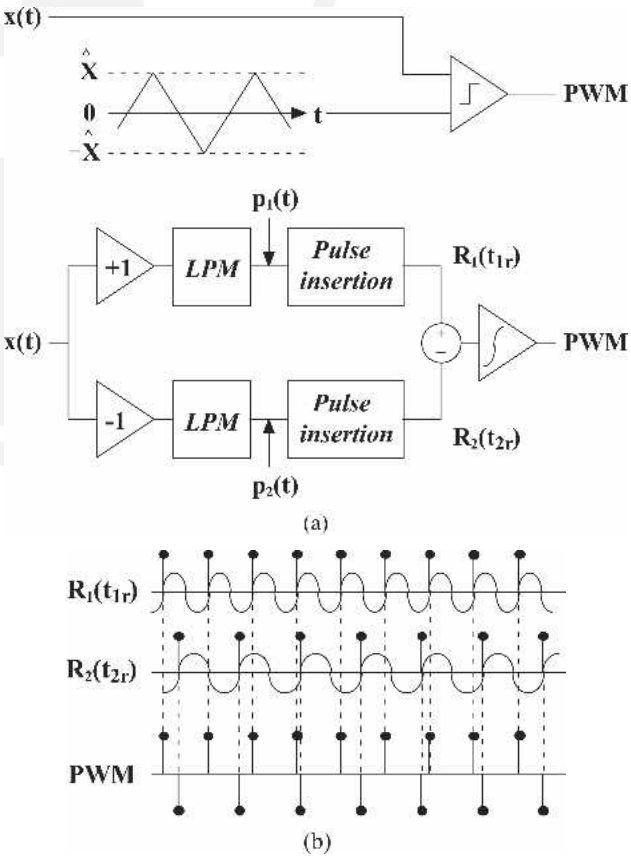


Fig. 3. (a) Non-time-quantized PWM using LPM. (b) Pulse insertion at PSZC for two LPM carriers with complementary modulation.

pears unlikely that such a simple structure can pass a signal with low intrinsic distortion. Following earlier work on uniformly sampled PWM [10], [11], it could be argued that as the input signal modulates the pulse width then, because the rth rectangular pulse of width τ_r transforms to the frequency domain with magnitude response $\tau_r \sin(\pi f \tau_r)/(\pi f \tau_r)$, the resulting nonlinear frequency response as a function of pulse width produces distortion. Therefore PWM appears to exhibit an intrinsic dynamic spectral modulation although, as will be shown, this does not occur with natural sampling.

The key to understanding naturally sampled PWM is the observation that each output pulse transition, both −1 to +1 and +1 to −1, undergoes linear modulation in time (provided the input amplitude does not exceed the peak-to-peak amplitude of the triangular wave) as a function of the instantaneous amplitude of the input signal. This differs from the SDM process described in Section 1.1 as there it was the frequency of the output pulses that was proportional to the input signal, as shown by Eqs. (3) and (4), which describe classic LFM. However, in PWM the process is closer to LPM, although the relationship between LFM and LPM is only the inclusion of linear integration of the input signal.

The proposed model for naturally sampled PWM, illustrated in Fig. 3, is described as follows. It is recognized that in naturally sampled PWM the positive and negative transitions of the PWM output are individually associated with the input signal and that there is no coupling between edges other than indirectly through filtering applied to the input signal. Hence in the non-time-domain quantized PWM model the pulse transitions can be determined individually by two independent linear-phase modulators. However, because for a given change in input signal amplitude, PWM edges move in opposite directions, these modulators are driven differentially by the input signal, mirroring the complementary LFM process used for modeling SDM. Also, because for a zero input signal the two output pulse sequences must be offset by a half-cycle to form a symmetrical interlaced sequence, the complementary dc signals are added to the phase modulator inputs to shift the respective phases of the carriers by $\pi/2$ and $-\pi/2$, as shown in Eqs. (8) and (9). Hence if the PWM carrier frequency is f_{pwm} Hz and the input signal $x(t)$ is normalized by $\hat{X}$, then the two phase modulator signals $p_1(t)$ and $p_2(t)$ are given by

$$p_1(t) = A \cos\left[2\pi f_{\text{pwm}} t + \frac{\pi}{2}\left(1 + \frac{x(t)}{\hat{X}} \right) \right] \tag{8}$$

$$p_2(t) = A \cos\left[2\pi f_{\text{pwm}} t - \frac{\pi}{2}\left(1 + \frac{x(t)}{\hat{X}} \right) \right]. \tag{9}$$

As with the LFM-SDM model, the PSZC time coordinates $\{t_{1\text{r}}, t_{2\text{r}}\}$ form natural time sampling instants derived from the two LPM carriers $p_1(t)$, $p_2(t)$. Unit-area pulses (such as Dirac) $R_1(t_{1\text{r}})$, $R_2(t_{2\text{r}})$ are then located at $\{t_{1\text{r}}, t_{2\text{r}}\}$, whereby $R_1(t_{1\text{r}}) - R_2(t_{2\text{r}})$ forms the composite data stream of positive and negative pulses. Finally the composite stream is integrated, which translates each Dirac pulse into a unit step function. However, because of the interleaved nature of the pulse sequence and since on average the number of positive and negative pulses must remain the same, a binary PWM square-wave output is formed, provided $|x(t)| < \hat{X}$. If this limit is exceeded, then the output pulse sequence becomes multilevel as it is possible to have two or more sequential pulses of the same sign. This compares directly with replacing the two-level quantizer normally used in PWM with a uniform multilevel quantizer. It should be noted that in order to achieve an amplitude-symmetric bipolar sequence typical of a practical PWM amplifier, a constant dc offset is subtracted from the integrated output equal to one-half the weight of the first Dirac pulse in the composite pulse sequence, thus achieving a long-term average of zero. Also, because of the relationship between phase and frequency, as stated in Eq. (2), it follows that LPM acts as a differentiator and therefore requires integration to correct the overall frequency response. In the LFM-SDM model the integrator precedes LPM, whereas for the LPM-PWM model it is located after LPM. The location of integration gives insight into modulator linearity as conventional wisdom describes PWM akin to a sample-and-hold function, where the hold period is modulated by the input signal, which when mapped into the frequency domain suggests a mechanism for dynamic spectral modulation. However, in the PWM model it can now be seen that the output square wave is actually formed by integration of two interleaved phase-modulated pulse sequences. As such there is no finite-duration hold function being used; it is an illusion. This is a critical observation, which reveals that system linearity depends solely on the characteristics of the two differentially driven LPMs, where any distortion is just frequency shaped by an integrator.

A system simulation was performed where for both LPM carriers every PSZC instant $\{t_{1\text{r}}, t_{2\text{r}}\}$ was determined using in iterative procedure similar to that used for the SDM model. The output spectrum of the time-integrated output pulse sequence was then calculated using a complex exponential method similar to that described by Eq. (7),

$$\text{out}_{\text{pwm}}(f) = \sum_{r=1}^{N} \left(\frac{e^{-j2\pi f t_{1\text{r}}} - e^{-j2\pi f t_{2\text{r}}}}{j2\pi f T} \right) \tag{10}$$

where the integration time constant $T = 1$ second such that a Dirac pulse maps to a unit step. A principal advantage of this approach is that multiple-sine-wave input signals can be generated so that the linearity of the overall process can be explored in detail in a way that can be problematic using pure analytical techniques. Also, the comparison between SDM and PWM using non-time-domain quantized models reveals that the principal differences are the location of the signal integration function and that SDM normally operates with a higher sampling rate. The higher sampling rate for SDM adopted in digital applications is a result of differing quantization strategies, as when the PWM time coordinates $\{t_{1\text{r}}, t_{2\text{r}}\}$ are quantized

in time, a finer quantization interval must be used compared to that set by the PWM natural sampling rate, that is, it is the higher PWM quantization related clock rate that should be compared to the SDM pulse repetition rate. To show the formation of the PWM output waveform, Fig. 4 illustrates phase-modulated Dirac pulses derived from the PSZC of the complementary phase modulators both before and after linear integration and including the constant dc offset required to achieve an amplitude-symmetric pulse distribution.

Examples of output spectra derived using the LPM model are shown in Fig. 5 for three sets of input signals with progressively higher signal frequency. As with the SDM simulations, the input signals adopt the same two tones to facilitate comparison, although the center frequencies of LPM and LFM differ. The example spectra pre-

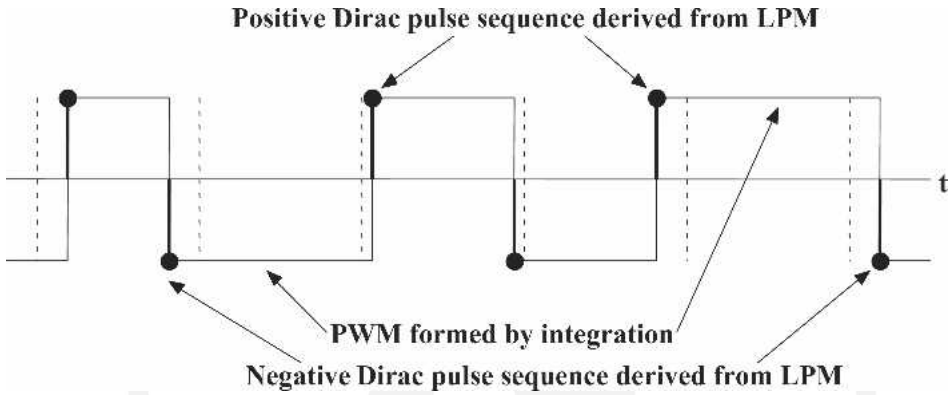


Fig. 4. PWM derived by integration of LPM.

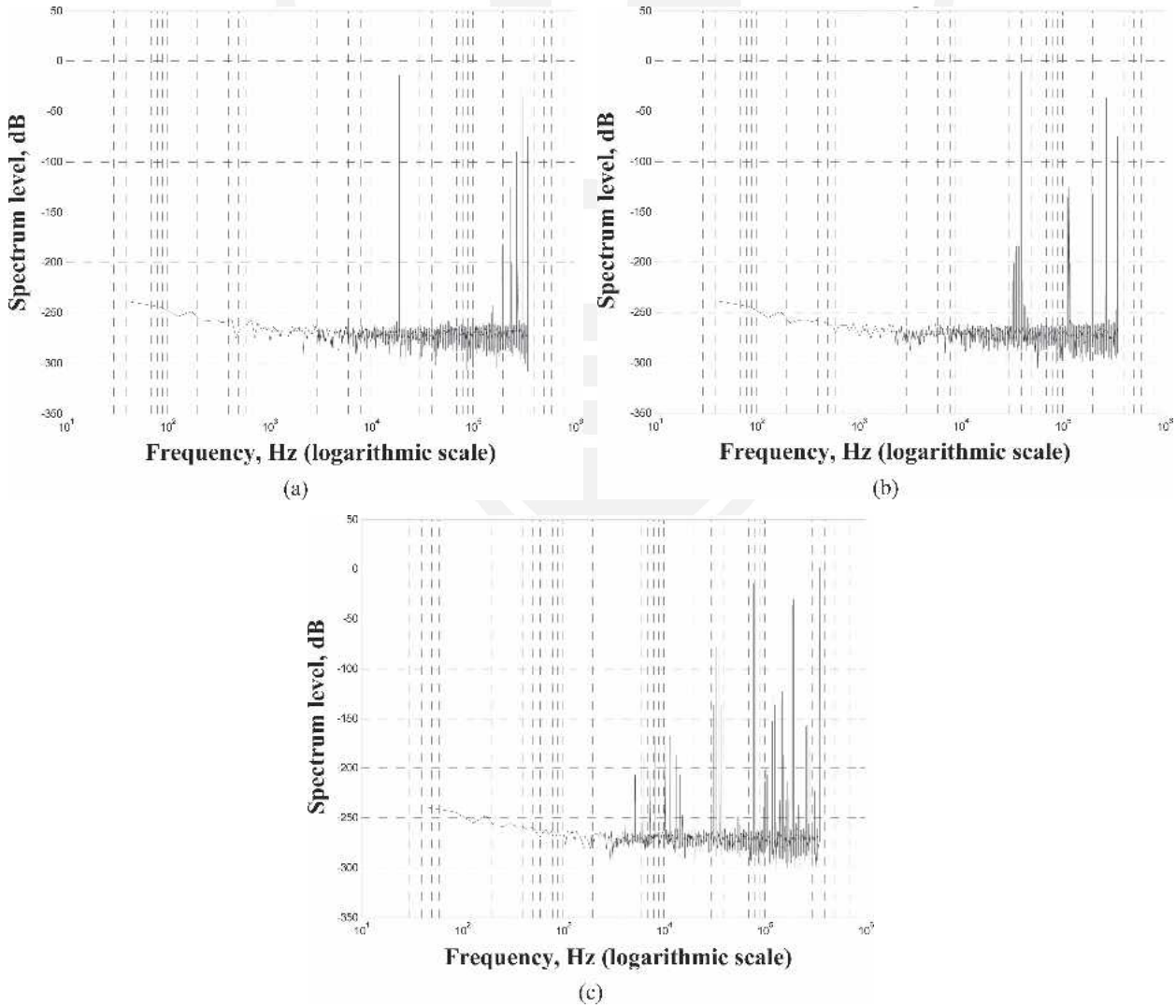


Fig. 5. PWM output spectra. (a) Example 1. (b) Example 2. (c) Example 3.

sented confirm the low inherent distortion achievable with LPM, although as the frequency of the input signal is increased, the migration of aliased frequency components toward the audio band can be observed. All PWM examples have $f_{\text{pwm}} = 8(44.1)$ kHz.

Example 1:
[A1 = 0.2, f1 = 19 kHz]
[A2 = 0.2, f2 = 20 kHz]
Example 2:
[A1 = 0.2, f1 = 39 kHz]
[A2 = 0.2, f2 = 40 kHz]
Example 3:
[A1 = 0.2, f1 = 79 kHz]
[A2 = 0.2, f2 = 80 kHz].

1.3 Analytical Derivation of Dirac Pulse Sequences

The method described for determining PSZC is well matched to the task of simulation. However, it is also possible to adapt analytically both LFM and LPM models such that short-duration pulses with appropriate polarity are formed at PSZC. Consider the uniform quantizer characteristic shown in Fig. 6 with quantum 2π, where the input function is $\phi(t)$ and the quantized output is $Q[\phi(t)]$. Adopting the procedure reported previously [9] but putting quantum $\delta \to 2\pi$, a series representation of the quantization process follows,

$$Q[\phi(t)]|_{\delta=2\pi} = \phi(t) + \frac{\delta}{\pi}\sum_{r=1}^{\infty}\left[\frac{1}{r}\sin\left(2\pi r\frac{\phi(t)}{\delta}\right)\right]$$
$$= \phi(t) + 2\sum_{r=1}^{\infty}\left[\frac{1}{r}\sin(r\phi(t))\right]. \quad (11)$$

Differentiating $Q[\phi(t)]$ with respect to time forms a time-domain series of pulses that are located at each quantizer transition,

$$\delta[\phi(t)]|_{R\to\infty} = \frac{d\phi(t)}{dt}\left\{1 + 2\sum_{r=1}^{R}[\cos(r\phi(t))]\right\}. \quad (12)$$

Fig. 6 illustrates the uniform quantizer expressed as a function of $\phi(t)$ whereas Fig. 7 shows by way of example the synthesis of pulses located at PSZC for 8, 32, and 256 harmonics and where $d\phi(t)/dt = 1$. The example for 256 harmonics reveals a narrow synthesized pulse width compared to the sequence period, although for the Dirac pulse to be formally represented $R \to \infty$. Hence applying Eqs.

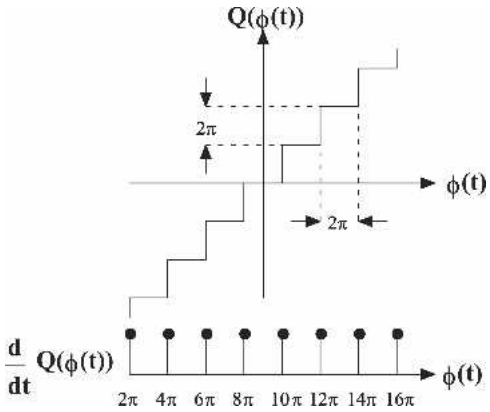


Fig. 6. Dirac pulse formation using differentiation of quantization characteristic.

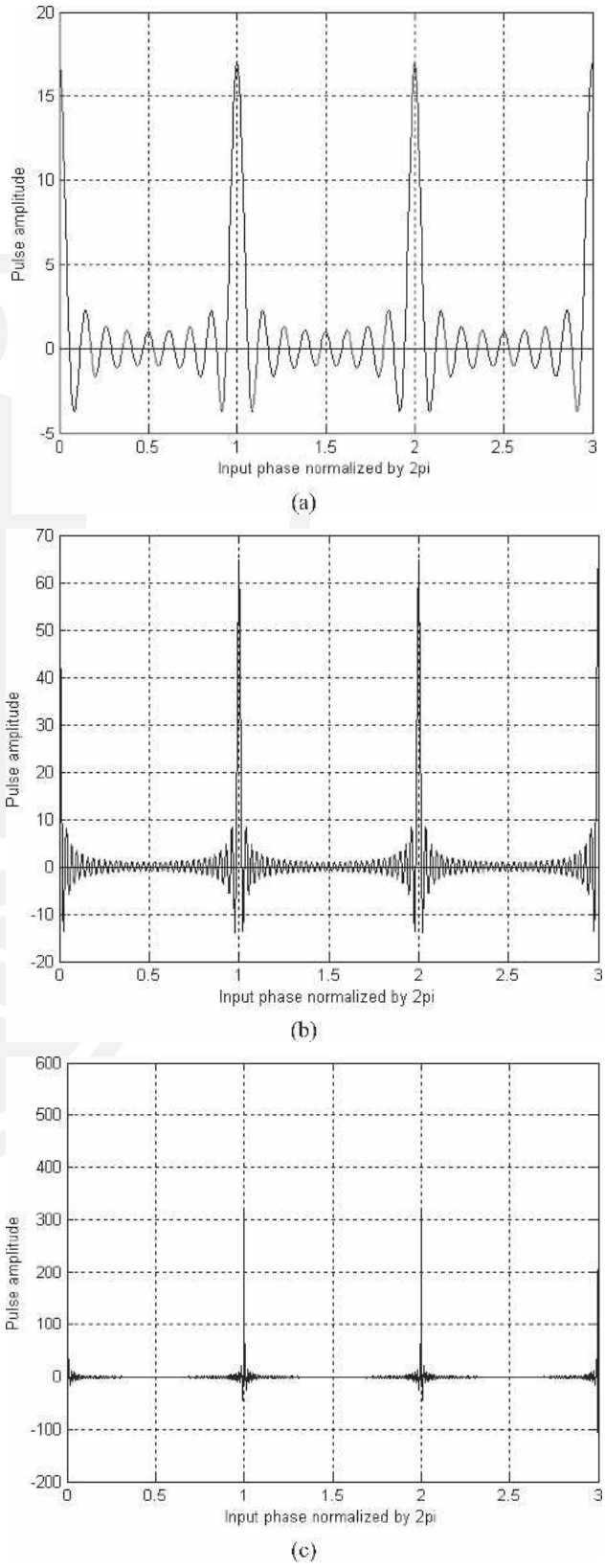


Fig. 7. Pulse synthesis. (a) 8 harmonics. (b) 32 harmonics. (c) 256 harmonics.

(3) and (4) to Eq. (12), the SDM pulse sequence δ_{sdm} follows,

$$\delta_{sdm} = \frac{y(t)}{\hat{Y}} + \left[1 + \frac{y(t)}{\hat{Y}}\right] \sum_{r=1}^{\infty} \left\{ \cos\left[r\left(2\pi f_{lfm} t + \frac{\pi}{2} + \frac{2\pi f_{lfm}}{\hat{Y}} \int_{u=0}^{t} y(u)\, du \right) \right] \right\} - \left[1 - \frac{y(t)}{\hat{Y}}\right] \sum_{r=1}^{\infty} \left\{ \cos\left[r\left(2\pi f_{lfm} t - \frac{\pi}{2} - \frac{2\pi f_{lfm}}{\hat{Y}} \int_{u=0}^{t} y(u)\, du \right) \right] \right\}. \quad (13)$$

Applying Eqs. (8) and (9) to Eq. (12) and integrating the output, the non-time-quantized PWM pulse sequence δ_{pwm} is given,

$$\delta_{pwm} = \frac{1}{\hat{X}} \frac{dx(t)}{dt} + 2\left[2f_{pwm} + \frac{1}{2\hat{X}} \frac{dx(t)}{dt} \right] \sum_{r=1}^{R} \left\{ \cos\left[r\left(2\pi f_{pwm} t + \frac{\pi}{2}\left(1 + \frac{x(t)}{\hat{X}} \right) \right) \right] \right\} - 2\left[2f_{pwm} - \frac{1}{2\hat{X}} \frac{dx(t)}{dt} \right] \sum_{r=1}^{R} \left\{ \cos\left[r\left(2\pi f_{pwm} t - \frac{\pi}{2}\left(1 + \frac{x(t)}{\hat{X}} \right) \right) \right] \right\}. \quad (14)$$

2 JITTER SENSITIVITY OF SDM AND PWM

From the respective non-time-quantized models of SDM and PWM described in Sections 1.1 and 1.2 it is straightforward to predict the relative sensitivity of each system to pulse jitter. In making this comparison the key difference can be derived from the position of the integrator, where it was shown that for SDM the input to the LPM is integrated, whereas for PWM the output pulse sequence is integrated. Jitter is represented in terms of output pulse time displacement with the conversion rule that the pulse area must remain invariant, a requirement normally met in SDM using switched-capacitor circuits [12].

Consider a general time-modulated impulse sequence of M samples $\sum_{r=1}^{M}\{\alpha_r \delta(t - t_r)\}$, where $\alpha_r = 1$ or -1, depending on the pulse polarity, derived from either the LFM or the LPM models and located at time t_r, displaced in time by instantaneous jitter Δt_r such that its actual time coordinate is $t_r + \Delta t_r$. The resulting instantaneous time-domain error sequence $\Gamma(t)$ is

$$\Gamma(t) = \sum_{r=1}^{M} \alpha_r \{\delta(t - t_r) - \delta(t - t_r - \Delta t_r)\}. \quad (15)$$

For SDM the corresponding spectral error Esdm(f) follows directly from Equation (15),

$$\mathrm{Esdm}(f) = \sum_{r=1}^{M} \alpha_r [e^{-j2\pi f t_r} - e^{-j2\pi f (t_r + \Delta t_r)}]$$

that is,

$$\mathrm{Esdm}(f) = j2 \sum_{r=1}^{M} [\alpha_r e^{-j2\pi f (t_r + 0.5\Delta t_r)} \sin(\pi f \Delta t_r)]. \quad (16)$$

Eq. (16) shows that in the lower frequency region the SDM spectral error is proportional to frequency since $\sin(\pi f \Delta t_r) \approx \pi f \Delta t_r$. However, for PWM the output pulses derived from LPM are spectrally weighed by an integrator with time constant T second, giving a PWM jitter spectrum Epwm(f) in terms of Esdm(f),

$$\mathrm{Epwm}(f) = \frac{\mathrm{Esdm}(f)}{j2\pi f T}. \quad (17)$$

Hence substituting for Esdm(f) from Eq. (16) and introducing a sinc function,

$$\mathrm{Epwm}(f) = \sum_{r=1}^{M} \left[\alpha_r e^{-j2\pi f (t_r + 0.5\Delta t_r)} \frac{\Delta t_r}{T} \mathrm{sinc}(\pi f \Delta t_r) \right]. \quad (18)$$

Eq. (18) reveals that because of integration, the PWM jitter spectrum is virtually constant with frequency and proportional to Δt_r since at low frequency $\mathrm{sinc}(\pi f \Delta t_r) \approx 1$. Exploiting Eqs. (16) and (18), two simulations were performed for SDM and PWM to reveal the output spectral error due only to jitter. The LPM center frequency for PWM was 8(44.1) kHz whereas the LFM center frequency for SDM was 64(44.1) kHz. Also both simulations used RPDF[3] jitter noise of 2 ns peak to peak. For both SDM and PWM the input signal consisted of 19- and 20-kHz sine waves, each with peak amplitude 0.1. This allowed both spectral shape and distribution about the respective carrier frequencies to be compared. The results are shown in Figs. 8 and 9. Overall the PWM results reveal greater sensitivity to jitter and also confirm the predicted spectral distributions. However, in making comparisons the different numbers of pulse transitions per unit time for SDM and PWM should be noted as the sampling rates were selected to reflect typical amplifier applications.

3 NEGATIVE-FEEDBACK-DEPENDENT DISTORTION IN PWM AMPLIFIERS

This section considers the problem of applying negative feedback in PWM power amplifiers to lower output-stage distortion and dependence on power-supply variations, both of which are major performance-limiting factors, especially in open-loop designs. Section 1 has confirmed the linearity of naturally sampled idealized PWM and has shown that an LPM model can generate equivalent non-time-quantized PWM signals. Here the only distortion fundamental to LPM occurs when the input signal has high-

[3]Rectangular probability distribution function.

frequency, high-amplitude content so nonlinear aliased components migrate downward from the sampling frequency, as illustrated in Fig. 5. The key observation is that the presence of high-frequency input signal components causes an increase in distortion. In a naturally sampled PWM amplifier that uses output-voltage-derived negative feedback, a critical performance factor is the influence on modulator linearity of the output switching signal residue following its subsequent filtering by the forward-path amplifier. The presence of these high-frequency switching components inevitably increases sideband generation in LPM and thus causes the distortion performance to degrade over what might otherwise be anticipated from the use of negative feedback. Fig. 10 shows a basic PWM modulator, which includes a negative feedback loop with feedback factor B and a forward-path amplifier with transfer function A. In practice an output reconstruction filter is required, although this is omitted here as the unfiltered output signal is to be analyzed. Also included in the loop is an analog output stage N or power switch, which is susceptible both to switching distortion and to power-supply voltage, shown here to generate an instantaneous error voltage V_ε.

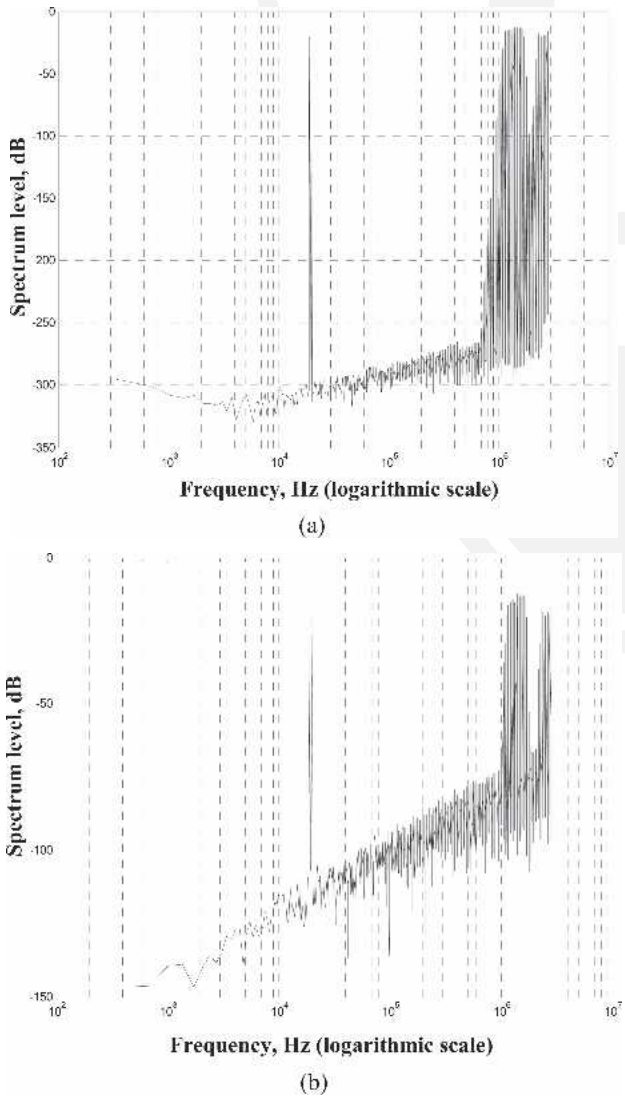


Fig. 8. SDM. (a) Without jitter. (b) With 2-ns peak-to-peak jitter.

In PWM the output voltage is a square wave and normally switches over almost the full range dictated by the power supply. It is therefore a high-amplitude signal containing high-frequency switching components. Following attenuation by B, the fed back signal is applied to the forward-path amplifier, where the transfer function A approximates

$$A = \frac{A_0}{1 + \mathrm{j}f/f_0} \Rightarrow \frac{A_0 f_0}{\mathrm{j}f} = \frac{f_T}{\mathrm{j}f} = \frac{1}{\mathrm{j}2\pi f T_A} \tag{19}$$

where A_0 is the dc gain, f_0 the 3-dB break frequency (dominant pole), $f_T = A_0 f_0$ the unity-gain bandwidth or the

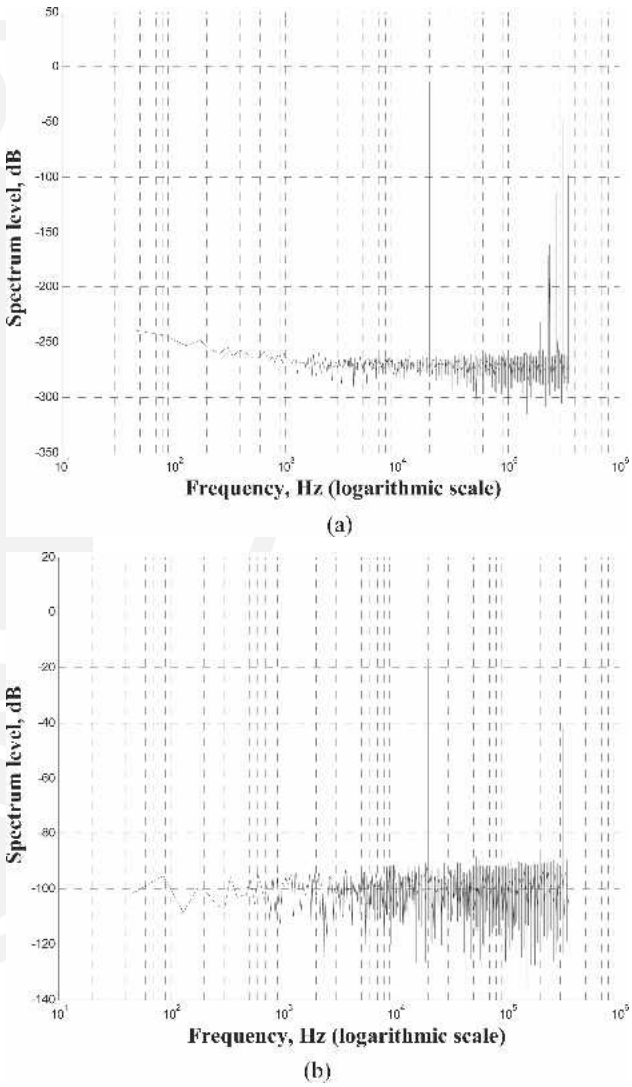


Fig. 9. Natural sampling PWM. (a) Without jitter. (b) With 2-ns peak-to-peak jitter.

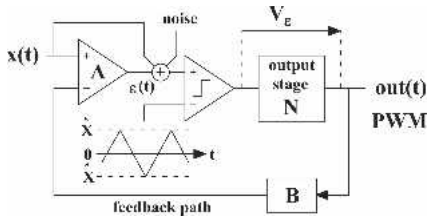


Fig. 10. Natural sampling analog PWM with feedback.

gain–bandwidth product, and T_A the corresponding time constant of the integrator (that is, amplifier A). T_A and f_T are then related as

$$f_T = \frac{1}{2\pi T_A}. \tag{20}$$

Because from Eq. (19) A tends to a first-order integrator when $A_0 \to \infty$, then $\varepsilon(t)$, the time-domain output of A, includes a triangular component produced by integration of the PWM square-wave output. Hence $\varepsilon(t)$ is a function of the PWM peak-to-peak output signal, feedback factor B, unity-gain frequency f_T, and PWM sampling rate f_{pwm}. Consequently there are high-frequency switching components present at the PWM modulator input, which can induce distortion by perturbing the natural PWM sampling instants.

In practice there are two principal factors that determine the maximum unity-gain frequency f_T of amplifier A. The first is related to the phase margin, which must include all phase shift within the amplifier loop, whereas the second, and more fundamental, is the triangular signal resulting from integration by amplifier A of the PWM square-wave output. To establish the maximum unity-gain frequency $\hat{f_T}$ for amplifier A in a simple negative-feedback PWM amplifier, assume in the forward path a unity-gain PWM stage with output power stage gain N, where both the triangular wave applied to the comparator and the PWM output square wave span −1 V to 1 V. Then allowing a 1-V margin for the signal headroom, the peak-to-peak range of $\varepsilon(t)$ is set to 1 V, that is,

$$\varepsilon(t) = \frac{NB}{T_A}\int_{t=0}^{0.5/f_{\text{pwm}}} \mathrm{d}t = 1$$

whereby the integrator time constant T_A is limited to

$$T_A \geq \frac{NB}{2f_{\text{pwm}}}. \tag{21}$$

Hence from Eqs. (20) and (21), and also confirmed by simulation to be a realistic estimate, the maximum unity-gain frequency $\hat{f_T}$ for amplifier A is

$$\hat{f_T} = \frac{f_{\text{pwm}}}{\pi NB}. \tag{22}$$

To gain further insight into PWM with negative feedback, observe that open-loop, naturally sampled PWM produces no discernable distortion in the absence of high-frequency input signals. It can therefore be said to produce an optimum PWM signal. Consequently if a closed-loop PWM amplifier were to correct completely for internal loop deficiencies then, in the absence of jitter, it should yield an output PWM waveform identical to that of the naturally sampled open-loop modulator. Any pulse relocation (other than pure delay) would represent degradation. However, because of the additional pulse-edge modulation caused by switching components present in $\varepsilon(t)$, even if the output stage N is perfect, then fundamentally a negative-feedback PWM amplifier as presented in Fig. 10, although improving on some performance aspects such as reduced power-supply sensitivity and output-stage dependence, could be anticipated to introduce distortion not present in open-loop PWM. Solutions to this problem therefore require signal processing targeted to eliminate (or significantly reduce) undesirable pulse-edge modulation.

Before presenting techniques to improve PWM closed-loop linearity, a precision simulation is presented to demonstrate distortion generation due to switching artifacts present within a PWM amplifier with overall negative feedback. A Matlab program was written based on the topology shown in Fig. 10 where, to minimize complexity, $NB = 1$. In the simulation amplifier A was modeled as an ideal z-domain integrator specified just in terms of its unity-gain frequency. Each PWM sample period was subdivided into 2^{14} increments (selected by experiment to reduce the simulation noise close to that set by finite-word-length artifacts in Matlab). For each computational increment an iterative procedure evaluated the state variables and estimated the time coordinate of each PWM transition. Time resolution was further enhanced by applying linear interpolation between computational increments. Since the topology included feedforward from the input to the PWM stage (see Fig. 10), changing the unity-gain frequency allowed the simulation to model amplifiers ranging from zero feedback, thus becoming an open-loop PWM amplifier, to maximum feedback where, From Eq. (22), $\hat{f_T} = f_{\text{pwm}}/\pi$. The simulation calculated the time coordinates $\{t_{1\text{r}}\}$ for the negative-to-positive transitions and $\{t_{2\text{r}}\}$ for the positive-to-negative transitions, where the output spectrum was evaluated using Eq. (10). To reveal noise shaping together with distortion generation as a function of feedback-loop unity-gain frequency, a noise source was added at the input to the comparator (see Fig. 10) to deliberately induce a small level of output jitter. In all other respects the amplifier had no other imperfections as the aim here was to expose only fundamental distortion mechanisms due entirely to the encapsulation of ideal, naturally sampled PWM with negative feedback.

All simulations used an input signal consisting of two sine waves, each of normalized amplitude 0.45, with respective frequencies of 17 and 20 kHz; the PWM sampling rate was set at $f_{\text{pwm}} = 16(44.1)$ kHz. The highest loop gain selected used $\hat{f_T} = f_{\text{pwm}}/\pi$, with subsequent gains reduced in increments of 20 dB up to a maximum of 60-dB attenuation. Also two simulations were performed with the amplifier gain set to zero (open-loop case), both with and without jitter noise, in order to benchmark the simulations with feedback and to observe the native resolution of the simulation. Computations were taken over 2^{14} PWM sample periods, with each period subdivided into a further 2^{14} increments. The open-loop spectral results are shown in Fig. 11 together with a histogram of the output jitter, indicating that noise induction produced a peak jitter of about 0.5 ps. The jitter level was deliberately kept low so as not to mask low-level distortion products. Spectral results for the case with feedback are shown in Fig. 12 for loop-gain attenuations of 0, −20, −40, and −60 dB. By way of comparison, Fig. 13 presents the results for uniform sampling PWM with loop gains of −20 and −60 dB.

The same simulator was used, but the a sample-and-hold was applied to the input signal with a sampling rate of $2f_{\text{pwm}}$ Hz.

The output spectra confirm that for zero-loop gain (that is, open-loop naturally sampled PWM) there is no evidence of in-audio-band intermodulation products. Also demonstrated is the extremely high resolution of the simulations with a spectral noise floor in excess of −250 dB. The inclusion of low-level jitter noise can also be seen where the histogram allows a direct observation of the jitter level that can be linked to the spectral noise level for calibration purposes. The level of time-domain jitter and the resulting spectral noise are in line with that discussed for PWM in Section 2. When feedback is applied, the

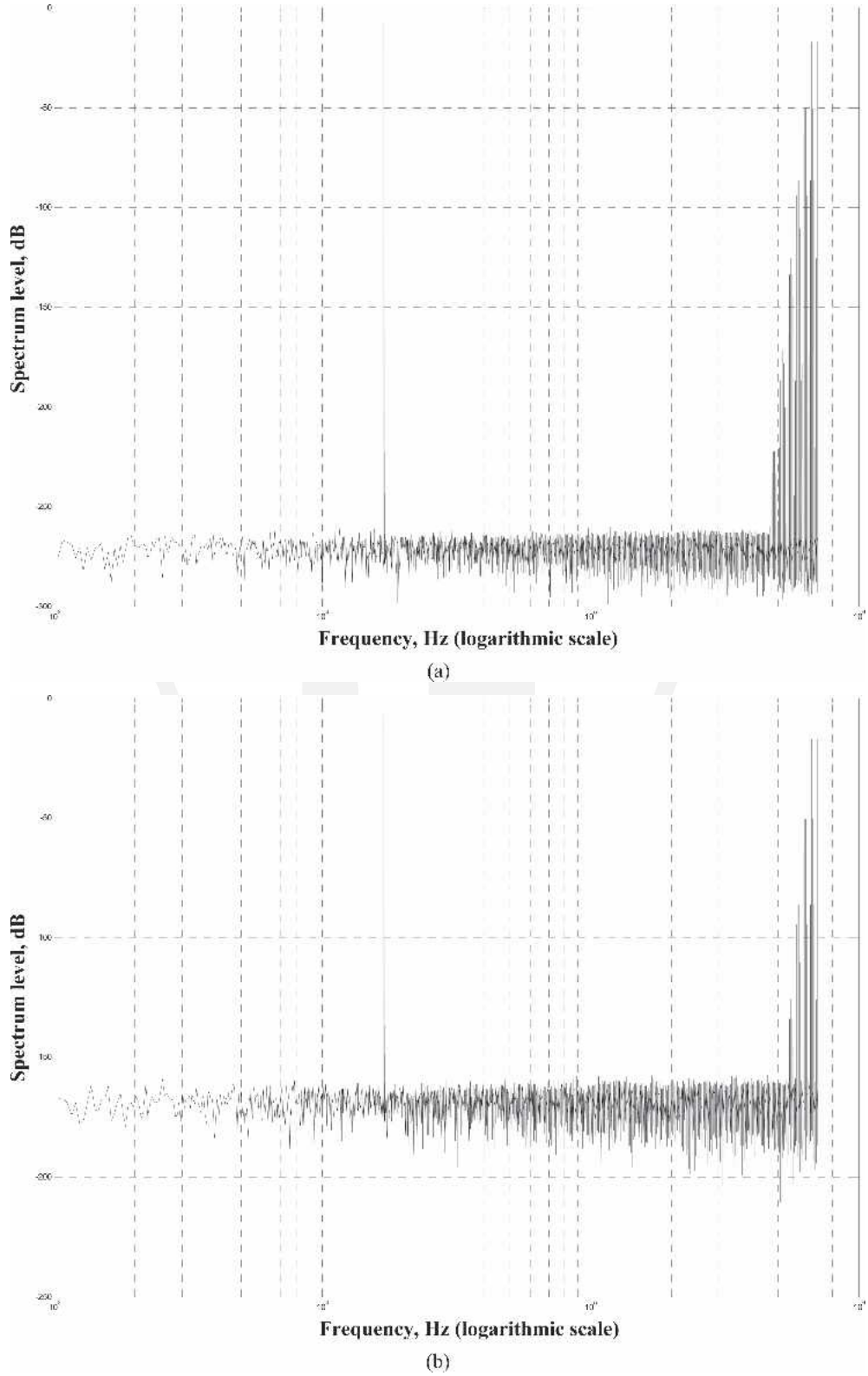


Fig. 11. Natural sampling open-loop PWM output spectra. (a) Without jitter. (b) With jitter. (c) Histogram of jitter present in PWM output.

spectral effect on jitter noise can be observed in terms of first-order noise shaping and follows theoretical predictions. However, of concern is the generation of intermodulation distortion, which for high-amplitude input signals is considered unacceptable in the context of a high-resolution amplifier. Also, contrary to normal feedback behavior and of specific interest in this paper, the distortion level is seen to rise progressively as the loop gain of the feedback amplifier is increased. Observe, however, that the results in Fig. 13 for uniform sampling PWM reveal a distortion similar to that induced by feedback, but here the distortion remains when the loop gain is reduced, a consequence of the input signal not being sampled at times coincident with the pulse transitions in the PWM output.

4 NODAL TRANSITION FILTERS (NTF) IN FEEDBACK PWM

In this section a technique using a loop filter to attenuate switching components is described to improve the linearity of feedback PWM. However, the inclusion of a filter introduces additional phase shift, which in the context of negative feedback presents problems of stability. The solution to stability has been solved here using a "constant-voltage" crossover filter [13], [14], which enables effectively the node in the circuit from which feedback is derived to be changed in a controlled frequency-dependent manner so as to bypass the internal PWM stage at high frequency. Constant-voltage filters are a specific class of loudspeaker crossover filters, where, for example, in a two-way crossover the high- and low-pass filter transfer functions sum to unity, thus exhibiting zero phase shift, a characteristic critical for stability when such filters are introduced within a negative-feedback loop. Because of the node-shifting property of a crossover filter, the process is called a nodal transition filter (NTF). Fig. 14 illustrates a negative-feedback PWM amplifier similar to that shown in Fig. 10, but using an NTF with a low-pass filter $\gamma(f)$ and a high-pass filter $1 - \gamma(f)$, where the constant-voltage properties follow from $\gamma(f) + [1 - \gamma(f)] \equiv 1$.

Fig. 15 presents an equivalent but more efficient topology than Fig. 14, where the filter $1 - \gamma(f)$ is derived from $\gamma(f)$. To demonstrate equivalence let the internal PWM and output stages have a nonlinear transfer function N. Referring to Fig. 14, the signal V_f fed back to the inverting input of A is

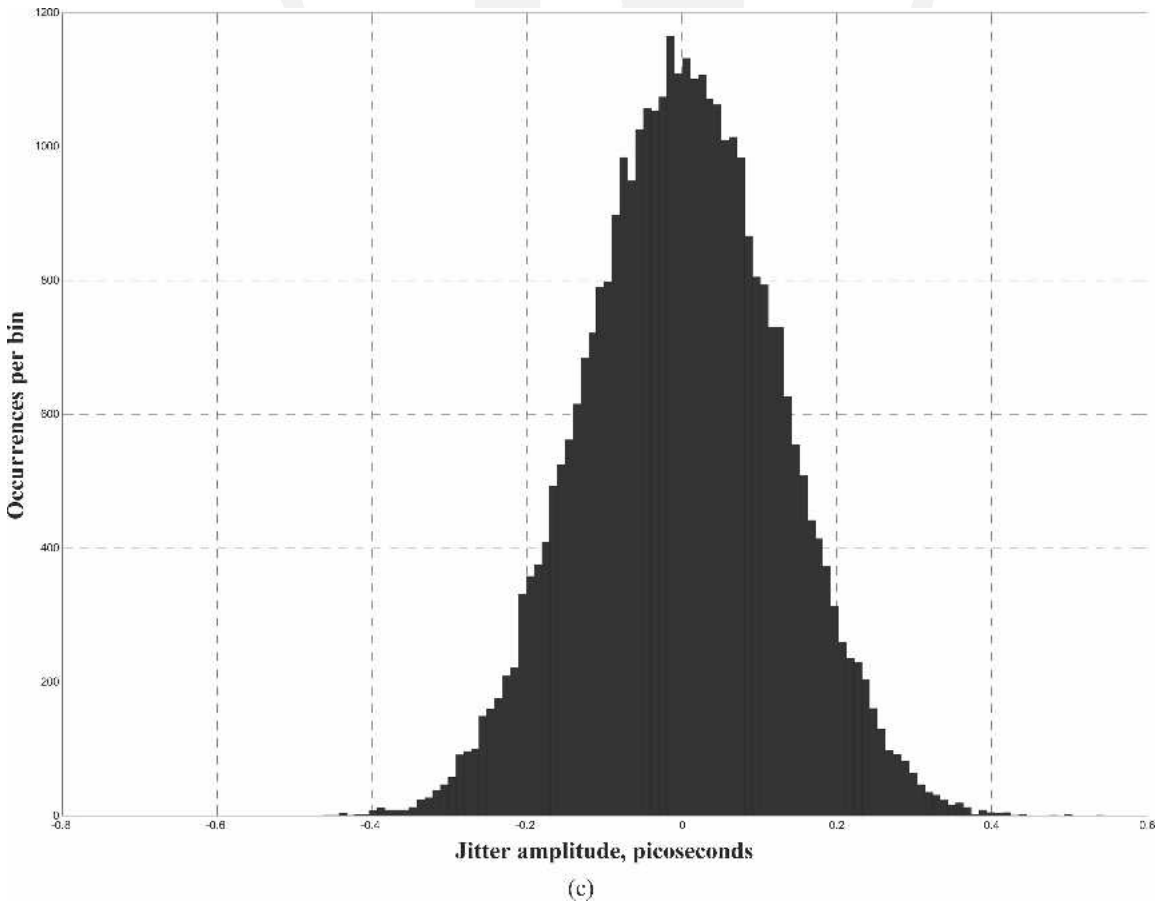
$$V_f = \mathrm{OUT}(f)\left[\gamma(f) + \frac{(1 - \gamma(f))}{N}\right]B. \tag{23}$$

Similarly, analyzing the topology of Fig. 15,

$$V_f = [\mathrm{OUT}(f) - V_\varepsilon + \gamma(f)V_\varepsilon]B$$

$$V_\varepsilon = \mathrm{OUT}(f)\left(1 - \frac{1}{N}\right).$$

Eliminating V_ε, equivalence is confirmed since V_f is again given by Eq. (23).



(c)

Fig. 11. *Continued*

To determine the properties of an NTF enhanced amplifier, the topology shown in Fig. 15 is analyzed to derive the closed-loop transfer function $G_{\mathrm{NTF}}(f)$,

$$G_{\mathrm{NTF}}(f) = \frac{AN}{1 + AB[1 + \gamma(f)(N-1)]}. \tag{24}$$

To observe NTF operation note that $\gamma(f)$ is a low-pass filter, where $\gamma(f \to 0) \to 1$ and $\gamma(f \to \infty) \to 0$. Thus

$$G_{\mathrm{NTF}}(f)|_{f\to 0} = \frac{AN}{1+ABN} \to \left.\frac{1}{B}\right|_{AB \gg 1} \tag{25}$$

and

$$G_{\mathrm{NTF}}(f)|_{f\to \infty} = \frac{AN}{1+AB} \to \left.\frac{N}{B}\right|_{AB \gg 1}. \tag{26}$$

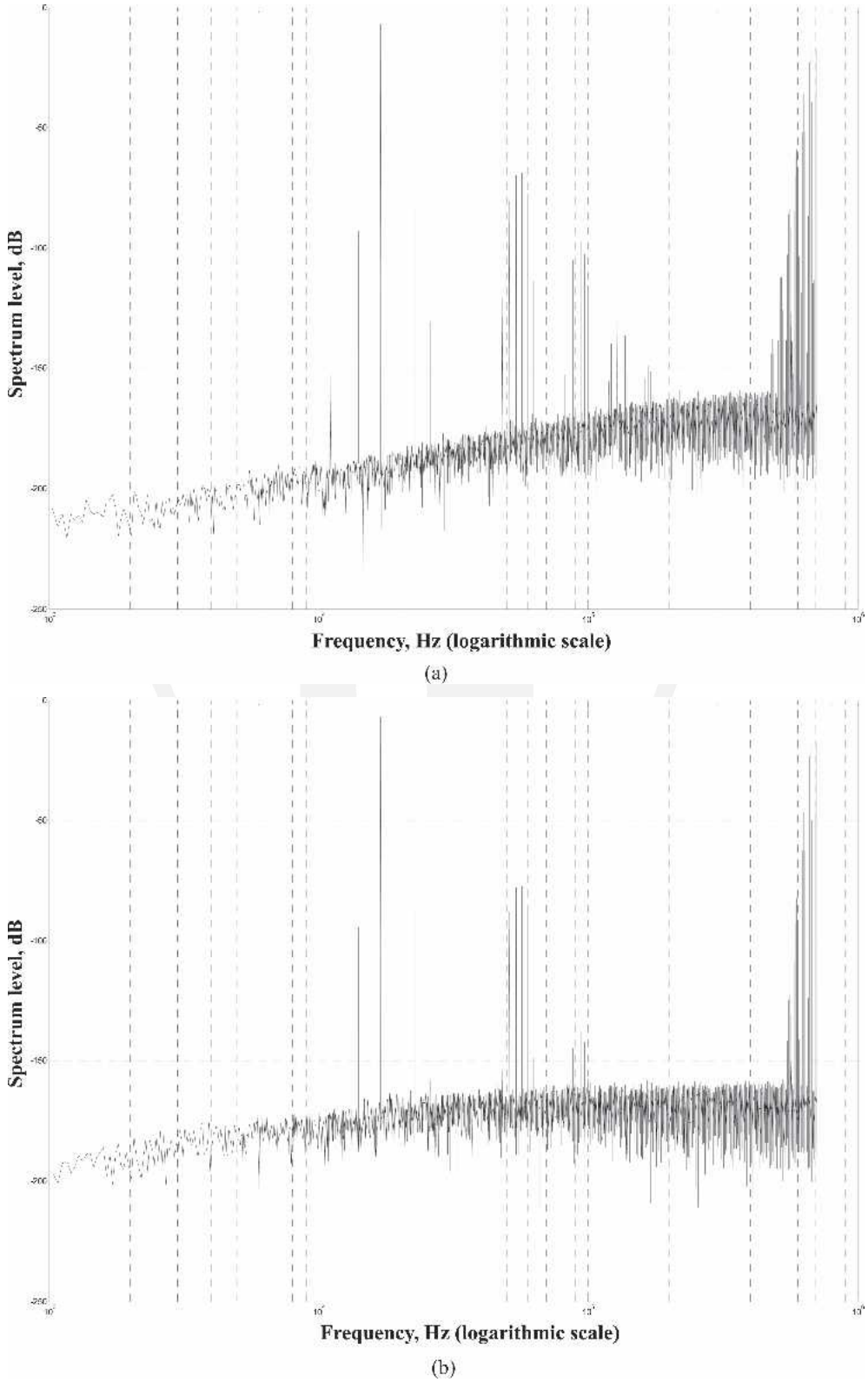
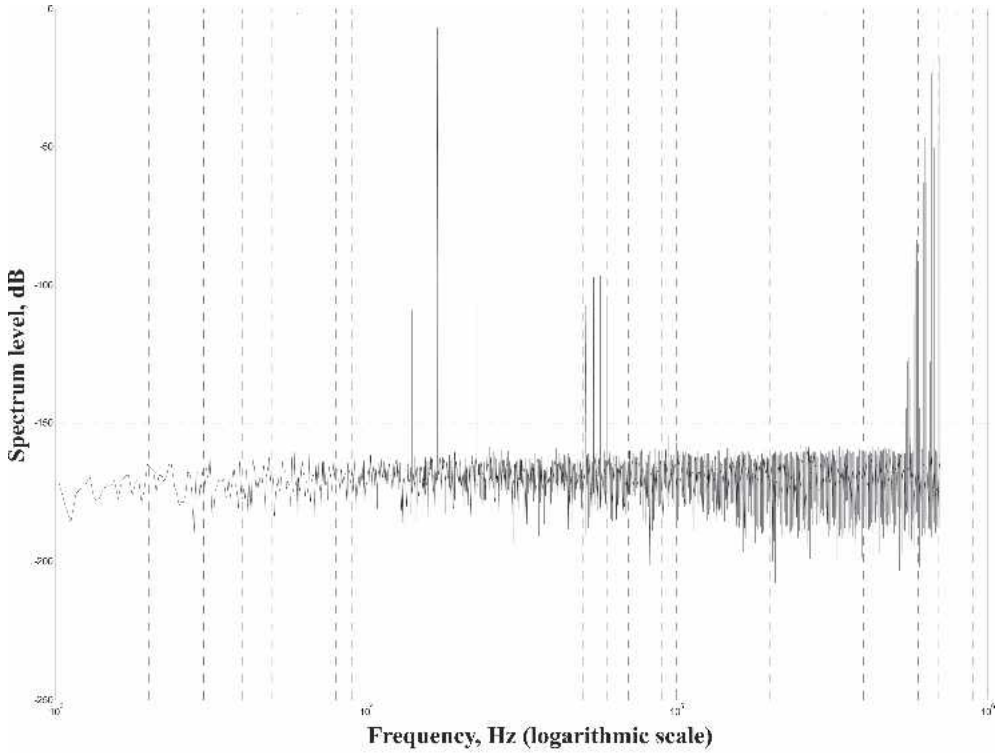


Fig. 12. Natural sampling closed-loop PWM output spectra. (a) With jitter, maximum loop gain. (b) With jitter, loop gain −20 dB below maximum. (c) With jitter, loop gain −40 dB below maximum. (d) With jitter, loop gain −60 dB below maximum.

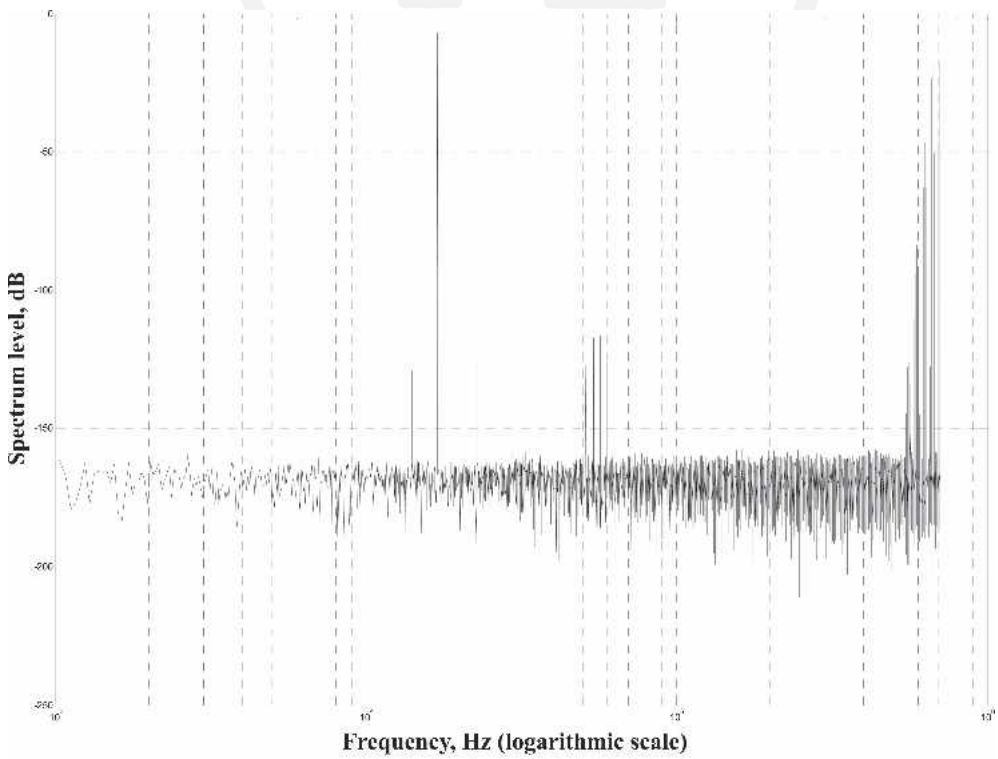
These limiting cases reveal that at lower frequencies the modulator and the output stage N are located within the overall feedback loop whereas at high frequency the feedback path is progressively transferred between nodes until feedback is derived from the output of the amplifier A. Thus PWM and the power output stages are excluded and appear open loop. Consequently with appropriate NTF design, $\gamma(f)$ can filter most of the switching components produced by the PWM stage while $1 - \gamma(f)$ maintains closed-loop stability by allowing feedback directly from the output of A at high frequency.

To investigate the distortion reduction performance of an NTF-enhanced PWM amplifier, let

$$\gamma(f) = \frac{1}{1 + \sum_{r=1}^{R} a_r (\mathrm{j}2\pi f)^r}. \tag{27a}$$



(c)



(d)

Fig. 12. *Continued*

Here $\gamma(f)$ is a low-pass filter of order R with filter coefficients $\{a_1, \ldots, a_R\}$, where at high frequency

$$\gamma(f)|_{f\to\infty} \to \frac{1}{a_R(\mathrm{j}2\pi f)^R}. \tag{27b}$$

To determine the sensitivity of $G_{\mathrm{NTF}}(f)$ to N, an error function $E_{\mathrm{NTF}}(f)$ is defined [15],

$$E_{\mathrm{NTF}}(f) = 1 - \frac{G_{\mathrm{NTF}}(f)}{G_{\mathrm{NTF}}(f)|_{\mathrm{target}}}. \tag{28}$$

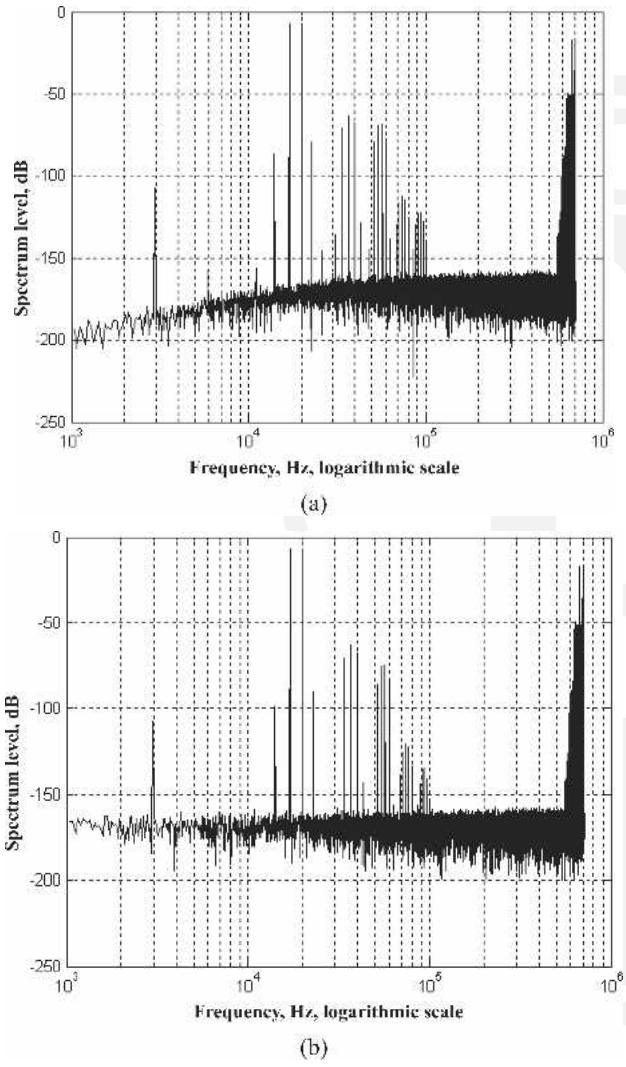


Fig. 13. Uniform sampling closed-loop PWM output spectra. (a) With jitter, loop gain −20 dB below maximum. (b) With jitter, loop gain −60 dB below maximum.

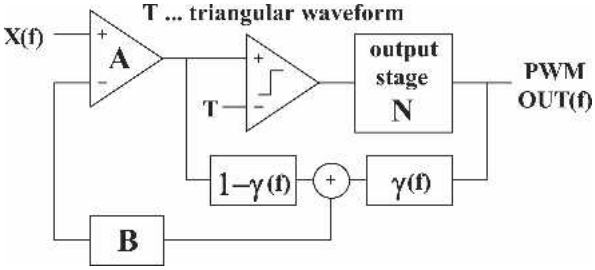


Fig. 14. Analog NTF feedback PWM.

The target transfer function $G_{\mathrm{NTF}}(f)|_{\mathrm{target}} = G_{\mathrm{NTF}}(f)$ for $N = 1$, and from Eq. (24),

$$E_{\mathrm{NTF}}(f) = \frac{(1-N)\{1+AB[1-\gamma(f)]\}}{1+AB[1+\gamma(f)(N-1)]}. \tag{29a}$$

Assuming $N \approx 1$ and $AB \gg 1$, $E_{\mathrm{NTF}}(f)$ then simplifies to

$$E_{\mathrm{NTF}}(f) \approx \left.\frac{(1-N)[1-\gamma(f)]}{1+\gamma(f)(N-1)}\right|_{AB\gg 1} \approx (1-N)[1-\gamma(f)]|_{N\approx 1}. \tag{29b}$$

Eq. (29b) reveals that the output-stage error is shaped in frequency by the high-pass filter $1 - \gamma(f)$ and not the loop gain AB. However, because the NTF is a constant-voltage crossover filter when high- and low-pass transfer functions sum to unity, the high-pass filter is limited to (pseudo) first order even if the low-pass filter has a high rate of attenuation. This constraint on the filter order is demonstrated in the following.

From Eq. (27a) the derived high-pass filter $1 - \gamma(f)$ follows,

$$1-\gamma(f) = \frac{\sum_{r=1}^{R} a_r(\mathrm{j}2\pi f)^r}{1+\sum_{r=1}^{R} a_r(\mathrm{j}2\pi f)^r}. \tag{30a}$$

Eq. (30a) shows that at high frequency $[1 - \gamma(f)]|_{f\to\infty} \to 1$, but at low frequency the asymptotic rate of attenuation is dictated by the term with the lowest power in the numerator, namely, as $f \to 0$,

$$[1-\gamma(f)]|_{f\to 0} = \frac{\sum_{r=1}^{R} a_r(\mathrm{j}2\pi f)^r}{1+\sum_{r=1}^{R} a_r(\mathrm{j}2\pi f)^r} \to a_1 \mathrm{j}2\pi f. \tag{30b}$$

A limit on the asymptotic slope applies also to the derived high-pass transfer function when the high- and low-pass filter orders are interchanged, that is, for a high-pass filter of order R,

$$[1-\gamma(f)]|_{f\to 0} = \frac{a_R(\mathrm{j}2\pi f)^R}{1+\sum_{r=1}^{R} a_r(\mathrm{j}2\pi f)^r} \to a_R(\mathrm{j}2\pi f)^R. \tag{31a}$$

Then the derived low-pass filter high-frequency asymptotic response is

$$\gamma(f)|_{f\to\infty} = \frac{1+\sum_{r=1}^{R-1} a_r(\mathrm{j}2\pi f)^r}{1+\sum_{r=1}^{R} a_r(\mathrm{j}2\pi f)^r} \to \frac{a_{R-1}}{a_R \mathrm{j}2\pi f}. \tag{31b}$$

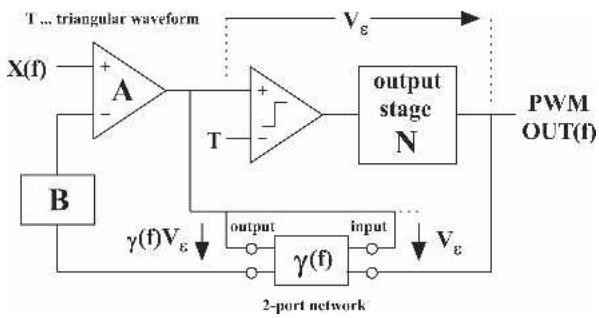


Fig. 15. Equivalent NTF feedback PWM.

Although a high-order high-pass filter increases the noise-shaping advantage within the bound of loop gain, Eq. (31b) shows that there is less attenuation of the switching components due to the first-order response of the derived low-pass filter. Conversely, even with a high-order low-pass filter and with $AB \gg 1$, the derived filter cannot exceed a first-order response. Here the low-frequency asymptotic noise-shaped error is derived from Eqs. (29b) and (30b),

$$E_{\mathrm{NTF}}(f)|_{f\to 0} \approx a_1 \mathrm{j}2\pi f(1-N). \tag{32}$$

To conclude this section the error function $E_{\mathrm{NTF}}(f)$ is compared against the noise-shaping transfer function (NSTF) used, for example, in the study of SDM. Fig. 16 shows the same NTF amplifier topology, but where the PWM output stage has been replaced with a linear gain stage N followed by an additive noise source V_{d}. Analyzing this topology, the output signal $\mathrm{OUT}(f)$ expressed as a function of input $X(f)$ and noise source V_{d} then follows,

$$\mathrm{OUT}(f) = \frac{1-AB[1-\gamma(f)]}{1+AB[1+N\gamma(f)-\gamma(f)]}V_{\mathrm{d}} + \frac{NA}{1+AB[1+N\gamma(f)-\gamma(f)]}X(f). \tag{33}$$

If the output stage gain is set to $N = 1$ to match the optimum conditions in the NTF amplifier, Eq. (33) reduces to

$$\mathrm{OUT}(f) = \frac{1}{1+AB}V_{\mathrm{d}} - \frac{AB}{1+AB}[1-\gamma(f)]V_{\mathrm{d}} + \frac{A}{1+AB}X(f) \tag{34a}$$

that is, when the loop gain is large,

$$\mathrm{OUT}(f)|_{AB\gg 1} \to \frac{1}{AB}V_{\mathrm{d}} - [1-\gamma(f)]V_{\mathrm{d}} + \frac{1}{B}X(f). \tag{34b}$$

Alternatively, expressed in terms of the closed-loop transfer function $H(f)$ and the noise-shaping transfer functions $D_{f_1}(f)$, $D_{f_2}(f)$,

$$\mathrm{OUT}(f) = D_{f_1}(f)V_{\mathrm{d}} - D_{f_2}(f)V_{\mathrm{d}} + H(f)X(f). \tag{34c}$$

Eqs. (34) reveal the closed-loop transfer function $H(f)$ as the standard canonic expression for a negative-feedback amplifier, whereas the additive output noise is shaped by two distinct NSTFs. $D_{f_1}(f)$ follows normal feedback theory and shows the noise shaped by the inverse of the amplifier loop gain; however, the second function, $D_{f_2}(f)$ [see Eq. (34b)], tends to $1-\gamma(f)$ when the loop gain is large. In practice noise shaping is dominated by $D_{f_2}(f)$. Thus the NSTF is equivalent to the high-pass transfer function of the NTF and compares to $E_{\mathrm{NTF}}(f)$ defined in Eq. (29b). This analysis confirms that increasing the loop gain AB offers little advantage with respect to the noise-shaping output error as the NTF limits the NSTF to a first-order response.

To demonstrate the validity of the NTF methodology proposed in this section, a circuit simulation was performed on the PWM amplifier presented in Fig. 17, which is derived from the conceptual topology shown in Fig. 15. In this example a six-stage RC ladder network was used, where $R = 1\ \mathrm{k}\Omega$ and $C = 1\ \mathrm{nF}$. Results are presented both with and without NTF, so changes in waveform detail can be observed. In both simulations the PWM output and the output of amplifier A were computed. The results shown in Fig. 18 are for a simple feedback amplifier where high-frequency signal components are present within the feedback loop, whereas in Fig. 19 the inclusion of the NTF suppresses these elements to reveal a waveform segment virtually free of high-frequency artifacts. Fig. 20 presents an alternative NTF using a second-order LCR filter, and Fig. 21 shows a simulation to demonstrate high-frequency suppression. In Fig. 22 a variant of the NTF-enhanced PWM amplifier is illustrated. Here the low-pass filter $\gamma(f)$

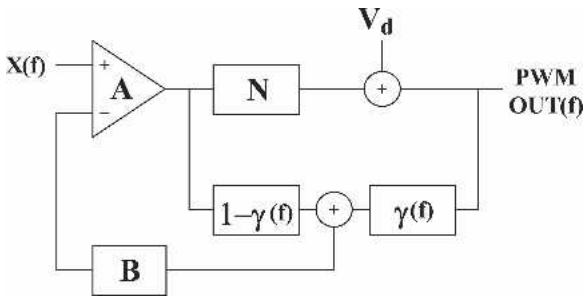


Fig. 16. NTF-type loop with PWM output stage replaced with noise source.

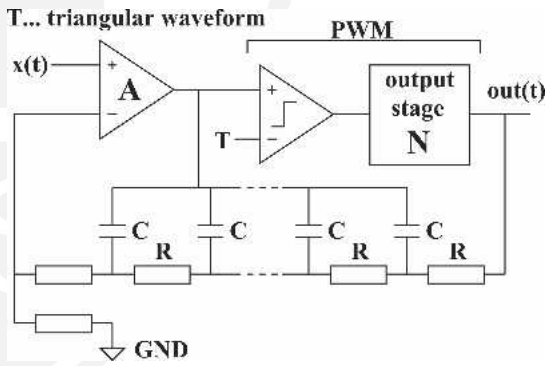


Fig. 17. Simulation of NTF PWM amplifier.

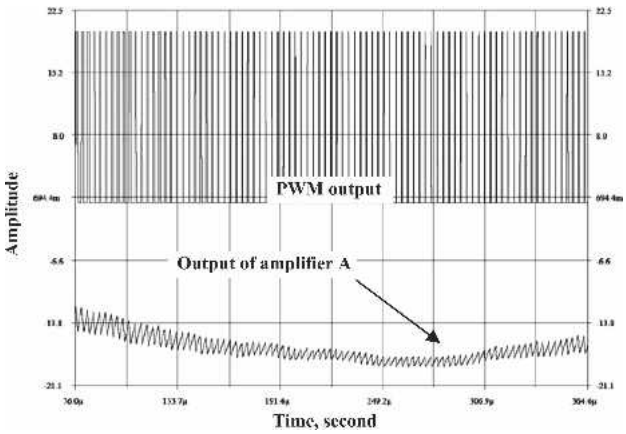


Fig. 18. PWM amplifier without NTF.

is now positioned after amplifier *A* to attenuate switching artifacts prior to the PWM stage and the high-pass filter $1 - \gamma(f)$ is placed in a feedforward path to the output where the composite signal forms the feedback signal.

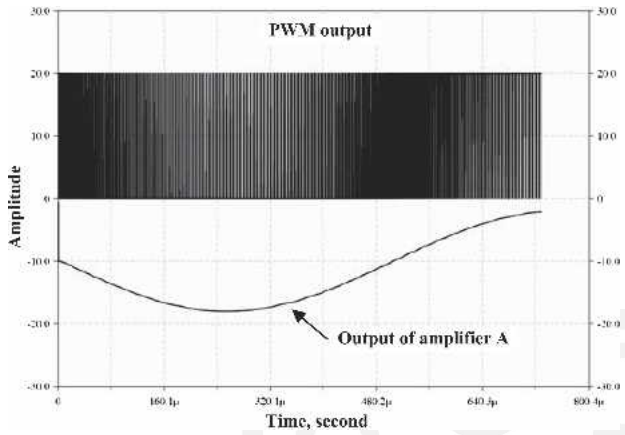


Fig. 19. PWM amplifier with NTF.

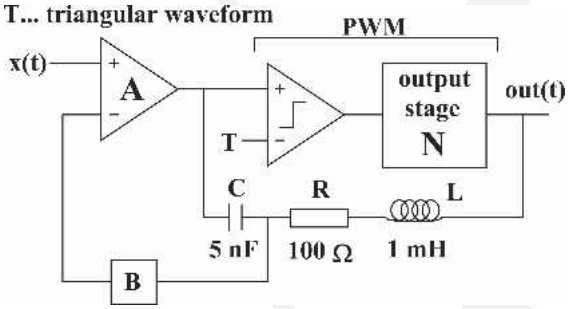


Fig. 20. PWM amplifier with *LCR* NTF.

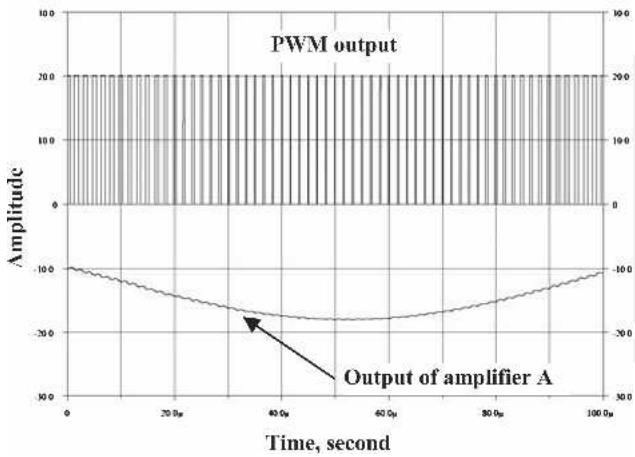


Fig. 21. Signal segment for PWM amplifier with *LCR* NTF.

To summarize this section, the technique of using an NTF within the closed loop of a PWM feedback amplifier achieves two principal performance enhancements:

- The low-pass filter $\gamma(f)$ of unconstrained order suppresses high-frequency switching artifacts present within the feedback loop and thus lowers distortion as the input signal to the modulator has reduced high-frequency content.
- Feedback achieves first-order distortion reduction as described by Eqs. (29), (32), and (34), similar to a conventional feedback amplifier incorporating a first-order forward-path amplifier.

5 INTERNAL SWITCHING COMPENSATION IN NEGATIVE-FEEDBACK PWM AMPLIFIERS

Section 3 analyzed the problem of distortion resulting from switching components within a PWM feedback amplifier and in Section 4 a method of filtering within the loop was discussed. As simulation revealed, the suppression of high-frequency signals within the feedback loop enables PWM to realize its full linearity potential within the bounds of comparator switching performance, output signal level control and both power supply and output-stage-induced switching artifacts. In this section an alternative approach is presented where a reference (namely, very low distortion) naturally sampled PWM side chain is located within the amplifier feedback loop and used to cancel the switching components of the main PWM stage. In this scheme filtering of the PWM signal fed back from the output of the amplifier can be omitted so that full-bandwidth closed-loop control is retained. The proposed scheme is shown in Fig. 23, where the loop contains two modulators, both driven from the output of amplifier *A*.

In the side chain the output of the reference PWM is subtracted from the output of amplifier *A* to form the idealized PWM error signal, which under optimum alignment contains no low-frequency information and thus carries only high-frequency switching distortion. This signal can be filtered, although there are constraints imposed by phase distortion which may affect adversely the process of compensation. The high-frequency error signal is then added to the fed back signal, where the summation process is designed so that, ideally, the switching signal in the main PWM output and the derived side chain PWM error cancel. As such the fed back signal matches closely the output of amplifier *A,* necessary to maintain proper loop behavior, but remains sensitive to the error signal between the main PWM output stage and the side chain reference

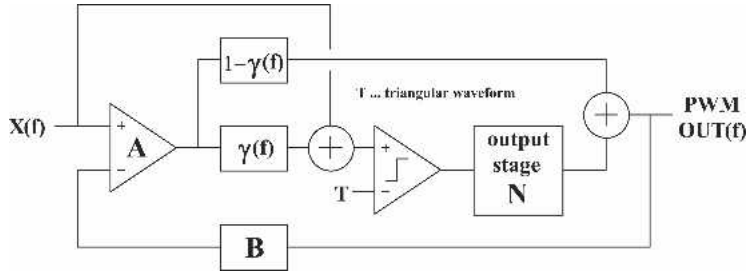


Fig. 22. NTF PWM amplifier reconfigured with feedforward path.

PWM, the latter having to be optimized for low distortion. It is critical for PWM signal levels to be matched by scaling as in practical amplifiers there will be differences between the main PWM output and the low-level reference PWM output. It is therefore desirable to calibrate the amplifier to achieve maximum switching cancellation.

To demonstrate the effectiveness of compensation, circuit simulation was performed for an optimally aligned PWM feedback amplifier. Fig. 24 presents an example output waveform for amplifier *A,* where an absence of switching-induced ripple is revealed. The corresponding PWM output waveform is also shown. However, for a nonideal output stage (and that includes gain error) there can be additional distortion resulting from unsuppressed high-frequency switching artifacts degrading the PWM process. Nevertheless, switching compensation can still offer substantial improvement, even if cancellation is imperfect. To summarize,

- Nonideality in the output stage adds distortion to the output voltage, as occurs in conventional analog amplifiers.
- Nonideality including simple gain error also implies error between the main PWM output stage and the reference PWM stage. This results in high-frequency switching ripple being added to the fed back signal, which can degrade modulator linearity.

These factors need to be considered when feedback is applied to PWM, otherwise some of the distortion reduction capability of feedback is impaired. For a practical amplifier design the following strategy is suggested:

1) Introduce an NTF to further reduce residual switching artifacts, as shown in Fig. 25. However, because this filter no longer has to suppress the whole of the switching distortion of the PWM, it can be less invasive and designed to facilitate an increase in loop gain. For example, it was shown in Section 4 [see Eqs. (31)] that if an NTF high-pass filter is high order then the derived low-pass filter is first order. Hence if switching suppression has already been implemented, a first-order low-pass filter allows some additional switching attenuation whereas a high-order high-pass filter enables improved noise shaping according to Eqs. (34). In practice the NTF would be selected using experimental data, including the phase margin and taking into account the PWM sampling rate to achieve the best overall performance compromise.

2) It is recommended that the side-chain process has an embedded controller (see also Section 6) to optimize the gain of the output stage so that, on average, switching artifacts are minimized. Fig. 25 shows a possible basic system topology incorporating variable gain and controller.

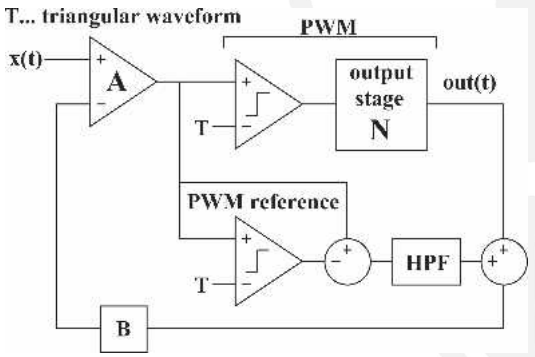


Fig. 23. PWM amplifier including reference PWM configured for switching compensation.

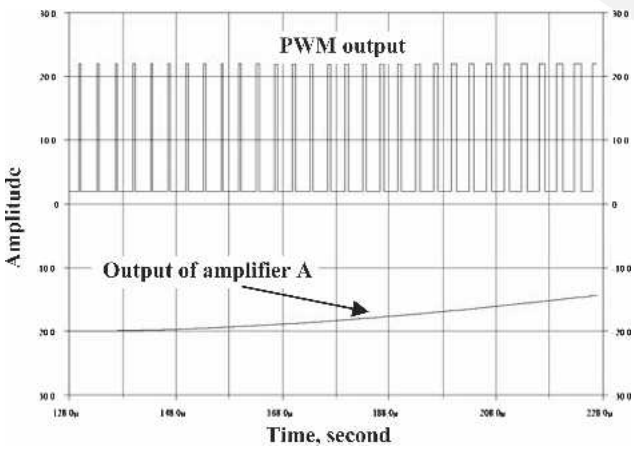


Fig. 24. Signal segment for PWM amplifier with side-chain switching compensation.

6 PREDICTIVE SWITCHING COMPENSATION IN NEGATIVE-FEEDBACK PWM AMPLIFIERS

An alternative approach to correct for switching distortion in PWM feedback amplifiers is to use a predictive side-chain based on open-loop PWM located in the input path to the main amplifier. In this scheme high-precision open-loop PWM first generates a naturally sampled PWM signal that predicts the optimum output of the feedback PWM amplifier. This signal is then subtracted from the input to produce a prediction of the switching error, which if performed accurately carries only the high-frequency signal components that normally reside well above the audio band. Now if the main feedback PWM amplifier were to output optimum naturally sampled PWM, then by subtraction, using the predicted switching error now present in the input signal to the feedback amplifier, the switching components in the amplifier feedback path are canceled, resulting in the desired low-ripple signal at the output of amplifier *A*. This scheme is shown in Fig. 26, where it is mandatory for both predictive PWM and inter-

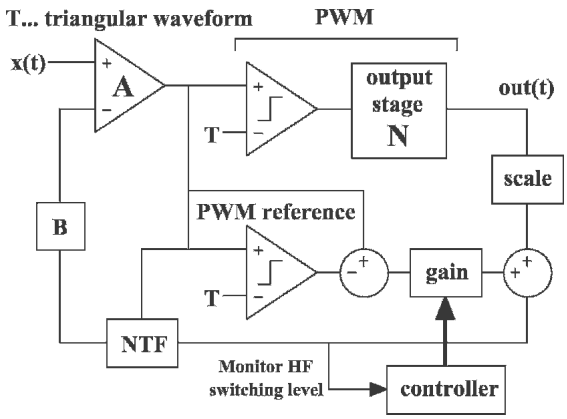


Fig. 25. PWM with NTF and auto balance.

nal-loop PWM to use the same reference triangular wave to match the gains and to time synchronize the two modulators. To confirm the efficacy of this technique, a circuit-level simulation achieved near-perfect ripple suppression using a relatively high loop gain established by a forward path amplifier with a time constant of 0.4 μs, a feedback factor of unity, and a PWM sampling rate set at 300 kHz.

Inspection of the signal flow in Fig. 26 reveals that the predictive process can be simplified to a cascaded open-loop PWM stage, as shown in Fig. 27, a result that initially may not have been anticipated. For this system to yield optimum switching waveform suppression requires identical comparators with matched output voltage ranges, consequently defining N_i as the gain of the predictive input PWM stage and N as the voltage gain of the internal PWM stage used to scale signals to the full output voltage swing. Then

$$NB = N_i. \tag{35}$$

Hence if the feedback factor B is constant in order to establish a well-defined closed-loop gain, N should be programmable so that loop conditions can be optimized dynamically according to Eq. (35). Thus, for example, changes in power-supply voltage and component tolerances can be compensated. A practical means to vary N is to modulate its power-supply voltage as in PWM this has a multiplicative function. Fig. 28 illustrates a forward control loop that monitors the short-term (say rms) power of the error $\varepsilon(t)$. This information can then be used to control the output-stage gain N by modulating the output-stage power-supply voltage with the aim of minimizing the power of $\varepsilon(t)$, where this implies $NB = N_i$. Since this strategy minimizes the switching error, the generation of secondary distortion as discussed in Section 3 is also minimized. Consequently if the PWM output stage and feedback path together have a gain equal to 1, assuming $N_i = 1$, then the loop gain of the negative-feedback amplifier is A. This process can be viewed as a secondary error correction strategy compensating for nonideal gain in the power output stage and also allowing a degree of regulation for the associated power supply. Such a procedure then integrates power-supply regulation within the PWM output stage rather than as an independently controlled subsystem that has to be calibrated. However, in systems where the power-supply voltage is regulated independently an alternative method to control the output stage gain is to modulate the amplitude of the triangular wave used in the forward path PWM.

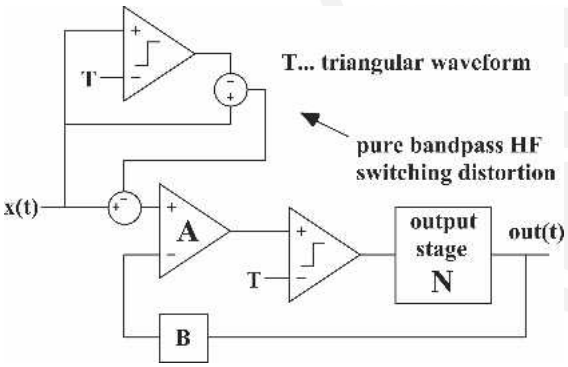


Fig. 26. PWM feedback amplifier with feedforward switching compensation.

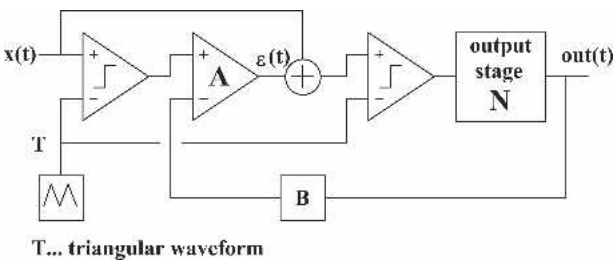


Fig. 27. Simplified PWM feedback amplifier with feedforward switching compensation.

The following analysis describes the distortion reduction characteristics of the PWM amplifier shown in Fig. 28 as well as the sensitivity to the magnitude of the error signal $\varepsilon(t)$. The following Fourier transforms are assumed:

$$\varepsilon(t) \Rightarrow E(f), \qquad x(t) \Rightarrow X(f), \qquad \text{out}(t) \Rightarrow \text{OUT}(f).$$

Also let input and output PWM stages have transfer functions N_i and N, respectively, where the optimum alignment is $NB = N_i$, with $N_i \to 1$. In this system the function of the feedforward path from amplifier input to comparator input should be observed since it makes $\varepsilon(t)$ a true error signal that is zero under optimum conditions. Hence if $A = 0$, the structure reverts to open-loop, naturally sampled PWM. From the PWM topology in Fig. 28 $E(f)$ can be expressed as

$$E(f) = A[X(f)N_i - \text{OUT}(f)B]$$

with $\text{OUT}(f)$ given via the output stage by

$$\text{OUT}(f) = N[X(f) + E(f)].$$

Substituting for $E(f)$ the input–output transfer function then becomes

$$\frac{\text{OUT}(f)}{X(f)} = N\left(\frac{1 + N_i A}{1 + NAB}\right) \to \left.\frac{N_i}{B}\right|_{NAB \gg 1}. \tag{36}$$

Similarly $E(f)$ can be determined by eliminating $\text{OUT}(f)$,

$$\frac{E(f)}{X(f)} = \left(\frac{NAB}{1 + NAB}\right)\left(\frac{N_i}{NB} - 1\right) \to \left.\left(\frac{N_i}{NB} - 1\right)\right|_{NAB \gg 1}. \tag{37}$$

Eq. (37) confirms that $E(f)$ is zero for $A = 0$ and for $NB = N_i$, as implied by Eq. (35). Hence amplifier A only

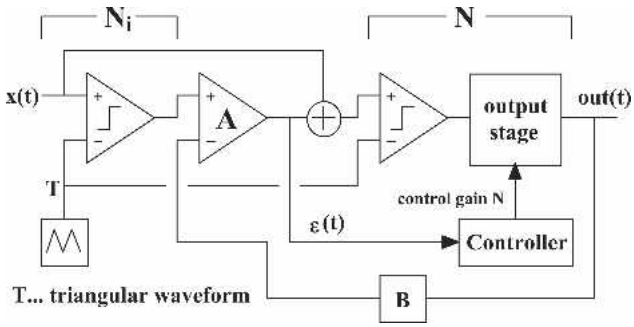


Fig. 28. PWM feedback amplifier with feedforward switching compensation and dynamic loop gain control.

produces a finite output when the input PWM and output PWM stages differ in their performance due to gain errors and output-stage imperfections. The analysis also shows that the ultimate performance of this type of PWM feedback topology is bounded by the linearity of the open-loop input PWM stage, where Eq. (36) confirms that even for finite gain A, with optimum alignment $NB = N_i$, then the overall transfer function is N_i/B. Also, it should be observed that the closed-loop gain performance is not sensitive to the condition stated in Eq. (35). Thus even if a secondary control loop modulates N to seek optimum loop conditions, this will not strongly affect the output other than to fine-tune the distortion performance by suppressing switching artifacts present in $\varepsilon(t)$.

6.1 Output Stage for Digital PWM

As a corollary to this section, because the amplifier with predictive compensation requires a PWM input signal, it follows that it can be used not only with a naturally sampled PWM code but also with a code derived from uniformly sampled, digitally derived PWM where the output is not directly amenable to analog control to take into account output-stage imperfections. Digital PWM normally uses linearization and noise shaping [10], [11] with pulse transitions calculated algorithmically. Consequently the pulses can be applied directly via a two-level DAC to the analog feedback power output stage because digital PWM forms the predictive stage required to reduce switching ripple. To implement a practical power amplifier there are signal-processing factors to consider. Fig. 29 shows a conceptual system together with a simplified signal flow diagram. In interpreting the functionality of this system, its mixed signal architecture must be considered as some processes are digital while others are analog. The basic function of subsystems N_1, N_2, N_3, N_4 are summarized as follows: N_1 represents uniformly sampled digital PWM, where the output is a time-domain quantized PWM signal sampled at a high clock frequency so that noise shaping realizes an acceptable signal-to-noise ratio (SNR). Including time-delay compensation T_x, an error signal is derived from across N_1 and fed forward to the amplifier input via a high-order digital low-pass filter N_2 designed to attenuate switching components. Significantly the filtered error signal is of low level; consequently the DAC and the analog reconstruction filter in this path require high accuracy but only low resolution. N_3 also includes a high-order, low-pass digital filter and DAC to form the feedforward signal summed with the PWM comparator input. Finally N_4 represents the transfer function of the PWM output stage. In addition there are two delay networks T_x and T_y to compensate for digital filter and process time delays. Analyzing this system the overall transfer function G_d of the simplified digital PWM amplifier is

$$G_d = \frac{N_4 A}{1 + N_4 AB}\left[\frac{N_1 N_3}{A} + N_1(e^{-j\omega T_y} - N_2) + N_2 e^{-j\omega T_x}\right]. \quad (38)$$

Eq. (38) shows that when $AB \gg 1$, then there is low sensitivity to N_3 and N_4, implying that the DAC performance in the feedforward path is noncritical and that output-stage distortion is reduced by feedback. Assuming high loop gain, Eq. (38) reduces to

$$G_d|_{AB\gg 1} \rightarrow N_1(e^{-j\omega T_y} - N_2) + N_2 e^{-j\omega T_x}. \quad (39)$$

If within the low-frequency pass band $N_2 \Rightarrow e^{-j\omega T_y}$, then Eq. (39) reduces further to just a pure time delay of $T_x + T_y$, revealing the most critical process is the low-pass filter N_2 associated with feedforward error correction about the uniformly sampled PWM stage, noting that T_x is chosen to minimize the level of error in this path. However, if N_1 includes linearization then even this condition for N_2 is noncritical. To summarize, the following features are highlighted:

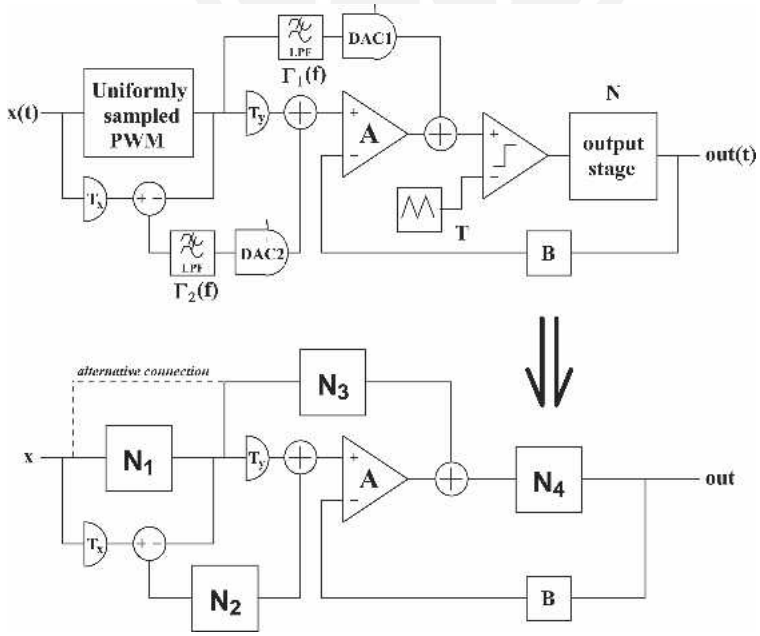


Fig. 29. Conceptual amplifier configuration for uniformly sampled digital PWM.

- The digital PWM stage N_1 normally includes linearization; therefore it produces low distortion.
- N_2 is critical to desensitizing system dependence on N_1, but because it processes only very low-level signals and is defined mainly by the passband of a digital low-pass filter, N_2 can be extremely accurate.
- N_3 is noncritical where any associated distortion is lowered because of feedback.
- If N_2 is optimized then the system has low sensitivity to the linearization process in the digital PWM stage N_1; possibly this feedforward correction procedure renders linearization unnecessary.
- The PWM output of N_1 drives the power amplifier directly, forming a predictive switching signal that facilitates the lowering of switching ripple in the feedback loop of the power amplifier.

This summary concludes the discussion on PWM. In the next section SDM digital power amplification is considered, using a highly stable coder and an output stage with low commutation losses.

7 SDM POWER AMPLIFIER SYSTEMS

Earlier work [4] has presented a discussion on switching power amplifiers based on a quantized SDM code. In this scheme the output of an SDM directly controls a power switch which, as with a PWM power amplifier, drives the loudspeaker via an analog low-pass filter. Variants of the output stage topology are reviewed in Section 7.2, whereas Sections 7.1.1 and 7.1.2 revisit SDM loop design as there are the following specific requirements.

- Adequate audio band SNR determined by the choice of NSTF.
- High-level input signal coding required as conventional high-order SDM normally limits maximum modulation depth, thus inhibiting maximum output from being achieved for a given power-supply voltage.
- Robust stability [16], that is, a low probability of instability at high input signal levels, ideally allowing a modulation index of unity.

7.1 Robust Loop Stability in SDM with High-Amplitude Input Signals

A conceptual digitally addressed SDM power amplifier scheme is shown in Fig. 30. Here, because of its wide adoption as a professional SDM coding algorithm and by way of illustration, the Sony FF SDM[4] topology [17] is included as the front-end coder. In practical SDM systems the sampling rate conversion is required to match the source signal to the SDM sampling rate, although for clarity that is not included in Fig. 30. The output stage is configured as a standard H bridge, where a method for circuit efficient ac-coupled interfacing with self dc restoration has been reported [4]. The design concept can use a resonant-mode power supply where the overall amplifier gain is modulated by scaling the H-bridge supply voltage. The use of a resonant supply locked in frequency and phase to the SDM clock can allow the H bridge to commutate during zero voltage transitions and thus reduce both switching loss and EMC interference significantly.

7.1.1 One-Sample SDM Look-Ahead

Theoretically the Sony FF SDM coder can yield more than sufficient SNR for power amplifier applications, especially in the context of the additional signal distortion inherent in power switches and also power supply noise. However, a weakness of the standard Sony topology is stability, especially under high levels of input signal,

[4]Sony FF SDM is a proprietary coding algorithm using local feedforward and local feedback paths; see [16] for a description.

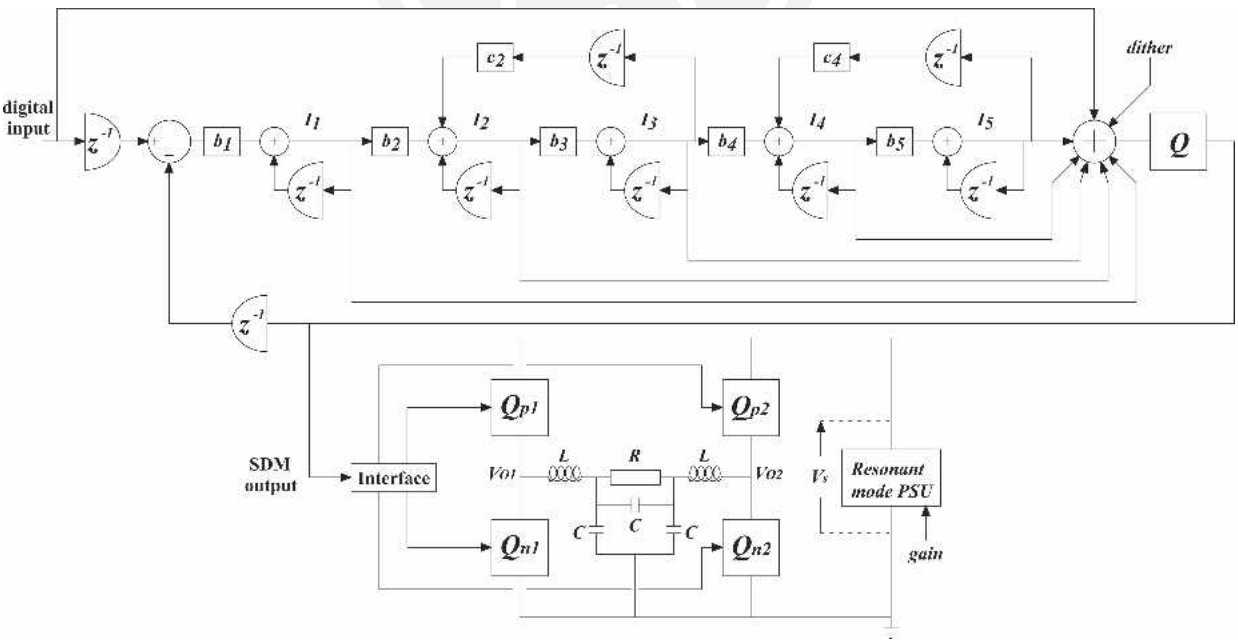


Fig. 30. Sony FF SDM with Type 2 bridge output stage.

which ultimately limits the maximum achievable modulation index. In this section a modified algorithm is explored based on the look-ahead principle [18] and an energy-balancing comparator. It is shown that with modest look-ahead significant gains in stability robustness can be achieved together with a low and constant coding latency. A further extension compresses under high-level excitation a subset of the state variables and is shown to allow stable operation up a modulation depth of unity. To demonstrate the probability of instability of the standard Sony FF SDM, results are derived using an earlier reported coding scheme [19], which stabilized the loop using a step-back-in-time procedure. A virtue of this method is that the step-back activity is an inverse measure of stability, that is, the more robust the loop, the lower the step-back activity. The simulation used the standard Super Audio CD (SACD) [3], [17] sampling rate f_{DSD} of 2.8224 MHz together with a 1-kHz input signal of amplitude 0.5. The corresponding SDM output spectrum [including a 24-bit, 88.2-kHz linear pulse-code modulation (LPCM) reference spectrum], quantizer input, and related amplitude histogram plots are shown in Fig. 31 with individual step-back activity events indicated by asterisks in Fig. 31(b). The step-back activity is relatively frequent, and if the input is increased further in level, catastrophic failure occurs.

In the standard feedback SDM algorithm a simple threshold decision is made on sample n as to whether the corresponding output sdm(n) is 1 or −1. Using Matlab notation, a typical threshold decision statement takes the following form:

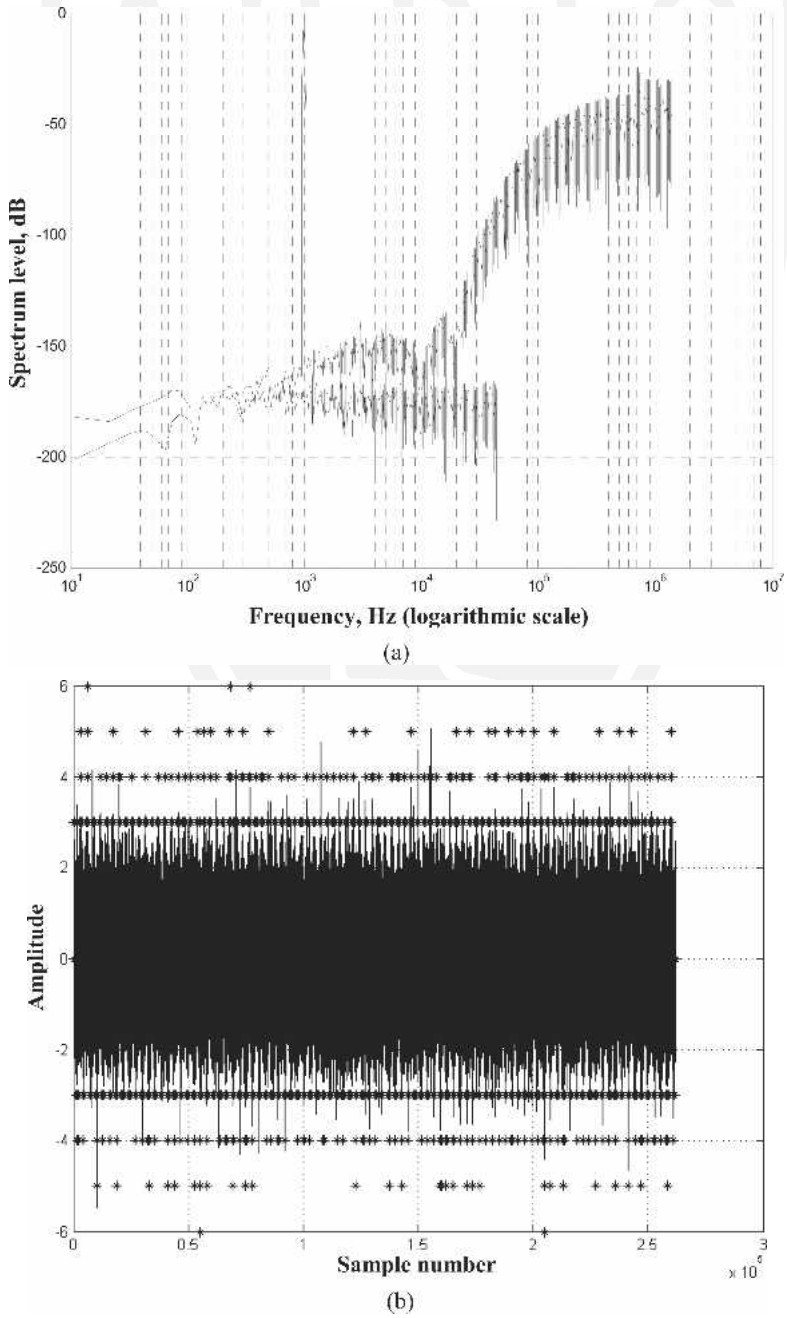


Fig. 31. Standard Sony FF SDM with step-back-in-time correction; input 0.50. (a) SDM output and 24 bit @ 88.2 kHz LPCM reference spectra. (b) Quantizer input with step back. (c) Quantizer input histogram.

```
% calculate input to quantizer
ss(n) = sum(I(1:order))+ax(n)+0.35*(rand(1,1)-.5);
% form SDM output using 2-level quantization
sdm(n) = sign(ss(n));
```

The first statement calculates the quantizer input ss(n) which includes a summation of the integrator outputs I(1:order), as shown in Fig. 30. It also includes RPDF dither, with a peak-to-peak amplitude range of 0.35 and a sampled input signal $ax(n)$. The second statement then performs a binary threshold comparison about a level of zero and generates the required SDM binary code.

A modification of the standard Sony FF SDM loop was investigated, where initially a one-sample look-ahead was implemented. In this scheme at sample n, the two pathways corresponding to sdm(n) = 1 and sdm(n) = −1 were computed and the corresponding states for sample n + 1 evaluated for the two options. The look-ahead algorithm is summarized as follows: Taking the present integrator states as I(1:5), the two possible state updates are calculated as $I1x, \ldots, I5x$ for sdm(n) = 1 and $I1y, \ldots, I5y$ for sdm(n) = −1, where, in Matlab notation,

```
% integrator update for sdm(n) = 1
I1x = I(1)+ax(n)-1;
I2x = I(2)+b2*I1x+c2*I(3);
I3x = I(3)+b3*I2x;
I4x = I(4)+b4*I3x+c4*I(5);
I5x = I(5)+b5*I4x;

% integrator update for sdm(n) = -1
I1y = I(1)+ax(n)+1;
I2y = I(2)+b2*I1y+c2*I(3);
I3y = I(3)+b3*I2y;
I4y = I(4)+b4*I3y+c4*I(5);
I5y = I(5)+b5*I4y;
```

Next, for each pathway signals PX and PY are calculated, representing the total energy of the five integrators states, although for integrator 5 this must include the summation of integrators 1 to 4 to represent the mandatory feedforward paths shown in Fig. 30 of the fifth-order Sony FF encoder, namely,

```
% look-ahead decision based on energy balance
```

$$PX = (I1x + I2x + I3x + I4x + I5x)^2 + I1x^2 + I2x^2 + I3x^2 + I4x^2 \quad (40)$$

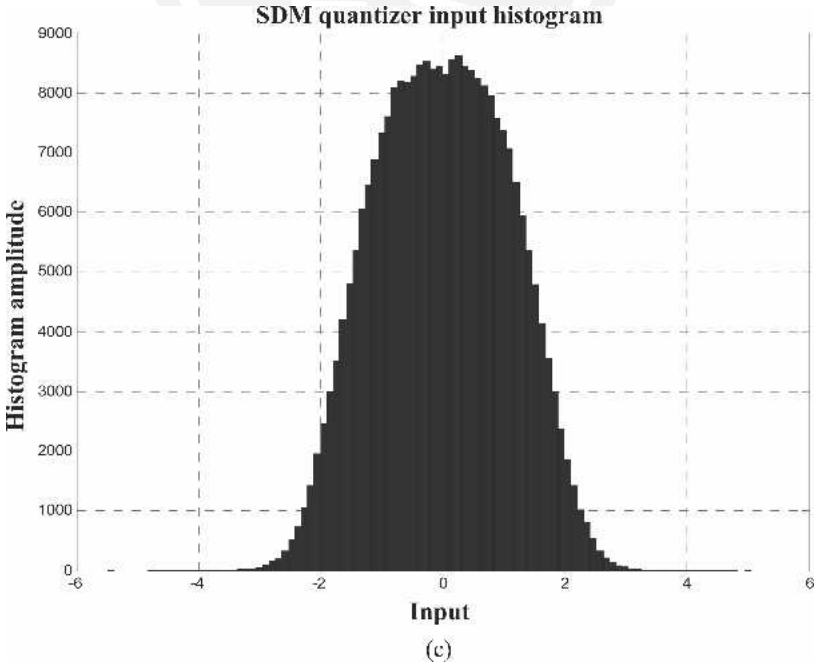
$$PY = (I1y + I2y + I3y + I4y + I5y)^2 + I1y^2 + I2y^2 + I3y^2 + I4y^2. \quad (41)$$

Having calculated the state energy estimates corresponding to the two pathways, an energy balance that includes dither rd(n) is used to select the actual SDM output sdm(n) for sample n,

$$\text{sdm}(n) = \text{sign}(\text{rd}(n) + PY - PX). \quad (42)$$

The loop integrators are then updated using the SDM output calculated by Eq. (42) in preparation for the next computational cycle. In addition, to help control stability using this modified algorithm, the loop (in this example) retained a step-back-in-time procedure [19] instigated by a simplified amplitude threshold comparator operating on the quantizer input.

A similar set of simulations was performed as for the standard Sony-FF SDM but using elevated peak input levels of 0.70 and 0.73. The corresponding results are shown in Figs. 32 and 33. Close inspection of the two output spectra reveals a slight overall reduction in SNR although the spectral shapes are almost identical. However, the quantizer input waveforms, and in particular the step-back activity, are changed. Although the input level has been



(c)

Fig. 31. *Continued*

increased from 0.5 to 0.7, activity is actually reduced with only two step-back incidents recorded in a window of 2^{18} samples. Increasing the input level to 0.73 caused the step-back activity to increase further, as might be anticipated for such a high-level signal, although it still falls within acceptable bounds and allowed the loop to remain stable. Consequently the energy-balancing equation with one-sample look-ahead achieves more robust SDM coding, especially with higher level input signals, a characteristic better matched to digital power amplifier applications.

7.1.2 Two-Sample SDM Look-Ahead

The look-ahead procedure was then extended from one to two samples to ascertain whether a further coding advantage is possible. The algorithm first calculated the state variables for a one-sample look-ahead for both a one- and a zero-output decision, as described in Section 7.1.1. Then from these two decisions two further sets of state variables were calculated, giving four sets for paths [1 0], [1 1], [0 0], and [0 1], as follows:

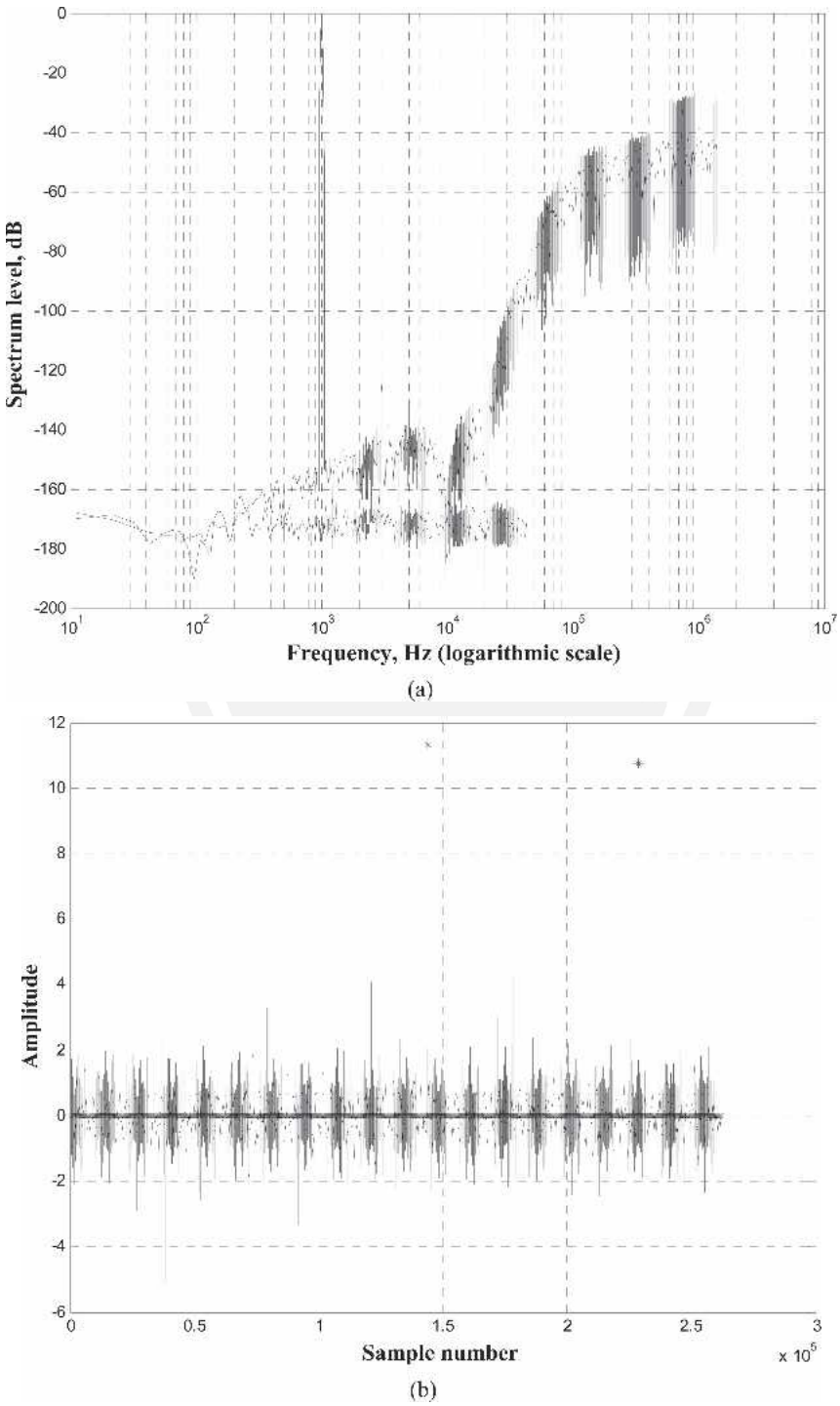


Fig. 32. Sony FF SDM with one-sample look-ahead; input 0.70. (a) SDM output and 24 bit @ 88.2 kHz LPCM reference spectra. (b) Quantizer input with step back. (c) Quantizer input histogram.

{I1×0, I2×0, I3×0, I4×0, I5×0} for path [1 0]

{I1×1, I2×1, I3×1, I4×1, I5×1} for path [1 1]

{I1y0, I2y0, I3y0, I4y0, I5y0} for path [0 0]

{I1y1, I2y1, I3y1, I4y1, I5y1} for path [0 1].

Energy estimates $mx0$, $mx1$, $my0$, and $my1$ were then determined for each of the four sets of state variables,

$$mx0 = (I1x0 + I2x0 + I3x0 + I4x0 + I5x0)^2 + I1x0^2 + I2x0^2 + I3x0^2 + I4x0^2 \quad (43)$$

$$mx1 = (I1x1 + I2x1 + I3x1 + I4x1 + I5x1)^2 + I1x1^2 + I2x1^2 + I3x1^2 + I4x1^2 \quad (44)$$

$$my0 = (I1y0 + I2y0 + I3y0 + I4y0 + I5y0)^2 + I1y0^2 + I2y0^2 + I3y0^2 + I4y0^2 \quad (45)$$

$$my1 = (I1y1 + I2y1 + I3y1 + I4y1 + I5y1)^2 + I1y1^2 + I2y1^2 + I3y1^2 + I4y1^2. \quad (46)$$

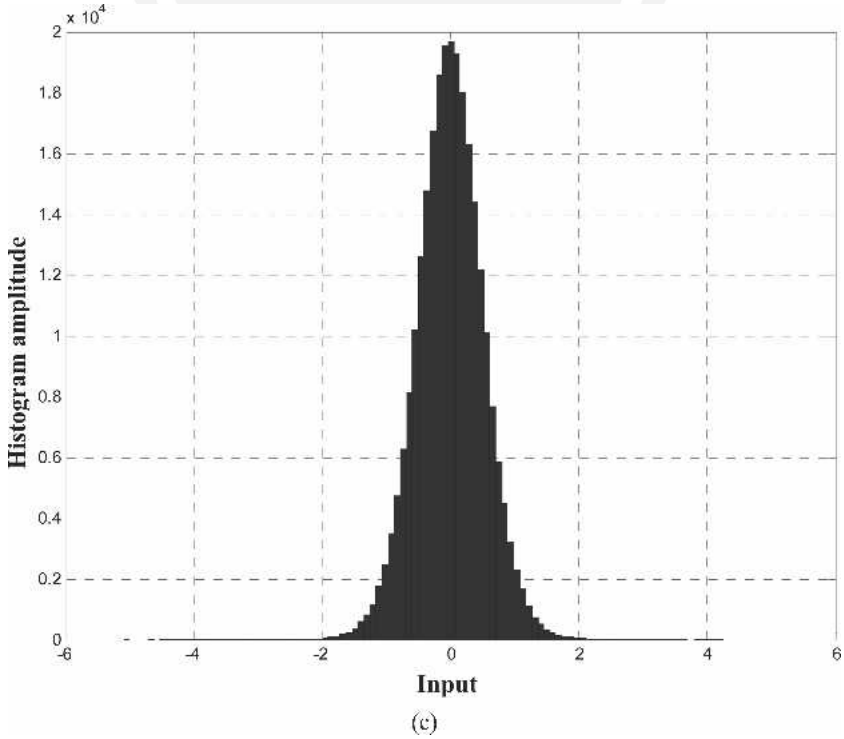
Note that x corresponds to a +1 output and y to a −1 on the first sample look-ahead. Finally, the SDM output decision sdm(n) was made based on the four look-ahead energy states by extracting initially the minimum of [$my0$ $my1$] and the minimum of [$mx0$ $mx1$] and then forming an energy difference equation,

$$\mathrm{sdm}(n) = \mathrm{sign}(rd(n) + \min([my0 \quad my1]) - \min([mx0 \quad mx1])). \quad (47)$$

Note that in Eq. (47) a dither source rd(n) is again included in the decision process. To show the complete procedure, a Matlab SDM simulation program without step-back correction for the two-sample look-ahead coder is presented in the Appendix. Simulation results confirm further improvement, such that with a peak input signal amplitude of 0.75 no step-back events were recorded. Also the simulation (using the code in the Appendix) without step-back correction remained stable over 2^{20} samples. This was repeated several times without problem other than mild nonlinear distortion similar to that displayed in Figs. 32(a) and 33(a). However, this simulation, although stable, proved to be close to the overload threshold such that increasing the peak input signal to 0.77 caused failure.

7.1.3 Dynamic State-Variable Compression with Two-Sample Look-Ahead SDM

A further modification is to incorporate dynamic compression of the state variables to enable two-sample look-ahead SDM to remain stable up to the maximum modulation index of unity. The method also removes the need for step-back correction so latency is no longer variable. Loop activity is monitored by observing the signal ss(n) described in Section 7.1.1, which in non-look-ahead SDM forms the input to the comparator Q shown in Fig. 30. If the magnitude of ss(n) exceeds a predetermined threshold, then the accumulated outputs of selected integrators $I_1(n)$, $I_2(n)$, . . . , $I_5(n)$ are replaced by compressed values $\beta_1 I_1(n)$, $\beta_2 I_2(n)$, . . . , $\beta_n I_5(n)$, where β_1, . . . , β_5 are the five integrator compression coefficients. The threshold level was determined by experiment but was set to be greater than the maximum signal normally encountered when the two-sample look-ahead loop was operating with an input of 0.75, that is, just within its stable regime as described in Section 7.1.2. This ensured that noise-shaping performance remained undisturbed for input signals up to a level of at least 0.75. Two variants of dynamic compression were tested with similar results as follows: Define T_h



(c)

Fig. 32. *Continued*

as the threshold level and D_p as damping. The compression algorithm then takes the generic form of a conditional loop. If abs(ss(n)) > T_h, then

$$I_1(n) \Rightarrow \beta_1 I_1(n), \quad I_2(n) \Rightarrow \beta_2 I_2(n), \quad I_3(n) \Rightarrow \beta_3 I_3(n),$$
$$I_4(n) \Rightarrow \beta_4 I_4(n), \quad \text{and} \quad I_5(n) \Rightarrow \beta_5 I_5(n)$$

else if abs(ss(n)) $\leq T_h$, then no change to the integrator outputs.

Variant 1: fixed compression
$\{\beta_1 = \beta_2 = 1$ and $\beta_3 = \beta_4 = \beta_5 = 0.5\}$

Variant 2: variable compression
$\{\beta_1 = \beta_2 = 1$ and $\beta_3 = \beta_4 = \beta_5 = f(\text{ss}(n), T_h, D_p)\}$

where the variable compression function is defined,

$$f(\text{ss}(n), T_h, D_p) = e^{-[|\text{ss}(n)|-T_h]D_p}.$$

Typical parameters found by experiment are $T_h = 20$ and $D_p = 0.1$. In the first variant a simple fixed substitution is made for the integrator outputs when the threshold is exceeded whereas in the second variant the degree of compression is progressive, being controlled by an exponential law. These modified integrator output signals are retained and carried forward to the next step in the loop. If on the next cycle ss(n) still has a magnitude above the threshold, then the compression procedure is again implemented. This repeats until ss(n) falls within the detection window,

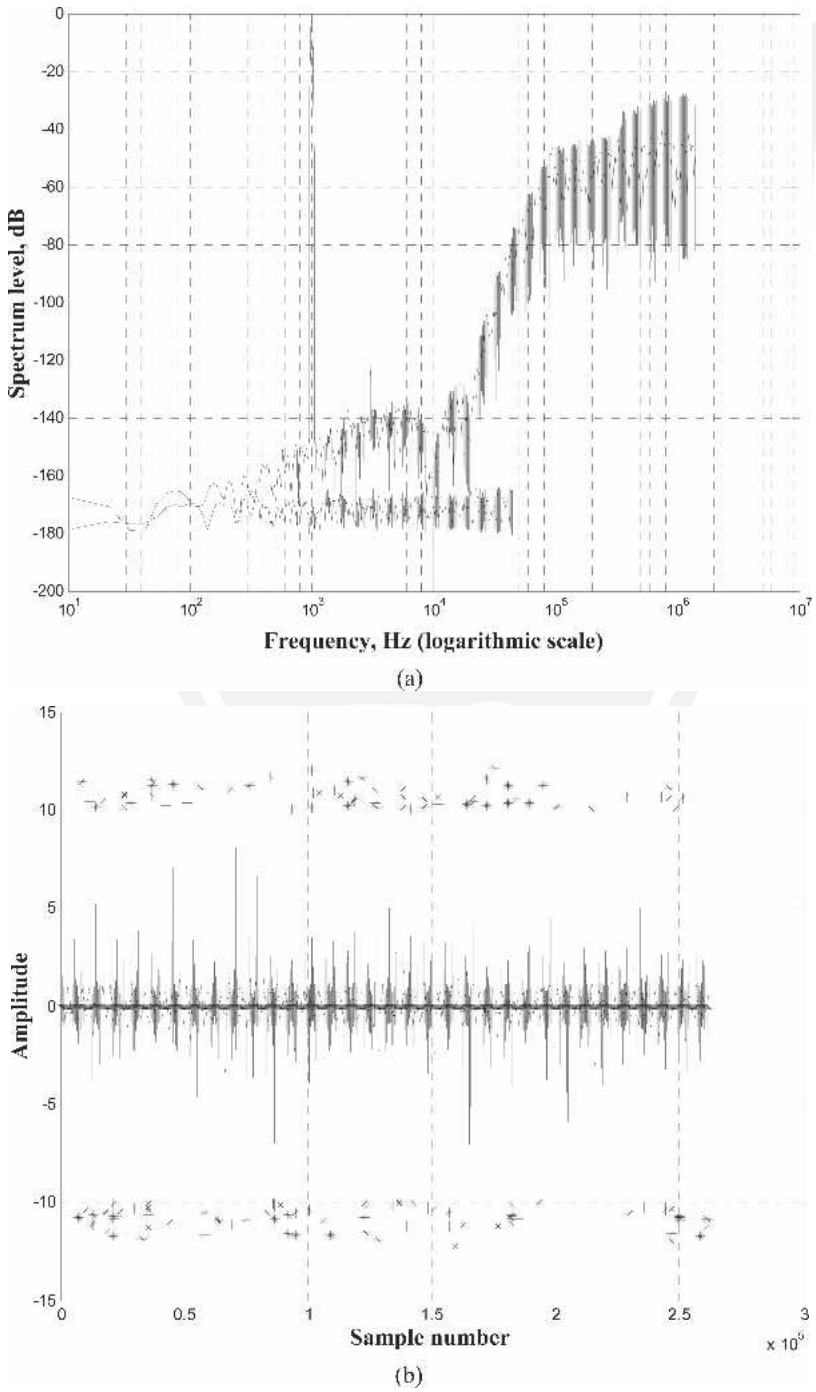
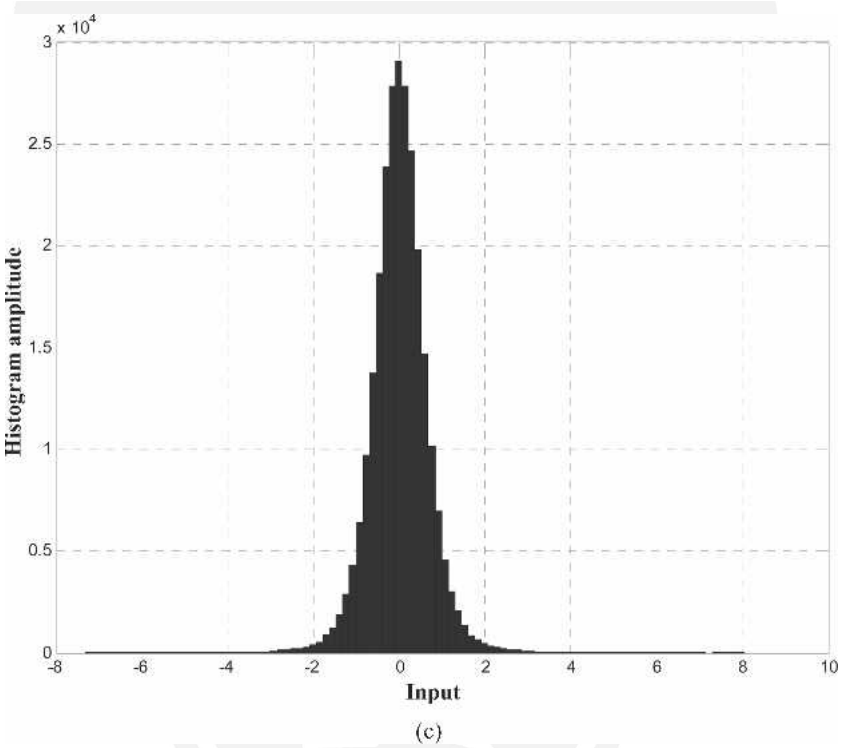


Fig. 33. Sony FF SDM with one-sample look-ahead; input 0.73. (a) Output spectrum. (b) Quantizer input with step back. (c) Quantizer input histogram.

whereupon normal loop behavior is resumed. Critical to the method is that the compression rate of the integrator outputs rises progressively with the magnitude of ss(n) and thus the input level. Also when the process is in a near permanent state of compression, then the first two integrator outputs remain unmodified while the other three are attenuated each loop cycle. Thus the loop tends to second-order behavior, which is known to be unconditionally stable, even with signals up to a modulation depth of unity. The dynamic modification of state variables is included in the SDM program presented in the Appendix. To demonstrate performance, the program was used to compute two example output spectra, shown in Fig. 34 for a 1-kHz input signal of amplitudes 0.78 and 1.0, respectively, where considering the high-input levels excellent noise shaping is revealed.

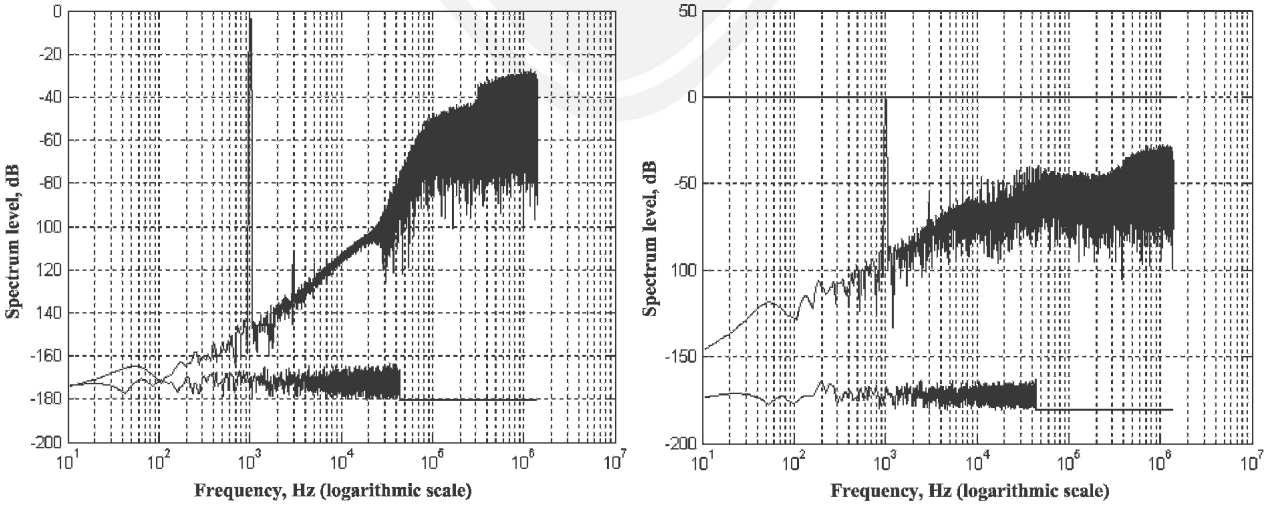
7.2 Review of SDM Power Amplifier Output-Stage Topologies

To conclude this section on switching amplifiers using SDM, three variants of an output stage topology [4] are discussed which reveal that a square-wave output is not mandatory and that reduced switching losses are possible compared to the standard H bridge with a constant voltage supply. The variants designated Types 1 to 3 offer progressive performance improvements in a number of critical areas, with Types 2 and 3 exploiting resonant-mode power supply techniques. The Type 1 stage uses an H-



(c)

Fig. 33. *Continued*



(a) (b)

Fig. 34. Sony FF with look-ahead and dynamic control of state variables. (a) Output spectrum; 0.78 input, 1 kHz. (b) Output spectrum; 1.00 input, 1 kHz.

bridge configuration powered by a constant-voltage power supply where this supply voltage determines the gain of the power amplifier, as changing the supply voltage directly modulates the amplitude of the output waveform [4]. However, Types 2 and 3 are new schemes that employ a resonant-mode power supply to produce a pure sinusoidal output superimposed on a constant-voltage component such that a raised-cosine repetitive waveform V_S of frequency f_{DSD} Hz is presented to the H-bridge stage, where

$$V_S = \text{gain}\left[\frac{1+\cos(2\pi f_{DSD}t)}{2}\right]. \quad (48)$$

The parameter "gain" defines the peak amplitude of the supply voltage and therefore determines the gain of the amplifier. Eq. (48) describes a voltage that swings between zero and "gain" volt, that is, in synchronism with the SDM sample clock. Also, V_S is phase locked so that the zero voltage instants are aligned precisely to the SDM sampling instants, thus virtually eliminating switching losses as all power transistors now switch at zero voltage. The Type 2 output stage is shown in Fig. 30, where the power supply uses the raised-cosine power supply V_S. Consequently the rectangular output pulses of the Type 1 amplifier are replaced with raised-cosine pulses whereby the output pulse stream is effectively prefiltered and has lower spectral content above the SDM sampling rate.

On first encounter it may appear that modifying the pulse waveform will introduce nonlinear distortion. However, this does not occur provided the modified output symbol shape is the same for all pulses in the data stream and where any further waveform changes resulting, for example, from low-pass filtering only affect the sample ensemble as a whole and do not give rise to intersample differences or pulse-sequence-dependent memory effects. To confirm this observation, simulations were performed for both Type 1 and Type 2 amplifier configurations. Normalized time-domain output waveforms are illustrated in Fig. 35 with the corresponding output spectra shown in Fig. 36. Using oversampling techniques, nonrectangular output pulses can be accommodated and output spectra calculated for frequencies in excess of the SDM sampling rate to show how using raised-cosine pulses reduces high-frequency content. Similar spectra to those presented in Figs. 32(a) and 33(a) can be observed with no evidence of additional in-band distortion resulting from waveform

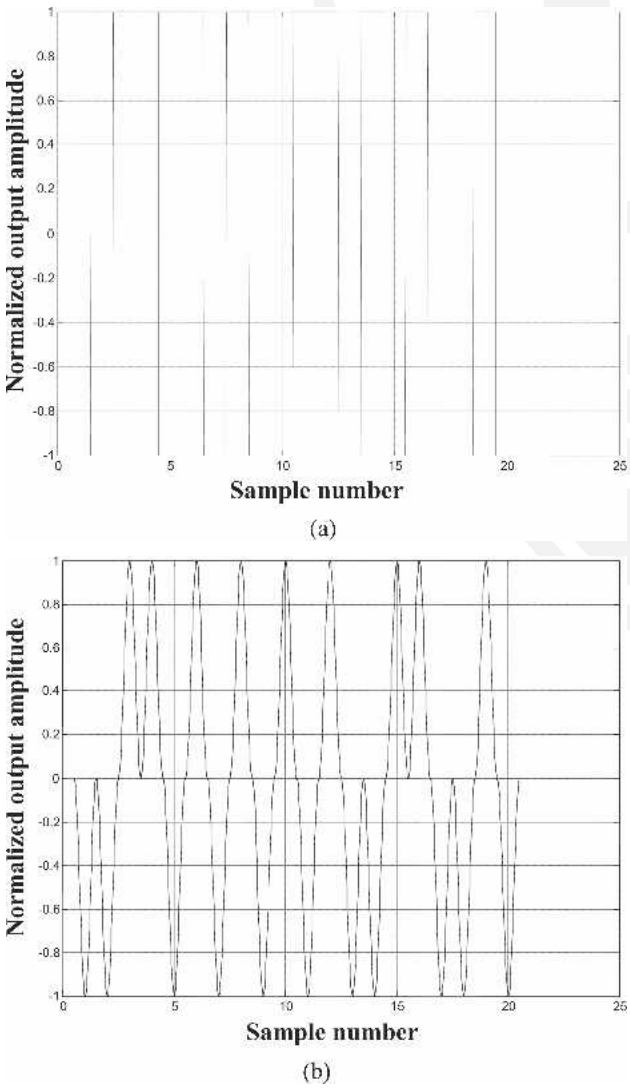


Fig. 35. SDM time-domain outputs. (a) Type 1. (b) Type 2.

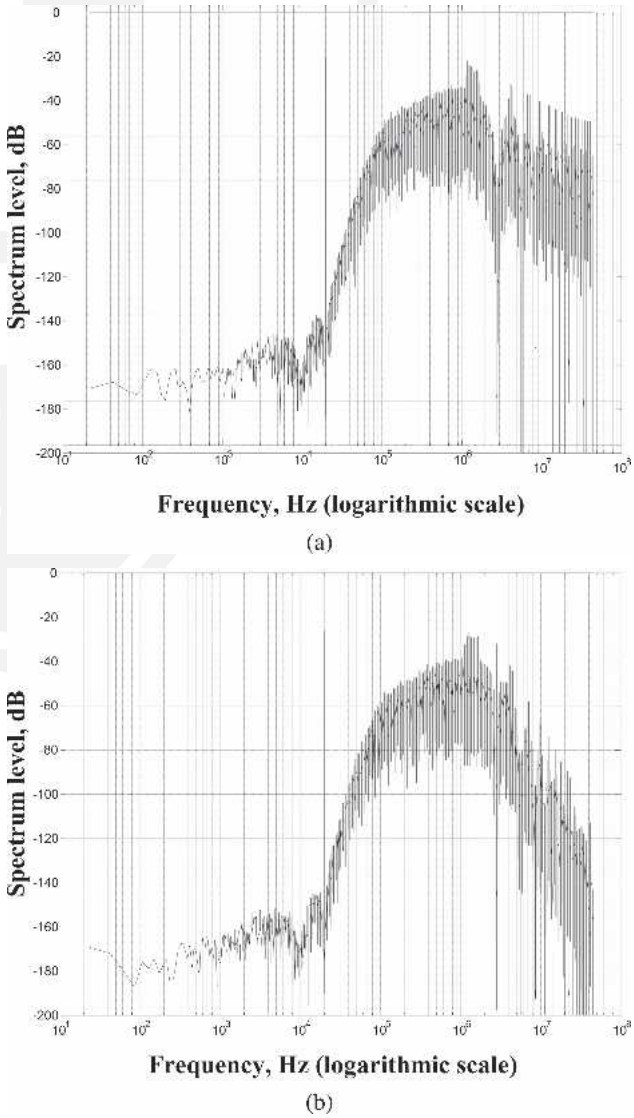


Fig. 36. SDM frequency-domain outputs. (a) Type 1. (b) Type 2.

modification. However, it is evident that significant attenuation of high-frequency components produced by the Type 2 amplifier output spectrum has been achieved. In these simulations a computation vector length of 2^{17} was elected and an additional oversampling factor of 32 applied to the SDM output code in order to accommodate the raised-cosine pulse shape.

Although the Type 2 amplifier topology addresses the problem of switching loss and to some extent alleviates the problem of EMC, it has the disadvantage in that the ratio of the low-pass-filtered output signal amplitude to the peak SDM output voltage is relatively poor compared to a PWM amplifier. This is exacerbated by the fact that even when a sequence of all-1 or all-0 pulses is generated, the differential output signal of the bridge stage always returns to zero between samples, as shown in Fig. 35(b), thus lowering the short-term average of the waveform. To overcome this deficiency a further modification to the amplifier is made and shown conceptually in Fig. 37. A similar raised-cosine power-supply voltage is used, but an additional constant-amplitude voltage source is introduced with its amplitude set precisely to that of the peak value of the raised-cosine waveform. As with the Type 1 and 2 amplifiers these waveforms are controlled by the parameter "gain" as the pulse amplitude modulation method is retained for gain control. Also in addition is a switch that can select either the constant voltage supply or the dynamic power supply. The positions of this switch are defined in Fig. 37 as positions 1 and 2, respectively. Fig. 37 includes an illustration of the relative levels of both the static and the dynamic power supply output waveforms with respect to the control parameter "gain."

The operation of the Type 3 amplifier is as follows: Normally when a change from 1 to 0 or 0 to 1 occurs then switch position 2 is selected and the amplifier operates as a Type 2 amplifier. However, if a burst of all-1 or all-0 pulses occurs in the SDM data stream then the switch changes to position 1 and the amplifier reverts effectively to a Type 1 configuration. This operation then forces the differential output signal of the H bridge to remain constant throughout the period of the burst. An example output sequence is shown in Fig. 38. The effect of this process is to retain constant-amplitude output pulses during a burst of like-valued SDM data. But where a data transition occurs then instead of a rectangular output, the waveform follows a smoothed path determined by the raised-cosine power supply. Consequently Type 3 is a hybrid of Type 1 and 2 amplifiers. Finally, to confirm that the output pulse processing used in the Type 3 amplifier does not introduce additional in-band distortion, a simulation was performed, and the output spectrum is shown in Fig. 39.

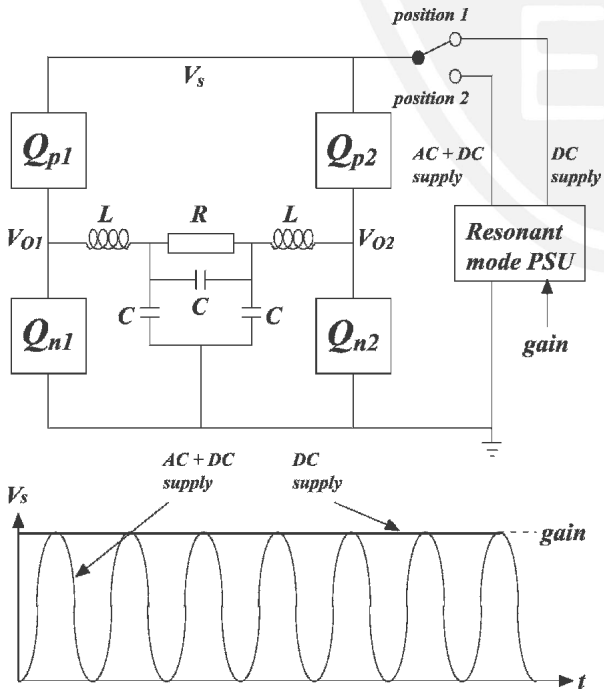


Fig. 37. Type 3 output stage powered by variable-output, resonant-mode power supply with ac and dc supply commutation.

8 CONCLUSIONS

This paper has endeavored to bring together topics on switching amplifiers that are relevant to both SDM and PWM power amplifier systems. It is evident that with the growing interest in SDM this type of modulation has ap-

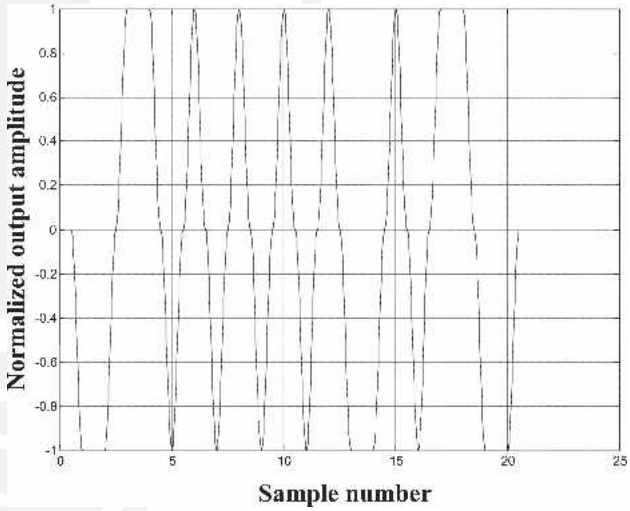


Fig. 38. Type 3 time-domain output.

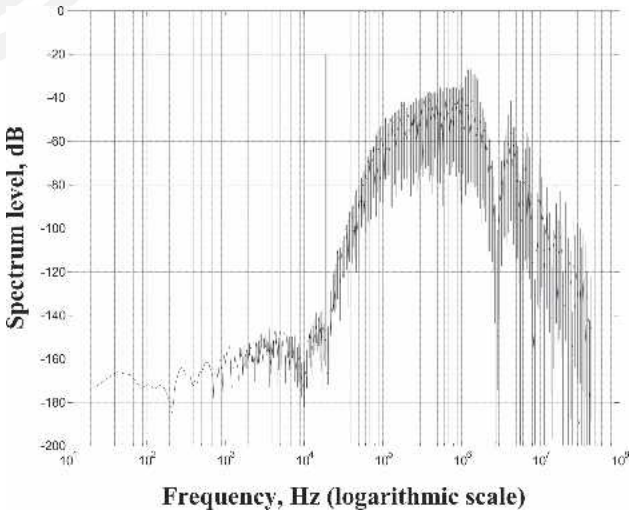


Fig. 39. Type 3 frequency-domain output.

plication, especially as it forms a natural digital system with a sampling rate that is compatible with available digital signal processors. Also, it offers a logical partner for DSD systems, especially as programmable supply voltages can be used to implement power-efficient gain control and thus eliminate the need for additional DSD signal processing.

Theoretical issues of natural sampling were presented, and the linkage between SDM and PWM was discussed for the case without quantization. Here it was shown that both systems may be modeled in terms of linear angle modulation, but the linear process of integration is repositioned from input to output to account for subtleties in waveform construction.

The application of negative feedback to improve PWM linearity was investigated. It was demonstrated by precision simulation that the presence of switching artifacts within the feedback loop causes high-frequency components to appear at the input of the PWM stage, degrading performance, and it was shown that these elements must be suppressed to render the PWM stage linear. As such the use of negative feedback cannot guarantee to reduce all distortion as its presence, without proper corrective procedures, can actually increase distortion as a function of loop gain. Three methods of suppressing switching components within the feedback loop were presented, namely, the NTF, a reference PWM stage within a feedback loop, and predictive correction using an open-loop PWM stage. All methods were shown capable of reducing switching distortion. The discussion of PWM was concluded by adapting the technique of predictive switching compensation for use with digital PWM applications, which also incorporated both feedback and feedforward error correction strategies.

In considering the SDM switched power amplifier it was recognized that robust SDM encoding is critical, especially with the requirement to encode high-amplitude signals. A look-ahead SDM coder was explored, which incorporated an energy-balancing binary decision threshold. This gave robust encryption up to a modulation index of about 0.75. However, by incorporating dynamic compression of the state variables which comes into operation only at high signal levels, stable operation up to the maximum modulation index of unity can be achieved. Simulation results showed that in this high-level operation region, excellent noise-shaping characteristics were retained whereas at lower levels there is no performance compromise. The technique is therefore ideal for power-amplifier applications. In this respect the use of SDM now competes with PWM in terms of peak signal handling. Finally, three configurations for SDM output stages were discussed. These revealed how the H-bridge topology can be adapted for SDM, and especially how switching losses could be lowered by incorporating both a resonant-mode power supply with a raised-cosine output voltage and a constant voltage supply, together with dynamic interpower supply switching related to the SDM code.

REFERENCES

[1] K. Nielsen, "High Fidelity PWM-Based Amplifier Concept for Active Loudspeaker Systems with Very Low Energy Consumption," *J. Audio Eng. Soc.,* vol. 45, pp. 554–570 (1997 July/Aug.).

[2] A. J. Magrath and M. B. Sandler, "Digital Power Amplification Using Sigma–Dalta Modulation and Bit Flipping," *J. Audio Eng. Soc.* (*Engineering Reports*), vol. 45, pp. 476–487 (1997 June).

[3] J. Verbakel, L. van de Kerkhof, M. Maeda, and Y. Inazawa, "Super Audio CD Format," presented at the 104th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 46, p. 570 (1998 June), preprint 4705.

[4] F. M. Prime and M. O. J. Hawksford, "Digital Audio Power Amplifier for DSD Data Streams," presented at the 117th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 53, p. 118 (2005 Jan./Feb.), convention paper 6303.

[5] A. C. Floros and J. N. Mourjopoulos, "Analytical Derivation of Audio PWM Signals and Spectra," *J. Audio Eng. Soc.,* vol. 46, pp. 621–633 (1998 July/Aug.).

[6] J. E. Flood and M. J. Hawksford, "Exact Model for Deltamodulation Processes," *Proc. IEE,* vol. 118, pp. 1155–1161 (1971).

[7] M. J. Hawksford, "Unified Theory of Digital Modulation," *Proc. IEE,* vol. 121, pp. 109–115 (1974 Feb.).

[8] M. J. Hawksford, "Application of Delta-Modulation to Television Systems," Ph.D. Thesis, University of Aston, Birmingham (1972).

[9] M. O. J. Hawksford, "Time-Quantized Frequency Modulation, Time-Domain Dither, Dispersive Codes, and Parametrically Controlled Noise Shaping in SDM," *J. Audio Eng. Soc.,* vol. 52, pp. 587–617 (2004 June).

[10] M. O. J. Hawksford, "Dynamic Model-Based Linearization of Quantized Pulse-Width Modulation for Applications in Digital-to-Analog Conversion and Digital Power Amplifier Systems," *J. Audio Eng. Soc.,* vol. 40, pp. 235–252 (1992 Apr.).

[11] M. O. J. Hawksford, "Linearization of Multi-Level, Multi-Width Digital PWM with Applications in Digital-to-Analog Conversion," *J. Audio Eng. Soc.,* vol. 43, pp. 787–798 (1995 Oct.).

[12] D. J. Naus, E. C. Dijkmans, E. F. Stikvoort, A. J. McKnight, D. J. Holland, and W. Bradinal, "A CMOS Stereo 16-bit Converter for Digital Audio," *IEEE J. Solid-State Circuits,* vol. SC-22 (1987 June).

[13] R. H. Small, "Constant-Voltage Crossover Network Design," *J. Audio Eng. Soc.,* vol. 19, pp. 12–19 (1971 Jan.).

[14] W. M. Leach Jr., "Loudspeaker Driver Phase Response: The Neglected Factor in Crossover Network Design," *J. Audio Eng. Soc.,* vol. 28, pp. 410–421 (1980 June).

[15] M. O. J. Hawksford, "System Measurement and Identification Using Pseudorandom Filtered Noise and Music Sequences," *J. Audio Eng. Soc.,* vol. 52, pp. 275–296 (2005 Apr.).

[16] L. Risbo, "Σ–Δ Modulators—Stability Analysis and Optimization," Ph.D. Thesis, Technical University of Denmark (1994 June).

[17] H. Takahashi and A. Nishio, "Investigation of Practical 1-bit Delta–Sigma Conversion for Professional

Audio Applications," presented at the 110th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 49, p. 544 (2001 June), convention paper 5392.

[18] J. A. S. Angus, "Tree Based Lookahead Sigma Delta Modulators," presented at the 114th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 51, p. 435 (2003 May), convention paper 5825.

[19] M. O. J. Hawksford, "Parametrically Controlled Noise Shaping in Variable State-Step-Back Pseudo-Trellis SDM," *Proc. IEE—VIS. Image Signal Process.,* vol. 152, pp. 87–96 (2005 Feb.).

APPENDIX
SONY FF MATLAB ENCODER WITH TWO-SAMPLE LOOK-AHEAD WITH IMPROVED STABILITY

```
% 5th-order Sony ff with 2-sample look-ahead
configured for digital power amplifier
% AESDM
% 31.12.05
% 2-sample look-ahead with minimum energy state
detection
% dynamic state variable compression to improve
stability
% based on 5th order Sony FF SDM
% fsdm: SDM sampling rate
% fin1, fin2: input frequency of sinusoidal components
% a1, a2: input amplitude of sinusoidal components
% input quantized to "bit" resolution reference Nyquist
sampling, then oversampled
% choice of output symbol, sigma sets samples per
symbol
% NN-tap FIR output filter
close; home; clear; colordef white
fprintf('5th-order Sony FF with look-ahead configured
for digital power amplification with output symbol
options\n')

% set output pulse symbol shape
group = 3;
while group>2
fprintf('\nSelect type of output pulse symbol:\n')
fprintf('group = 0 standard rectangular symbol\n')
fprintf('group = 1 raised-cosine symbol\n')
fprintf('group = 2 raised-cosine symbol with hold
applied between pulse groups\n')
group = input('Pulse symbol type: ');
end

% set input signal amplitude and frequency (2
sinewaves)
a1 = .4; a2 = .4; % input signal level
fin1 = 19000; fin2 = 2000; % input frequencies

% set data
fprintf('\nLoad input data\n')
sigma = 2^5; % samples per symbol of SDM output
code (suggest 32)
NN = 32; % number of taps in output filter (32)
over = 64; % SDM oversample ratio above Nyquist (64
for DSD)
order = 5; % SDM order (5 for Sony FF)
L = 2^18; % computational vector length
bit = 24; % input signal quantization (24 bit)
nyquist = 44100; % Nyquist sampling rate (44100 Hz)
ditheramp = .35; % select dither amplitude (0.35)
thresh = 20; % dynamic state variables threshold (20)
damp = 10; % dynamic state variables damping (10)

% calculate constants, Blackman window and dither
fprintf('Calculate constants, window and dither\n\n')
fsdm = over*nyquist;
Lx = L/over; Ls = L*sigma;
f0 = nyquist/Lx;
w1 = round(fin1/f0)*2*pi/Lx; w2 = round(fin2/f0)*2*
pi/Lx;
dither = 2*ditheramp*(rand(1,L)-1);
bl = blackman(Ls)'; bl = bl/mean(bl); % calculated over
Ls samples
sdb = zeros(1,L);

% quantize input at 44.1 kHz
axx = a1*sin(w1*(1:Lx))+a2*sin(w2*(1:Lx));
axx = round(axx*2^(bit-1)+rand(1,Lx)+rand(1,Lx)-1)/
2^(bit-1);

% interpolate input to a sampling rate of fsdm
ax = zeros(1,L);
for x = 1:Lx
ax((x-1)*over+1:x*over) = over*[axx(x) zeros(1,over-1)];
end
clear axx
win = [ones(1,Lx/2) zeros(1,(L-Lx)/2)];
win = [win 0 win(L/2:-1:2)];
ax = real(ifft(fft(ax).*win));

% set SDM initial conditions and constants
sdb = zeros(1,L); aa = 2*ditheramp;
I = zeros(1,5); ss = sdb;

% run Sony FF SDM loop with look-ahead
fprintf('Run Sony FF SDM: 2-sample look-ahead +
dynamic state variable modulation\n\n')
b1 = 1; b2 = .5; b3 = .25; b4 = .125; b5 = .0625;
c2 = -.001953125; c4 = -.03125;
% dither sequence
rd = aa*(rand(1,L)-rand(1,L));
for n = 2:L-1
% update integrators
I(1) = I(1)+ax(n-1)-sdb(n-1);
I(2) = I(2)+b2*I(1)+c2*I(3);
I(3) = I(3)+b3*I(2);
I(4) = I(4)+b4*I(3)+c4*I(5);
I(5) = I(5)+b5*I(4);
% look-ahead 1 sample D(n) = 1
I1x = I(1)+ax(n)-1;
I2x = I(2)+b2*I1x+c2*I(3);
I3x = I(3)+b3*I2x;
I4x = I(4)+b4*I3x+c4*I(5);
I5x = I(5)+b5*I4x;
```

```
% look-ahead 2 sample D(n)=1 D(n+1)=1
I1x1=I1x+ax(n+1)-1;
I2x1=I2x+b2*I1x+c2*I3x;
I3x1=I3x+b3*I2x;
I4x1=I4x+b4*I3x+c4*I5x;
I5x1=I5x+b5*I4x;
% look-ahead 2 sample D(n)=1 D(n+1)=-1
I1x0=I1x+ax(n+1)+1;
I2xo=I2x+b2*I1x+c2*I3x;
I3x0=I3x+b3*I2x;
I4x0=I4x+b4*I3x+c4*I5x;
I5x0=I5x+b5*I4x;
% look-ahead 1 sample D(n)=-1
I1y=I(1)+ax(n)+1;
I2y=I(2)+b2*I1y+c2*I(3);
I3y=I(3)+b3*I2y;
I4y=I(4)+b4*I3y+c4*I(5);
I5y=I(5)+b5*I4y;
% look-ahead 2 sample D(n)=1 D(n+1)=1
I1y1=I1y+ax(n+1)-1;
I2y1=I2y+b2*I1y+c2*I3y;
I3y1=I3y+b3*I2y;
I4y1=I4y+b4*I3y+c4*I5y;
I5y1=I5y+b5*I4y;
% look-ahead 2 sample D(n)=1 D(n+1)=-1
Ily0=I1y+ax(n+1)+1;
I2y0=I2y+b2*I1y+c2*I3y;
I3y0=I3y+b3*I2y;
I4y0=I4y+b4*I3y+c4*I5y;
I5y0=I5y+b5*I4y;
ss(n)=sum(I(1:5))*(sdb(n)+1)-sum(I(1:5))*(sdb(n)-1); %
calculate input to quantizer for sdb(n)=1 or -1
if abs(ss(n))>thresh % attenuate state variables when
threshold exceeded
%I(3:5)=[.5 .5 .5].*I(3:5);
I(3:5)=I(3:5)*exp((-abs(ss(n))+thresh)/damp);
end
mx0=((I1x0+I2x0+I3x0+I4x0+I5x0)^2+I1x0^2+I2x0^2
+I3x0^2+I4x0^2);
mx1=((I1x1+I2x1+I3x1+I4x1+I5x1)^2+I1x1^2+I2x1^2
+I3x1^2+I4x1^2);
my0=((I1y0+I2y0+I3y0+I4y0+I5y0)^2+I1y0^2+I2y0^2
+I3y0^2+I4y0^2);
my1=((I1y1+I2y1+I3y1+I4y1+I5y1)^2+I1y1^2+I2y1^2
+I3y1^2+I4y1^2);
sdb(n)=sign(rd(n)+min([my0 my1])-min([mx0 mx1]));
% 2 step look-ahead minimum energy
n=n+1; % increment sample value
end % SDM loop end ****************************
******************
plot(ss,'k')
title('Signal ss(n) applied to standard quantizer')
ylabel('Amplitude')
xlabel('Time')
grid; pause; close

% shape output pulses for digital amplifier application
using oversampling by sigma
sdbrc=zeros(1,Ls);
if group>0
fprintf('Shape SDM output pulses: raised cosine\n')
symb=(1-cos(2*pi*(0:sigma-1)/sigma))/2;
else
fprintf('Shape SDM output pulses: rectangular pulse
shape\n')
symb=ones(1,sigma);
end
for x=1:L
ss=(x-1)*sigma+1;
sdbrc(ss:ss+sigma-1)=sdb(x)*symb(1:sigma);
end

% detect groups of 2 or more pulses and hold pulse
amplitude at maximum
if group==2
fprintf('Shape SDM output pulses with hold function for
pulse groups\n')
for x=1:L-1
ss=(x-1)*sigma+1;
if sdb(x)-sdb(x+1)==0
sdbrc(ss+.5*sigma:ss+1.5*sigma-1)=sdb(x)*ones(1,
sigma);
%plot(sdbrc(ss:ss+2*sigma)); pause
end; end; end

% FFT routine
fprintf('Calculate output spectrum\n')
sdbf=abs(fft(sdbrc(1:Ls).*bl));
sdbf(1:Ls/2-1)=20*log10(10^-10+2*sdbf(2:Ls/2)/Ls);

% plot spectrum SDM to fsdm/2
fprintf('Plot SDM spectrum\n')
semilogx(f0*(1:L/2-1),zeros(1,L/2-1),'k')
hold
semilogx(f0*(1:L/2-1),sdbf(1:L/2-1),'k')
title('SDM output spectrum to fsdm/2')
xlabel ('Frequency, Hz (logarithmic scale)')
ylabel('Spectrum level, dB')
grid; pause; close

% plot spectrum SDM full
fprintf('Plot SDM spectrum\n')
semilogx(f0*(1:Ls/2-1),zeros(1,Ls/2-1),'k')
hold
semilogx(f0*(1:Ls/2-1),sdbf(1:Ls/2-1),'k')
title('SDM output spectrum to sigma*fsdm/2')
xlabel('Frequency, Hz (logarithmic scale)')
ylabel('Spectrum level, dB')
grid; pause; close

% plot example output time domain
kk=20;
plot((1:kk*sigma)/sigma+.5,sdbrc(1:kk*sigma),'k')
title('SDM output sequence')
xlabel('Sample number')
ylabel('Normalized output amplitude')
grid; pause; close

% NN-tap filtered output time domain plot
kk=Lx;
sdbrcav=sdbrc(1:kk*sigma);
for x=1:NN-1
```

```
sdbrcav = sdbrcav+sdbrc(1+x*sigma:(kk+x)*sigma);
end
sdbrcav = sdbrcav/NN;
plot((1:kk*sigma)/sigma+.5,sdbrcav(1:kk*sigma),'k')
title('Normalized filtered output over NN DSD
samples')
xlabel('Sample number')
ylabel('Normalized output amplitude')
grid; pause; close

% end of program
fprintf('\nProgram terminated\n')
return
% **********************************
```

THE AUTHOR

Malcolm Hawksford received a B.Sc. degree with First Class Honors in 1968 and a Ph.D. degree in 1972, both from the University of Aston in Birmingham, UK. His Ph.D. research program was sponsored by a BBC Research Scholarship and he studied delta modulation and sigma-delta modulation (SDM) for color television applications. During this period he also invented a digital time-compression/time-multiplex technique for combining luminance and chrominance signals, a forerunner of the MAC/DMAC video system.

Dr. Hawksford is director of the Centre for Audio Research and Engineering and a professor in the Department of Electronic Systems Engineering at Essex University, Colchester, UK, where his research and teaching interests include audio engineering, electronic circuit design, and signal processing. His research encompasses both analog and digital systems, with a strong emphasis on audio systems including signal processing and loudspeaker technology. Since 1982 his research into digital crossover networks and equalization for loudspeakers has resulted in an advanced digital and active loudspeaker system being designed at Essex University. The first one (believed to be unique at the time) was developed in 1986 for a prototype system to be demonstrated at the Canon Research Centre and was sponsored by a research contract from Canon. Much of this work has appeared in *JAES,* together with a substantial number of contributions at AES conventions. He is a recipient of the AES Publications Award for his paper, "Digital Signal Processing Tools for Loudspeaker Evaluation and Discrete-Time Crossover Design," for the best contribution by an author of any age for *JAES,* volumes 45 and 46.

Dr. Hawksford's research has encompassed oversampling and noise-shaping techniques applied to analog-to-digital and digital-to-analog conversion with special emphasis on SDM and its application to SACD technology. In addition, his research has included the linearization of PWM encoders, diffuse loudspeaker technology, array loudspeaker systems, and three-dimensional spatial audio and telepresence including scalable multichannel sound reproduction.

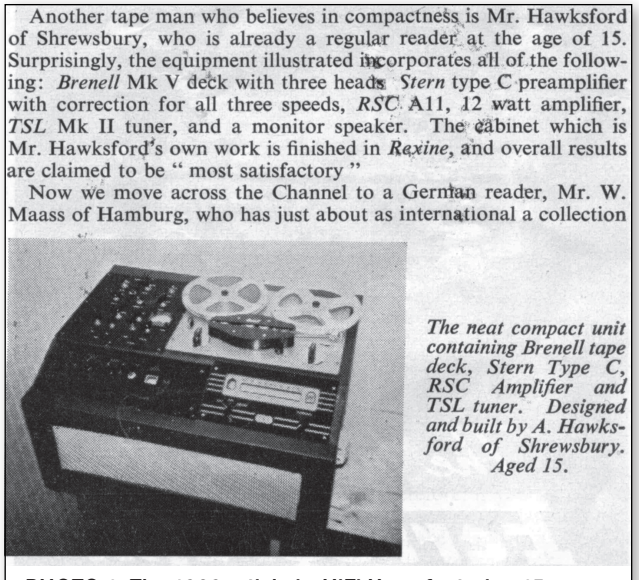
Dr. Hawsford is a chartered engineer and a fellow of the AES, IEE, and IOA. He is currently chair of the AES Technical Committee on High-Resolution Audio and is a founder member of the Acoustic Renaissance for Audio (ARA).

The Essex Echo:
Audio According to Hawksford, Pt. 1

Malcolm Hawksford had his first design, a console tape recorder with pre- and power amp and build-in loudspeakers, published in the prestigious *HiFi News* in June 1963 at age 15. It was an early milestone in a life and career marked by contributions to advance the state-of-the-art in audio. It led up to the Audio Engineering Society's Silver Medal Award "for major contributions to engineering research in the advancement of audio reproduction" in 2006, and a Doctor of Science degree for lifetime research achievements from Aston University (UK) in 2008. Jan Didden visited him at his lab at Essex University and spent many fascinating hours talking about audio technology.

Jan Didden (JD): Professor Hawksford, you published your tape recorder design at the age of 15 (**Photo 1**). When did you start to become interested in audio, and why?

Malcolm Hawksford (MH): Oh, I think it started when I was 8. You see, my father, and indeed his father before him back in the 1920s, were interested in music reproduction; my father had this 50s Grundig tape recorder, type TK820-3D I think it was. I was also interested in making puppet theaters, especially the lighting, and used to build control systems by winding variable resistors on wooden dowels, batten lighting, and spot lights using baking powder tins. It taught me a lot about lamps in series and parallel and the losses that could occur in connecting wires (smile). I became fascinated by those wires and lights and how they worked. So I got to play with tape recorders and circuits and also obtained electrical parts and old TVs from my father to experiment with, as at that time he worked for an electrical wholesaler.



Another tape man who believes in compactness is Mr. Hawksford of Shrewsbury, who is already a regular reader at the age of 15. Surprisingly, the equipment illustrated incorporates all of the following: *Brenell* Mk V deck with three heads *Stern* type C preamplifier with correction for all three speeds, *RSC* A11, 12 watt amplifier, *TSL* Mk II tuner, and a monitor speaker. The cabinet which is Mr. Hawksford's own work is finished in *Rexine*, and overall results are claimed to be " most satisfactory "

Now we move across the Channel to a German reader, Mr. W. Maass of Hamburg, who has just about as international a collection

The neat compact unit containing Brenell tape deck, Stern Type C, RSC Amplifier and TSL tuner. Designed and built by A. Hawksford of Shrewsbury. Aged 15.

PHOTO 1: The 1963 article in *HiFi News* featuring 15-year-old Malcolm's integrated tape machine.

I remember my first tape recorder didn't want to record: I had measured its recording head resistance with a multimeter, and, unknown to me, the head had become heavily magnetized. But I eventually figured it out, and, after degaussing the head (care of a Ferrograph de-gausser), it actually worked quite well. I was also very fortunate that my mum didn't really object when I turned my bedroom into a lab and the carpet developed a silver sheen of solder!

JD: The 60s were a great time for a youngster interested in electronics anyway.

MH: Absolutely. You could get parts and kits for almost any purpose. After building a few kits, I quickly started to develop my own circuits and topologies. I still enjoy circuit design. The way it works for me is that I sort of juggle or model the circuit in my mind, actually visualizing the circuit and various voltages and currents, before I put it on paper.

I also was lucky that by the time I was ready to go to university in 1965, the University of Aston in Birmingham offered one of the first electrical engineering courses that focused on what was then called "light-current electronics," as opposed to power electronics. When I graduated at the end of my third year with a first-class honors B.Sc., it was suggested that I stay on for research. I then applied, and received, a BBC Research Scholarship and did my thesis on the "Application of Delta-Modulation to Colour Television Systems" (available on my website).

Professor Hawksford has published several articles for the audio press under the cover title "The Essex Echo." The Echo was a local Essex newspaper published between 1887 and 1918.

Using the (then new) emitter-coupled logic from Motorola, I was able to get up to 100MHz clocks, and that was in 1968 mind you. This logic family had to be interconnected using transmission-line techniques with proper termination to prevent reflections! This choice of subject proved rather fortuitous, as it gave me a strong grounding in delta-modulation and its close relation sigma-delta modulation, technology that was later to have a massive impact on audio systems in the 1980s and 90s.

JD: Do you see circuit design as an art?
MH: Yes, I think it is to some extend an art, or a bridge between science and art, in the sense that you develop a "pictorial" solution without knowing exactly how you got there. It sort of develops itself. I have been doing circuit design most of my life and it has become a "sixth sense"; I'm thinking in circuit blocks, sort of. In those early days you would try out different topologies, thinking it through, and trying to picture the currents and voltages in your mind while trying to get to the optimal solution.

JD: Is there a personal style in circuit design? Is there a "Hawksford" style in circuit design?
MH: To a certain extent I think there is. Designers usually solve a circuit problem slightly different from each other, perhaps based on how they learned to solve certain problems earlier and probably also depending on their personality. If you are a digital designer, you might choose to plug some design spec into a program that puts it in an FPGA for you. Likewise, as an IC designer you may have a library of standard cells or modules that you can use to lay out your chip.

In each of these cases the designer seems a step or two removed from the detailed design, making it more anonymous, unlike an analog discrete circuit designer. That said, I think that also sometimes circuits are designed differently for other reasons than you might think. I firmly believe that if you design an amplifier, and you take care of both the critical factors and secondary effects, such designs will tend to "sound" very similar. . . hopefully implying the performance is accurate.

Now, of course, the topology isn't all of it. People often become preoccupied with topology, but there are many more issues required to make a circuit into a great piece of equipment. There's the power supply, the grounding layout, EMI issues, the quality of the components, the wire used—they all contribute to the final result. So, when you get the topology right—that is, get it to converge in terms of stability and linearity and such—then the secondary factors become important.

Let me give you an example. Most designers are aware that you must avoid sharing supply return paths between power and signal returns. The power return current could cause a "dirty" voltage across the return path that couples into the signal circuit. Even if you use a series supply regulator, you can still have this problem with a rock-stable and clean supply voltage, because the harm is done through the return current.

Now, if you use a shunt regulator, the "dirty" current can be localized and kept from signal returns, and that offers a major advantage. If it still isn't enough, you can use what is called an "active ground" or "dustbin" where the supply return current is not returned to the ground common at all but disappears into another, separate supply system (**Fig. 1A-C**).

JD: What is your view on the desirability and usefulness of blind testing to rate the performance of audio equipment?
MH: Well, I think you must use some kind of objective form of subjective testing method to isolate differences between components. Many people do not realize that they have a sort of internal perceptual model that determines how they perceive the auditive input. That internal perceptual model not only takes into account the sound feed from your ears into your brain, but also how you feel, your expectations, how bright is the environment, how relaxed you are, and many other factors. So if your internal perceptual model changes due to those other interference factors changing, your perception of sound can change.

I recall occasions where initially I perceived a certain difference between cables, and then I repeated it the next

day and my perception was often quite different. I think this was due, at least in part, to my changed internal perceptual model. So some kind of objective test is required, but that said, I'm not a strong advocate of ABX-style double-blind testing (DBT).

The limitation in sensing a change in sound can put us in an unnatural situation, and I'm not sure we then function so reliably or sensitively. Possibly a better approach is a blind method that allows a relaxed and holistic type of listening session. Of course, with DBT it's easy to get a null result, so it may be a good method if that is your agenda.

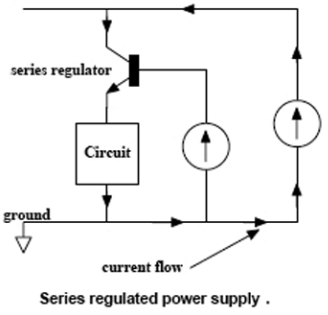


FIGURE 1A: Series regulator return current flow.

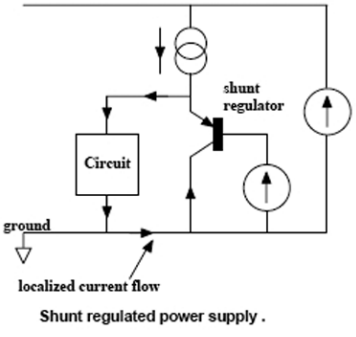


FIGURE 1B: Shunt regulator localizes return current.

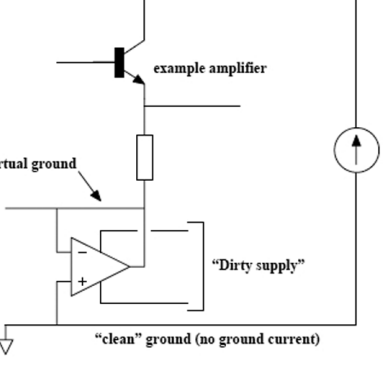


FIGURE 1C: "Active ground" dumps return currents in another supply.

My preference is to undertake listening tests in a completely darkened room. The fact that the equipment you listen to isn't hidden and could be identified with just a bit more light makes it much more natural and less stressful than being aware that the equipment is purposefully hidden from you. Also, being able to focus your senses purely on sound and not be distracted by uncorrelated visual input to the brain heightens your auditory perception. It is very easy to do and increases your sensitivity and acuity, especially in spatial terms.

In my experience it is not the same as closing your eyes. It seems that when you close your eyes when listening, you are sort of fooling yourself; it's artificial in a way and it still diverts some mental processing power away from your listening. You should try the dark room sometimes, although it's good to keep a small torch at your side!

JD: Another method correlating measurements with perception that gets some attention lately is trying to extract the difference between the "ideal" signal and the actual signal. In the past year I attended several AES presentations on systems to extract those differences and make them audible, such as the differences between unprocessed music and the MP3 version. Bill Waslo of Liberty Instruments, the makers of the Praxis measurement suite, even has a free version online (AudioDiffmaker).

MH: It is a very powerful technique which I explored formally in 2005[1], and we have employed the extraction of error signals over many years at Essex (see, for example, "Unification" articles on my website). The idea is that you have a system with both a target function (the design response) and the actual function with imperfections, so you can then represent the actual system in terms of the target function and an "error function." There's a lot to it, but as a simple example consider the frequency response of a high-quality CD player. You can assume that the target response here is a flat response to around 20kHz; if that is not the case, then, of course, you need to correct for the nonlinear target response in the extraction of the error.

If, for example, there are response irregularities below, say, -40dB, it would give around 0.01dB frequency response ripple. You barely see that in the frequency response, as it's actually less than the graph line thickness! However, if you assume a flat target response, extract the error and then plot it on the same graph, which tells you much more.

Figure 2 shows the minute ripples in the response resulting from an imperfect DAC reconstitution filter. This tells you how far below the main signal you have some kind of "grunge" in the system, where ideally it should be below the noise floor. I like this type of presentation because it can inform you of the actual low-level error resulting from small system imperfections (both linear and nonlinear) that may cause audible degradation. This frequency response example is relatively simple, but in the paper[1] I give some examples of using MLS or even music signals to extract the low-level errors from ADCs and DACs.

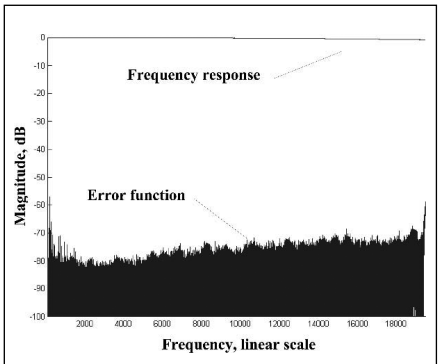


FIGURE 2: Example error graph for CD frequency response.

JD: If you can extract the error, can you then not compensate for it? Sort of "pre-distorting" the signal with the inverse of the error function? Possibly digitally?

MH: Well, compensating analog systems with numbers becomes complicated pretty fast. Most of these errors are dynamic or may arise from some interference of some kind, and although you can measure them accurately, you cannot predict them to any accuracy. The errors vary a lot with time and temperature and what have you.

It's been tried with loudspeakers, where you can develop a Volterra-based model to describe cone motion, for in-

stance, and use inverse processing to linearize it. But it is extremely difficult to keep the compensation model synchronized to the instantaneous cone position and movement. If you're just a little bit off, the results may be worse than without correction. It's much easier and cheaper to design a better driver!

There are some other techniques. There's a guy called David Bird, who used to work for the BBC and was using a current drive technique. One of my ex-Ph.D. research students, Paul Mills (who is now responsible for loudspeaker development at Tannoy), and I have also done some work on that subject. With current drive, the principal error is, in fact, the deviation of the $B\ell$ -product of the driver.

You can therefore measure the $B\ell$ deviation as a function of cone displacement, and if you then monitor the cone position, you can apply inverse $B\ell$ correction such that the force on the cone is proportional to the input current. We actually developed a transconductance power amplifier to current drive a loudspeaker, with several error correction techniques included in the design[2]. We solved the low-frequency damping problem in two different ways. One was to use an equalizer; you measure the hi-Q resonance and then preprocess the signal to obtain the required linear response.

The other approach was to wind a thin wire secondary coil onto the voice-coil former of the drive unit, just voltage sensing, and to process that signal and feed it back into the transconductance amplifier. There was some unwanted transformer coupling from the main voice coil into the sensing coil which we had to compensate for with a filter. But since the main coil was current driven, it didn't matter if it heated up, and since there was no current flowing through the sensing coil, it also did not matter if it heated up. It worked very well; I remember that even using current drive with a tweeter also significantly lowered distortion.

JD: It wouldn't help with things like cone breakup.

MH: No, it wouldn't. And it adds an extra layer of complexity and things that can go wrong. It is also only suitable for active loudspeaker systems.

JD: One issue that turns up in your work again and again has to do with jitter in some form.

MH: Well, yes, because it turned into an issue after we got the CD from Philips and Sony, and after the first euphoric reports, many people realized that what should have sounded perfect didn't. A major cause was jitter, which hadn't really been considered in those early years, probably because jitter is an "analog aspect" of a digital system. It can also manifest itself in different ways; it can disguise itself like noise (random and relatively benign) or as a periodic disturbance related to power supply ripple or clock signals, which is more objectionable, or it can be correlated with the audio signal, which also can sound quite bad.

So just saying "jitter" is not enough; its effect depends very much on how it manifests itself. In fact, I produced a paper at one time in which I designed a jitter simulator that allowed one to compute specific amounts and type of jitter, noise or periodic or correlated to the signal, and add that to the clean signal so you could listen to its effect (*see sidebar Hawksford on the Sound of Jitter*).

You can debate its significance, but at least you can point to a measureable and audible defect, whereas a traditional jitter picture with sidebands and what have you doesn't give you a "feel" for what it sounds like.

I did a study with research student Chris Dunn[3] (not the Chris Dunn who has published substantial work on jitter) which showed that the jitter introduced by the AES/EBU (or S/PDIF) interface protocol even depends on the bit pattern—in other words, on the music signal itself.

For instance, when you listen to the error signal of the phase-lock loop (PLL) on the digital receiver, you can actually hear the music signal that was transmitted through that digital link! It is distorted, of course, but this was clearly an example of music-correlated jitter.

Now there are known engineering solutions to eliminate that jitter later on,

but it doesn't always happen in equipment, so there is the possibility when you transmit digital audio through a band limited link (and it is always band limited), you can get correlated jitter just from that process. We also showed that if you code the L and R signals separately, invert one of them, and then send both over that interface, almost all of that signal-related jitter would disappear. But it wasn't picked up on; such is the law of standards!

JD: How would you design the "ideal" DAC?

MH: The DAC chips themselves nowadays are very good indeed. Where you see the differentiation in quality is in stages like the I/V converter, a seemingly innocent subject. The sharp switching edges from the DAC output can only be perfectly reproduced with an I/V op amp that has infinite bandwidth and no limit on slew-rate. Any practical circuit will have nonlinearity and slew rate limits such that a transient input signal can slightly modulate the open-loop (OL) transfer of the op amp.

Modulating the OL transfer function means you modulate the circuit's closed-loop (CL) phase shift. What is interesting is that it looks remarkably similar to correlated jitter; they share a family resemblance (**Fig. 3**). It also is similar to what people have been talking about as dynamic-phase modulation in amplifiers. Whenever an amplifier stage needs to respond very quickly, it tends to run closer to open loop and therefore is more susceptible to open loop nonlinearity. So the I/V stage is clearly a critical stage, and although the underlying processes are different, the resulting signal defects may manifest themselves as correlated jitter, especially as the timing errors occur close to the sampling instants where signal rate-of-change is maximum.

Anyway, I really think we should not talk about phase modulation here, as that is more appropriate for sine wave signals and linear systems. We should talk about temporal modulation instead. There are many ways you can solve these issues once you understand them, possibly to design your I/V converter to be very wide band, or using a very linear open-loop circuit, or maybe some low-pass filtering between DAC and I/V stage. What you end up with[4] is a discrete current-steering circuit that runs partially open loop and integrates the I/V conversion and low-pass filtering into one circuit, rather than bolting an I/V stage to a subsequent second-order filter as is normally done. Consequently, you minimize the active circuitry involved (**Fig. 4**).

If you think about it, theoretically we are trying to make circuitry work flawlessly up to infinitely high frequency, which in principle cannot be reached. So at one time I thought maybe we need a totally different way to solve the problem of critical timing issues in DACs.

One possible solution I came up with was to modulate the reference voltage of an R-2R ladder DAC with a synchronized raised-cosine waveform. Rather than the DAC output staircase signal jumping "infinitely" fast to a new level at every clock pulse, it effectively made the new level the same as the previous and then ramped it up, so to speak, to the new level using a raised-cosine shape with the same period as the clock (**Fig. 5A**, **B**). Consequently, adjacent samples were linked by raised-cosine interpolation rather than a rectangular step function. This also helps a little with signal-recovery filtering. So, the rate-of-change of the currents coming from the DAC was dramatically reduced. I built a prototype to proof the principle and it dramatically reduced the timing and transient errors in the I/V stage.

In many ways I view I/V conversion after a DAC as the digital-system equivalent of a MC phono preamplifier.

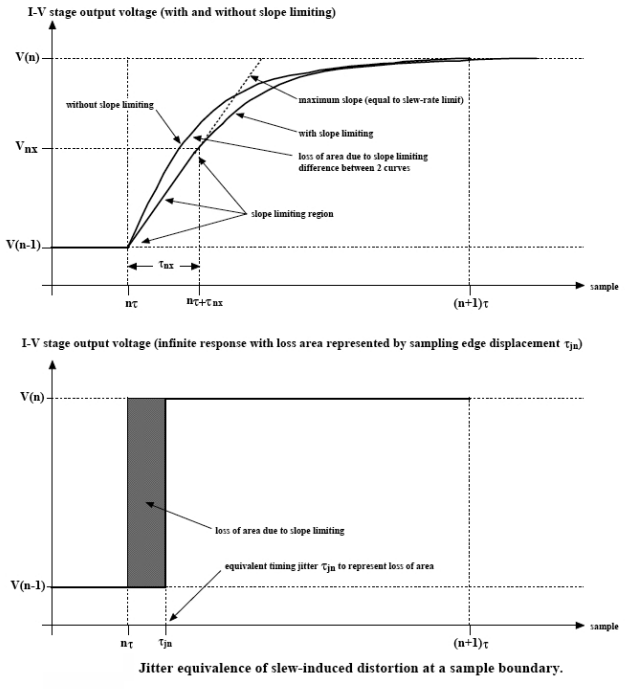


FIGURE 3: Slew-rate limiting in I/V converters has jitter equivalence.

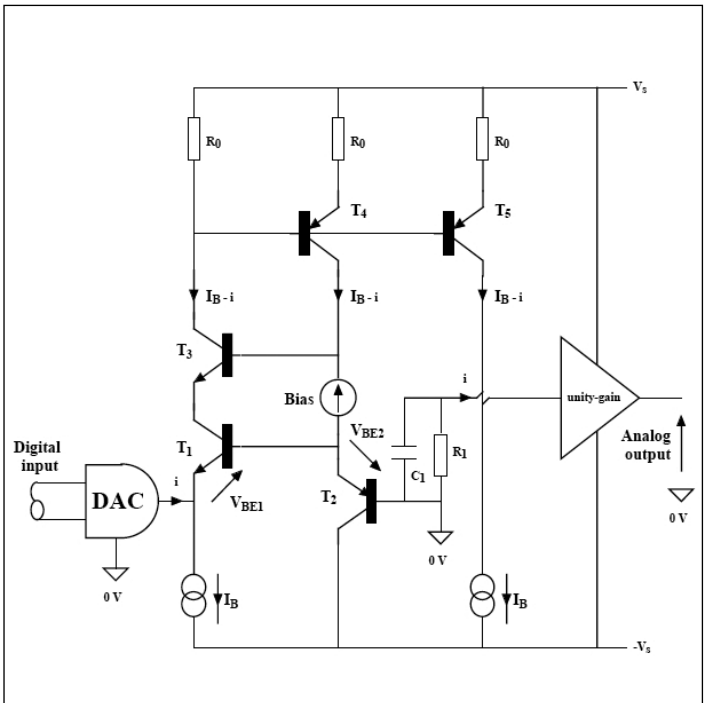


FIGURE 4: An open-loop I/V converter with integrated filter and input-stage error correction.

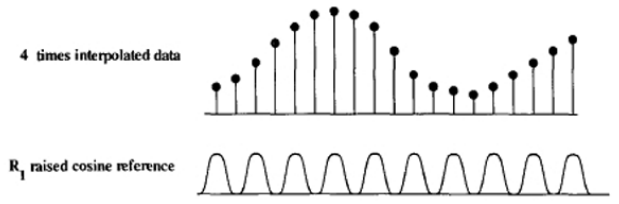


FIGURE 5A: Audio samples combined with raised-cosine DAC reference combine to. . .

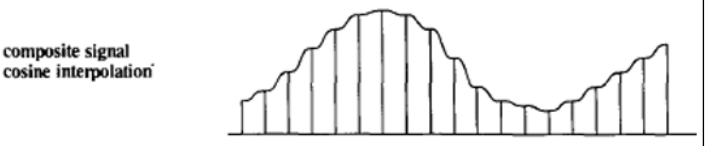


FIGURE 5B: ...dramatically reduce harmonics in analog output current reducing I/V slew rate requirements.

Although the application is totally different, I find that if you have learned to design a good MC preamp, that actually helps you to design a good I/V stage!

Another important issue is to locate the clocking source for the DAC very close to the DAC itself and slave everything, including the transport, to that clock. The clock should be free running, very pure and not controlled by a PLL; very often a PLL will only move the jitter to another frequency band and the frequency of oscillation is bound to wobble. There's nothing wrong with a free-running clock as long as you make sure that your data samples arrive on time, and you can do that with an appropriate buffer memory and data request protocol.

The CD player is a horrible RF environment, and you need to get the clock and DAC away from that source; just place a portable radio close to a CD player and do your own EMI testing! Even local supply bypassing of the DAC can couple noise into the supplies for the clock and increase jitter! So now you can list a few issues necessary to get it "right" in a digital playback system: the I/V; all the massive problems from EMI, supply, grounding, and so forth; and putting a clean clock right where you need it. I would speculate that if you gave a circuit topology to three different engineers to lay out a PCB and then build it, you would end up with three different results purely due to the differences in layout and component parts selection.

Now, how do you get a clean, stable DAC clock in your system? Suppose you have a transport and a DAC interconnected and you try to stabilize the DAC clock at the end of the digital interconnect; in principle you will succeed long-term, but in the short term that clock will wobble about and produce jitter. And even if you have your super DAC with clean clock and PLL with low filter cutoff, you still are faced with an input signal that is not necessarily clean. It can induce ground-rail interference and your supply may become contaminated, so that incoming jitter may then bypass all your hard work and still end up affecting the output of the DAC. Memory buffers can, of course, help in the smoothing process, but beware of power supply and ground-rail noise.

JD: Benchmark Media Systems claims that their DAC1 products succeed to almost get rid of jitter completely because they put a very clean clock next to the DAC with an option to slave the transport clock to it. Their USB interface apparently works the same in that it actually "requests" samples from the media player or PC, at a rate dictated by the clean DAC clock.

MH: Yes, network audio turns a lot of these issues upside down. It actually works the other way around. You put the DAC clock in charge, and it can be very clean and free running—no PLL—very low phase-noise. It is the way it's done in the Linn Klimax DS; the clock effectively "demands" audio samples from the network or NAS drive at its own pace to keep the buffer memory filled. I found the Klimax one of the cleanest and most articulate digital replay systems I've ever heard. For me, this is the way to go.

Now, I think that a good high-reso-

lution 24/96 or 24/192 audio file, delivered through a top-notch network DAC, can sound absolutely stunning, and I have some wonderful Chesky recordings at 192/24. But even a 16/44.1 CD recording, when played through a network DAC implemented correctly, can also sound pretty spectacular. Maybe not quite as good as the hi-res stuff, but very, very good nevertheless, and you would be hard-pressed to hear the difference. Of course, the CD recording quality has to be first rate, but that's a very different story!

JD: There are several companies out there trying to make this happen. Mark Waldrep's iTrax.com allows you to download music on a pay-per-download basis, where the price depends on the quality. You pay perhaps $2 for a 24/96 download, giving you actually the recording master, down to perhaps $0.69 for the MP3 version of the same music.

MH: Yes that's an extremely good way to do it, and if you look at the Linn website you'll see that they offer similar services. Linn also gives you the option to buy their hi-res content pre-loaded onto a NAS drive, which for some people is more convenient than the hassle of downloading and setting up playlists on the PC. *Chesky Records* is also a very excellent source of music, and they actually have some 24/192 material. The B&W model that you subscribe to and obtain regular downloads is also interesting.

I think if you make the price right and especially if you provide high-quality recordings, people won't cheat, generally. And these specialty music providers also are extremely careful about the recording quality of the music they list, so that is one more uncertainty removed from buying a CD, where you may like the music but maybe the recording quality isn't so good. So to me it looks that networked audio delivery is slowly coming of age, yes.

JD: Can we spend a few words on loudspeakers and their part of the audio performance? In your keynote speech to the Japan AES regional conference in 2001[5], you saw a great future to Distributed-Mode Loudspeakers (DML) to diminish the influence of room acoustics on music reproduction.

MH: Yes, indeed. You see, a DML has some great advantages. Rather than having a pistonic action like a traditional cone or panel speaker, a DML consists of a myriad of vibrating areas on a panel where in effect the impulse response of each of these small areas has low correlation with its neighbor. That is the significant thing which makes the polar response spatially diffuse. Now many people feel uneasy with that because it looks as though this will lead to a diffuse field, and it does!

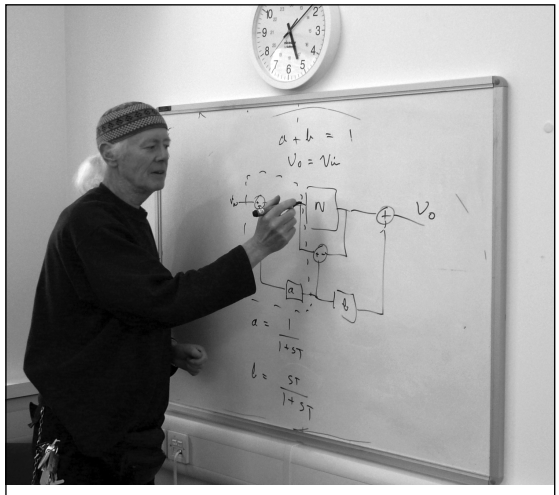


PHOTO 2: The Professor in his element: explaining feedback/feedforward concepts.

But, it does *not* lead to a significant breakdown of spatial sense or of instrument placing, because although the field becomes diffuse, the directivity characteristic does not. The major advantage is that the room acoustic reflections add with a significant degree of incoherence; they average out, so to speak, they are diffused. It helps to make an analogy between coherent light (from a laser) and incoherent light (from conventional lighting); in the latter the lack of interference results in much more even illumination without interference patterns.

To be honest, the sound stage itself does suffer a little bit, but the advantages can outweigh the disadvantages. You have no defined sweet spot, but you have no "bad" spot either when you move around the room. Furthermore, you could construct a DML as a flat panel, make it look like a painting, for instance, which makes 5.1 or 7.1 surround so much friendlier in the living room! You could even make them an integral part of your flat-panel video screen or, in principle, weave them into the fabric of the room architecture.

But as far as I know only NXT has taken up the technology for use in specific circumstances, and successfully, I might add. Now, for regular stereo use, DMLs may not quite give you the sharp holographic image traditional loudspeakers can achieve, but in practice that will not often happen anyway. People seldom place their loudspeakers in the correct position in the room to realize the full potential for imaging.

Now that we are discussing the diffuse characteristics of DML loudspeakers, it reminds me of something similar I have done with crossover filters[6], where the crossover transfer functions have a kind of random component added to them in the crossover region. The issue is: If you add the responses in a crossover on-axis, they add up and you should get a flat combined response. But if you add them off-axis, you normally would get a dip at the crossover frequency due to interference from non-coincident drivers. But with noise-like frequency responses, then for the off-axis sum, the interference is dispersed and the dip spreads out over some frequency band around the crossover frequency and becomes less pronounced. You diffuse the problem, so to speak. It's similar to what DMLs do: I call them *stochastic crossovers* (**Fig. 6A, B**).

JD: You would favor active speaker systems?

MH: Yes. I believe that active loudspeakers have a number of advantages due to using separate amplifiers for each frequency range. Intermodulation, either directly or through the power supply, is much easier to avoid, as different amplifiers handle different regions of the audio band. Amplifier peak power requirements are also relaxed, in turn making it a bit easier to build high-quality amplifiers. And, assuming close proximity between amplifier and its associated driver, then those pesky loudspeaker cables are largely removed from

the equation (smiling).

In the 80s I consulted on what I believe was the first digitally corrected active speaker, developed by Canon. In fact, Canon funded a research project at Essex where we (that is, Richard Bews, now proprietor of LFD Audio, and me) produced a system that was ultimately demonstrated at their research facility in Tokyo. [JD: I have listened to that system in a large room in a General's castle in Belgium in 1984 or thereabouts; it left a vivid memory!]. There are two key aspects to digital loudspeaker processing. First, you can use it as a digital crossover filter, which will allow you to very easily correct any loudspeaker response errors as part of the crossover code. This is much simpler and less costly than using high-quality analog crossovers.

But once you have the capability, the urge is often to use it for room correction as well. I'm not a fan of that, simply because (ideal) room correction can typically be done for only one specific listener location, the ubiquitous "sweet spot." At any other location, the response, including the phase response, goes down the drain. The problem is very much wavelength dependent, so accurate correction tends to be limited to low frequency with less precise frequency shaping being applied at higher frequency. So, I'd use digital loudspeaker processing only for crossovers and loudspeaker correction. You should deal with room influences (other than low-frequency modal compensation) through other methods, where intelligent loudspeaker placement is one powerful way to improve your stereo reproduction.

Now, there's another aspect to digital loudspeaker equalization. Loudspeakers are a bit like musical instruments really; they have their own coloration and character where often you chose what appeals to you. Now if you equalize that loudspeaker, you may compromise the attribute that you liked, so, although being more accurate you may have the impression something is missing or wrong. You should therefore consider digital loudspeaker correction an integral part of the design, just as is a passive crossover.

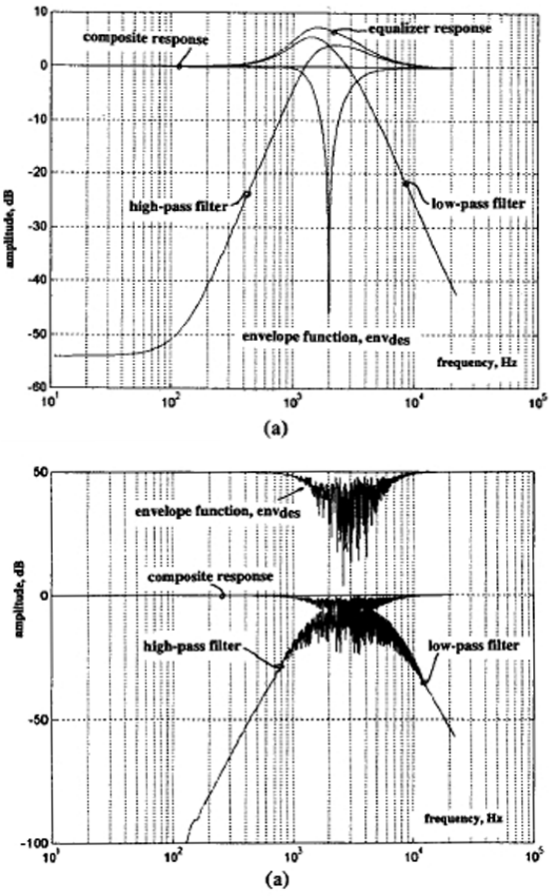


FIGURE 6A/B: Stochastic crossover has randomized filter characteristics that diffuses the off-axis crossover dip.

You shouldn't try to play with the correction or have switchable multiple corrections, just as you wouldn't want switchable multiple passive crossovers (apart from maybe some slight level correction in the low- or high-frequency range).

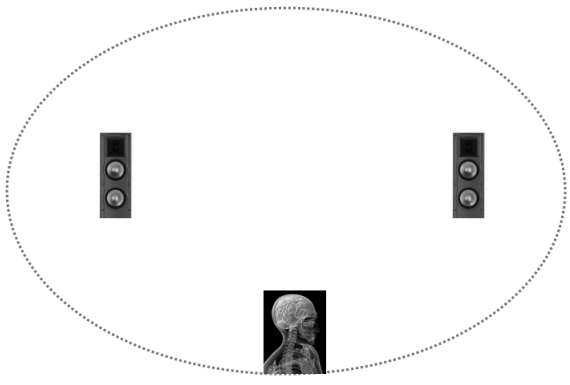


FIGURE 7: Idealized speaker placing (Joachim Gerhard, founder of Audio Physic).

JD: That's the philosophy of AudioData in Germany. They sell one of these digital speaker/room correction systems. They will come to your home and set the system up to your liking with the corrections and all. They do not encourage you to play around with it. They put your particular correction files on the Internet, so when something goes wrong in your system you can download and re-install them. But in a practical sense it is a one-time thing—to your room, your speakers, and your taste, if you will.

MH: Yes, that's sensible. People should listen to the music, not to their loudspeakers or correction processing! There's one more thing I'd like to mention about placement. In the past, I have worked closely with Joachim Gerhard (founder of loudspeaker company Audio Physic in Germany), who came up with one of the best placement schemes I know. You need to avoid reflections coming from the same direction as the direct sound, because this distorts your spatial perception (it messes with the head-related transfer functions we use in sound localization). Joachim drew an ellipse that just touched the inside of the room boundary. You then place the loudspeakers at the foci of this ellipse and place yourself at the middle of a long wall boundary (**Fig. 7**).

So, not only are the reflections now remote from the direct sound direction, they are also separated more in time, where both these effects have a major impact on localization and perception of the recording venue acoustics. In the context of a high quality two-channel audio system (using Audio Physic loudspeakers), it achieved one of the finest stereo soundstages I have ever heard. The sound seems to hang in there between the widely spaced loudspeakers; you can hear all detailed venue acoustics, very convincing, especially when

the room is darkened! Also, having the loudspeakers widely spaced increases the difference signal between our ears, which helps to produce a more 3-D like image. Very interesting.

JD: Siegfried Linkwitz makes the point that you should place the speakers such that there is a minimum of 8ms temporal separation between direct and reflected sound, so that your brain can separate out the recording venue acoustics from the room acoustics.
MH: Yes, I very much agree with that. With most stereo placements, you add room reflections to the sound which "dilute" the spatial properties. So you may think that you have a larger image, but that is because it is blurred! The ellipse-based placement I just mentioned separates the direct and reflected sound both in direction and timing, and so helps your brain to keep the original spatial properties intact.

JD: And then there's the issue of the speaker cables. I remember this paper you wrote in Marrakech I believe.
MH: I wrote a lot of papers in Marrakech! I like to get away now and then to a quiet place, away from daily distractions. I would get up at 5:00 AM and then work for three or four hours. Those hours can be very productive, what you would call "quality time."

But you probably refer to my article on cable effects and skin depth[7]. That one attracted a lot of criticism, and although there was a degree of speculation in it, I stand behind the major conclusions to this day. If I write something like that I always try to indicate what is fact, as we electronic engineers understand it, and what is more of a gut feeling.

In that article I addressed the topic of skin effect in the context of audio; however, it seems what I said was widely misunderstood and misquoted. Say you have a coaxial cable, consisting of lossless conductors (i.e., zero resistivity). All AC-current would then flow only on the two opposing inner surfaces as electromagnetic forces would push the charge carriers away from each other; the current would not penetrate the conductor and skin depth would tend to zero.

Here all the electromagnetic energy would flow only in the dielectric space between the two conductors, propagating in an axial direction along the cable close to the speed of light with the conductors acting as guiding rails. Here the electric field is radial, while the magnetic field is circumferential with power flow in a direction mutually at right angles to these two fields that is along the cable axis. Now because all practical conductors are lossy, you inevitably get potential differences along each conductor, and this means that at the cable surface there must be a component of the electric field in an axial direction; however, the surface magnetic field is still circumferential.

When you consider these two fields, the direction which is mutually at right angles is now directed in a radial direction into the interior of each conductor. As a consequence, there is a propagating electromagnetic wave (loss field) within the conductor itself. Think of it as energy spilling out into the guiding rails which are now partially lossy and therefore must dissipate some energy.

When you solve Maxwell's equation for propagation in a good conductor, you obtain a decaying wave because some energy is converted into heat. Also, the velocity is very slow and frequency dependent. It is this slowly propagating wave that determines the internal current distribution in the conductor and is the basis of skin depth; it also explains why skin depth increases with decreasing frequency. The "loss field" is at maximum at the surface and decays exponentially into the conductor.

So your current is no longer confined to the conductor surface but penetrates into the conductor; it depends on frequency and decays exponentially. Therefore, when you consider the series impedance of a cable, you find it is made of two principal parts. There is the inductive reactance due to the magnetic field within the dielectric between the conductors, and this, as you would expect, rises as 6dB/octave. However, the magnetic flux trapped inside the conductors has both a resistive and an inductive component. If the skin depth is such that the current has not fully penetrated all the way to the center of the conductor, then this component of impedance approximates to 3dB/octave.

What happens in practice depends

HAWKSFORD ON THE SOUND OF JITTER

There is a lot of talk about the effect of jitter on reproduced music. To help people to get a feel for it, I prepared some test files with well-defined amounts of jitter. Basically, what I did was to calculate the variation in digital sample values when a specific jitter signal would be present, and alter the samples accordingly. The tracks are on the *audioXpress* website and can be listened to or downloaded for your own use. Those of you adventurous enough to go through the details are referred to the reference below.

Track 0 is the original music, and the following tracks are the resulting amplitude-normalized "distortion" or error signals resulting from the types of jitter as listed:
Track 1: TPDF (triangular probability distribution function) noise-based jitter
Track 2: 2 equal-amplitude sinewaves (44100 - 50) Hz and (44100 + 50) Hz based jitter
Track 3: 3 sinewaves 50Hz, 100Hz, and 150Hz, amplitude ratio 1:0.5:0.25 based jitter
Track 4: sinewave 0.2Hz based jitter
Track 5: sinewave 10Hz based jitter
Track 6: 3 equal-amplitude sinewaves 1Hz, 50Hz, and 44100/4Hz based jitter
Track 7: All of the above 6 jitter sources combined.

NOTE: In a real-world situation these error signals require amplitude scaling to match the system jitter level; they have been normalized here to allow them to be auditioned.

Enjoy!

Reference: Jitter Simulation in high-resolution digital audio, Presented at the 121st AES Convention, October 5-8, 2006, San Francisco, Calif.

on the actual cable geometry and therefore which aspect of the impedance is dominant. I could go on, but I suggest you download "Unification" from my website for more information. So to conclude, at lower frequency the penetration is deeper while as frequency rises, the internal conductor impedance increases as the current becomes more confined to the surface layer, just as it would be if the conductor was lossless to begin with.

I also put some numbers to it and it turns out that when your conductor diameter is less than about 0.8mm, there are almost no skin effects even up to 20kHz.

Now, going back to loudspeaker cables, ideally you would want them to have just a very low value of resistance over the audio frequency band with no reactance. Due to the phenomena described above, that may not always be true, but there lies the art of loudspeaker cable design!

However, in understanding the problem with loudspeaker cables that can impact their perceived subjective performance, there is another important factor. Even if cables are completely linear, they still feed loudspeaker systems that offer a nonlinear load due to drive unit impedances changing dynamically with cone displacement, suspension nonlinearity, and possibly saturation effects in crossover components. As a result, the current entering the loudspeaker is a nonlinear function of the applied voltage; this, in turn, means that any voltage drop across the (even perfectly linear) cable also has a nonlinear component which must be added to the loudspeaker input voltage. It is interesting to audition these error signals in real-world systems where distortion can be clearly audible. So in this sense cables do impact the final sound where this process is probably responsible for perceived differences in character or coloration. aX

This interview with Professor Hawksford continues next month.

REFERENCES

1. "System Measurement and Identification Using Pseudorandom Filtered Noise and Music Sequences," *JAES*, Vol. 53, No. 4, 2005 April.

2. "Transconductance Power Amplifier Systems for Current-Driven Loudspeakers," *JAES*, Vol. 37, No. 10, 1989 October.

3. C. Dunn and M. O. J. Hawksford, "Is the AES/EBU/SPDIF Digital Audio Interface Flawed?," presented at the 93rd Convention of the Audio Engineering Society, *JAES (Abstracts)*, vol. 40, p. 1040 (1992 Dec.), preprint 3360.

4. Discrete integrated I/V.

5. Ultra high-resolution spatial audio technology for HDTV on DVD, keynote speech at AES 10th regional convention, Tokyo, Japan, June 13-15, 2001.

6. "Digital Signal Processing Tools for Loudspeaker Evaluation and Discrete-Time Crossover Design," *JAES*, vol. 45 no. 1-2, 1997 Jan./Feb.

7. Electrical Signal Propagation & Cable Theory, Malcolm Omar Hawksford, October 1995, available online at www.stereophile.com/reference/1095cable/.

Note: All papers referenced here and many more are available at Professor Hawksford's website, www.essex.ac.uk/csee/research/audio_lab/malcolms_publications.html

The Essex Echo: Audio According to Hawksford, Pt. 2

Jan Didden continues his discussion about audio technology with Professor Malcolm Hawksford.

JD: Let's move to amplifier electronics, because one thing that comes across clearly from your publications is that you enjoy electronic circuit design.

MH: Is it that clear? But it is true. My first amplifiers were tube-based, of course, and I still have a certain fondness for them. Most were simple, first-order circuits, with some pleasant coloration usually added by self-induced microphonics and vibrations. Different manufacturers using different tubes even with similar circuits show up different issues, but they err benignly, so to speak. It is very seldom that a tube amplifier's sound can't be enjoyed despite its technical limitations; the errors tend to be quite musical.

JD: What triggered your interest in error correction (EC)?

MH: Peter Walker's *Current Dumping* concept did that. I thought it an extremely clever and elegant solution (still do), and a "thinking out of the box" amplifier design that was *en vogue* at the time. There are various ways of looking at Current Dumping, but I explained it as a combination of feedback and feedforward techniques. The clever bit, as I saw it, was that it allowed you to design a structure that didn't require infinite gain to obtain theoretically zero distortion over a fairly broad bandwidth. In a feedback amplifier, as you move up in frequency, the feedback decreases leading to increasing distortion.

In this (then) new concept, the feedforward path compensates for the loss of feedback with frequency, and in theory you can keep up the "zero distortion" over the audio band. Of course, it depends on what stage of the amplifier produces distortion. It started me thinking about some way to generalize the concept of combining feedforward (ff) and feedback (fb)—which, of course, is at the core of Current Dumping—and explore other trade-offs in ff and fb. As the most objectionable distortion in a power amplifier is generated in the output stage, would it be possible to locally correct that output stage so that the remaining distortion signals that are fed back from the output to the input stage would be much cleaner (i.e., devoid of output stage distortion) thus also contributing to lower input-stage distortion? As N (the uncorrected output stage gain) approximates to 1, the error tends to zero and this makes the difference (correction) amplifier much more linear as it only amplifies small signals, and this holds even when the output voltage swing is large.

The conceptual view (**Fig. 8**) made it clear that, in theory, combining ff and fb can completely eliminate the forward loop nonlinearity, without the need for infinite loop gain, simply by choosing suitable combinations of transfer functions *a* and *b* in **Fig. 8** providing (a + b) = 1. Practical ff or fb networks will most probably need to have some active components and will thus be at least first-order low-pass circuits. But, if the "*a*" network has a first order 1/(1 + sT) characteristic, you could make "*b*" a conjugate sT/(1 + sT), and the elimination of distortion independent of frequency still holds.

Now, for the feedforward component "*b*," there is the practical problem of combining the forward and feedforward signal in the output (power) stage, so that is less attractive. Therefore, one solution would be to use only the "*a*" fb path, as it is much easier to combine low-level signals at the amplifier input. Because you now can no longer compensate for the first-order rolloff, the full curative properties of the system break down at higher frequencies so zero distortion is out of reach.

Yet, employing this type of error correction locally in, for instance, output stages still has significant advantages. Such fast-acting local correction does a good job to linearize the output stage by one or two orders of magnitude and, as a bonus, give very low output impedance before global feedback is applied. I also showed that you can implement a correction circuit virtually without needing more components than those used for biasing, so it's essentially free.

The local loop does not impact stability much, so you can have your cake and eat it, too. You end up with a more linear power amplifier for the same parts investment

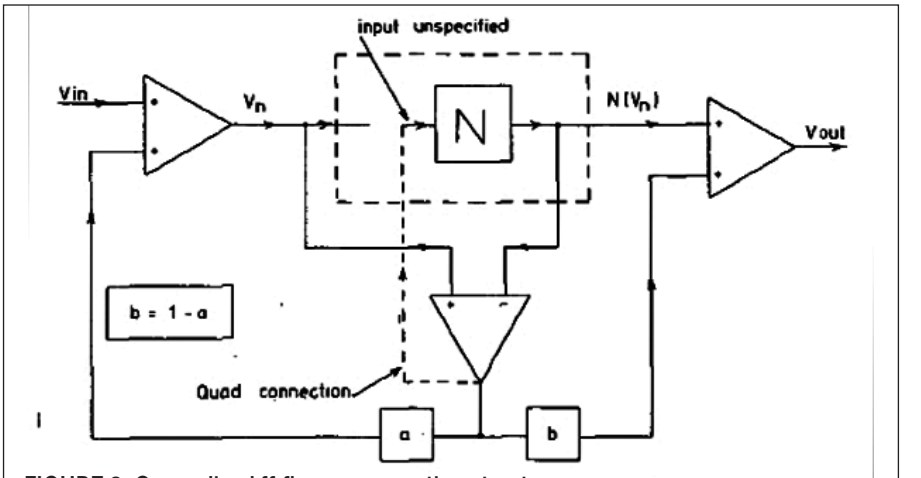


FIGURE 8: Generalized ff-fb error correction structure.

and that's always worthwhile. Bob Cordell had a very elegant implementation of this concept which I like very much[8].

JD: At one point there was a great discussion on diyaudio.com between Bob Cordell, yours truly, and other very smart circuit designers. The question was whether error correction is really a different circuit concept or whether it is another way of using negative feedback (nfb). That it was, to paraphrase evolutionary biologists, a matter of exploring the "space of all possible nfb implementations."

MH: Well, I guess that conceptually it is indeed a different way to apply nfb, but with some interesting different issues which also lead to more insight into this type of circuit. For instance, in **Fig. 8**, assuming that b = 0, then Vout/Vin = G = N/ (aN - (a - 1)). The target for Vout/Vin = 1, so now you can calculate the error function ε representing the overall input-to-output transfer function error, that is the deviation from "1," thus ε is defined as ε = 1 – G.

Substitution gives you ε = (a - 1)(N - 1)/(aN - (a - 1)). Now you immediately see that the error function has two zeros, i.e., (N - 1) and the balance condition represented by (a - 1). This succinctly explains the operation and power of EC, especially with near unity-gain output stages as you get two multiplicative terms in the error function which should both be close to zero. Half of the art of understanding and developing circuits lays in finding the right viewpoint!

JD: I know of at least one commercial implementation of what appears to be your EC concept, based directly on Bob Cordell's circuits, by Halcro. Presumably based on a patent by Candy, which came later in time than your publication.

MH: Yes, I am aware of that. At the time I sent Halcro my papers and wrote to them asking for some clarification, but never received a reply. So it goes. Anyway, life's too short to worry about such things. It's not my problem. Bob Stuart of Meridian Audio also used the circuit in his amplifier range for a period of time, which was most gratifying as he is a very gifted audio circuit and system designer.

There's analogy to error correction in the digital domain, and that is noise shaping. I wrote a paper with John Vanderkooy comparing digital noise shaping with nested differential feedback in analog circuits[9] and concluding that they can be seen as different views of similar issues! If you look at a first-order noise shaping configuration (**Fig. 9**), you see that, similar to EC, you take the difference between the forward block (the quantizer) input and output, which is the noise it generates, and feed it back to the input, properly shaped like $H = e^{(-sT)}$. Now, if you look at the noise shaping transfer function (1 - H), it looks very similar to the error reduction function of EC you showed before. So as you go lower in frequency, where the loop gain gets higher, the noise also gets lower.

Now this is a simple first-order case, but as you go to higher order noise shapers, your in-band noise gets lower at the expense of forcing more and more noise above the audio band. Now, if you put in a coefficient in (1 - H) of less than 1, then the reduction curve bottoms out at lower frequencies, so it is analogous to the bottoming out of your EC curve due to a less than 1 error-feedback coefficient. So, you could say that quantization noise shaping in sampled data systems is analogous to distortion-shaping in feedback or error correction in continuous signal systems. You often see that when the distortion is driven down by feedback or EC, it works for the first few harmonics at the expense of increasing higher harmonic components. Again, just like what we observe with noise shaping in digital systems!

You should look into the literature about Super-Bit Mapping (SBM). Michael Gerzon and Peter Craven in the UK worked on that as did Stanley Lipshitz and John Vanderkooy and also SONY. I well remember a rather heated argument between Michael and a Sony engineer during an AES convention some years ago! The idea with SBM is to apply noise shaping to a digital signal in the context of CD. Normally, with uniformly quantized and dithered 16-bit/44.1kHz LPCM, the

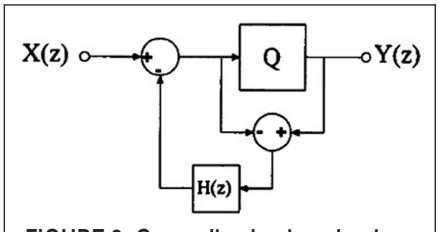


FIGURE 9: Generalized noise-shaping structure.

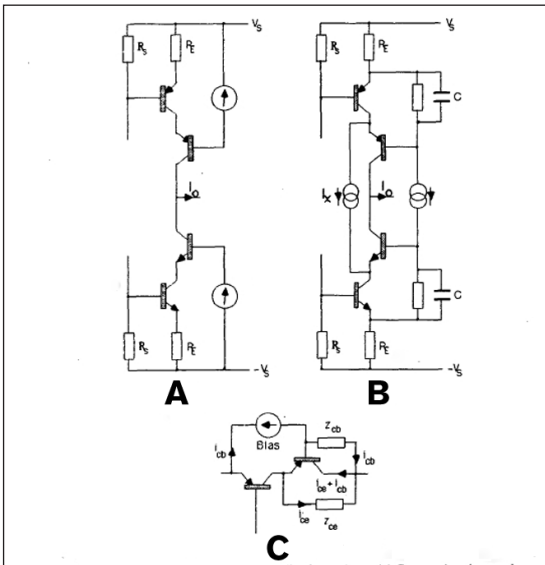
noise floor is essentially flat from DC to 22.05kHz.

Now, they asked, suppose we start with a 20 or 24-bit source, and we re-quantize and noise-shape the signal, can we somehow retain some of those additional bits of resolution below those 16 bits? Of course, the noise that you reduce in one part of the spectrum needs to go somewhere, and what SBM does is to decrease the noise in the mid band so you get perhaps 18-bit resolution in the frequency region where the ear is most sensitive.

The noise-shaping transfer function is designed to follow closely the Fletcher-Munson curves; consequently, the noise may rise by perhaps as much as 40dB at the very high frequencies, but because your ears are very insensitive in that area you cannot hear it. It is also important to realize that in a properly designed SBM system the noise is of constant level, and there should be no intermodulation with the signal. Also, the signal-transfer function is constant. So, provided that your DAC has at least 18-bit accuracy, you can perceive a subjective resolution of around 18 bit. And at its core, again, is a concept that you would recognize as an error-correction amplifier!

Your use of that AD844 current conveyor in your error-correction amplifier does remind me of a similar topology that I developed with two of my research students, Paul Mills and Richard Bews. This design, which led to the LFD moving-coil preamp, was published in *HiFi News* in May 1988. Richard subsequently developed this conceptual LFD pre-pre that used floating power supply circuitry by optimizing component selection and overall construction to achieve a very high level of performance. The reasoning behind the circuit is as follows: In a simple, single-ended emitter follower (**Fig. 11A**) the transconductance of the stage $G_m = 1/(r_e + R_E)$ where r_e is the intrinsic base resistance.

Since $r_e = 25/I_E$, you see that because r_e changes with signal current, this introduces distortion. You can improve on this (**Fig. 11B**), and now $G_m = 1/ (r_{e1} + r_{e2} + R_E)$, where, for example, when r_{e1} increases, r_{e2} falls. There is not perfect cancellation because the transistors of the long-tail pair are effectively connected in series in the AC-equivalent circuit, but it is much more linear than the previous case. You can further improve on that with **Fig. 11C**, where complementary transistors are now effectively in parallel for AC, so the changes in the respective r_es due to signal current are almost perfectly complementary such that the transconductance of the combined transistors is almost independent of signal current; that is, the circuit is linear.



Slope distortion reduction using feedback topology. (a) Conventional cascode. (b) Enhanced cascode. (c) Illustration of signal current paths i_{ce}, i_{cb} in z_{ce}, z_{cb}.

FIGURE 10: Enhanced cascode concept.

If you plot the nonlinearity (as an error function) versus the value of R_E and signal current (**Fig. 12**), you see that there is a point, with very low R_E, where the **Fig. 11C** stage is almost perfectly linear. So this is a valuable property, but as you can see there are some challenges in biasing it, especially with those very low-value emitter resistors. However, you can rework the circuit to retain the linearity yet make biasing somewhat easier.

Another most important aspect of the topology is the use of truly floating power supplies because even if the supply voltage were to vary or to exhibit noise, there is no signal path linking to the RIAA impedance, as related currents can only circulate in closed loops. Consequently, power-supply imperfections are dramatically reduced, which is very critical in MC applications where small signals can be sub microvolt in level.

Under large signal conditions, you have transistor slope resistances and slope capacitances which are being modulated by the signal, and that's potentially bad news. Some people call it phase modulation, going back to something Otala brought up many years ago.

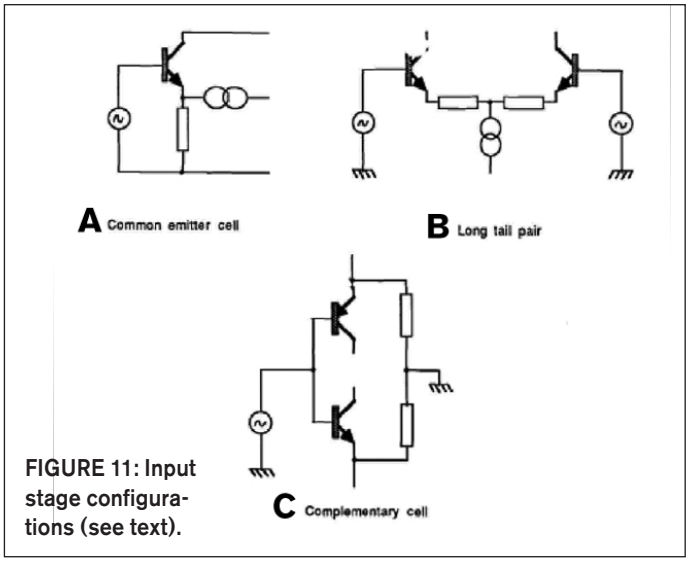
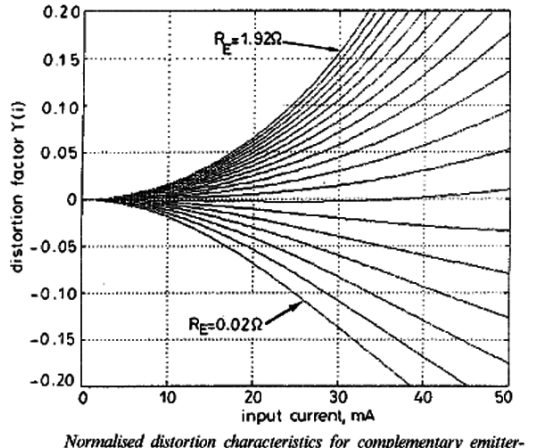


FIGURE 11: Input stage configurations (see text).



Normalised distortion characteristics for complementary emitter-coupled cell
Circuit data: I_{BQ} = 25mA; ΔR_E = 0.1Ω; input current 0–50mA

FIGURE 12: Input stage linearity versus RE.

It's more like a gain-bandwidth modulation, and I prefer to think of it as a time-domain modulation. For instance, in a feedback amplifier, this would slightly modulate the open-loop gain-bandwidth product and you can then calculate what it does to the closed-loop phase shift. It's like a signal-dependent phase shift, which manifests itself as jitter. It is analogous to a signal-dependent jitter, and it basically happens in all analog amplifiers. So, you have jitter in digital systems, you have jitter in I/V converters due to finite slew rate leading to slight modulation of the loop gain-bandwidth product, and you have these signal-dependent jitter-like phenomena in analog amplifiers in general, albeit that the modulation is time continuous rather than being instigated at discrete instants.

PHOTO 3: A younger Malcolm Hawksford showing off his speaker building skills.

You know, if you start to design a system, you need to have some sort of philosophy that drives you. For me, it is often the minimization of these timing errors, and I think that large-signal nonlinearity is less of a big deal than sometimes is believed. Most of the time you listen to low-level signals anyway, where linearity is very good. So then, you ask, what distinguishes one system from another, right in these low-level regions?

Now, I don't have any magic number, but let's assume that 100pS is the magic number for digital jitter, and suppose that you find similar numbers for what I call "dynamic timing errors" in analog amplifiers, the picture sort of comes together. It just *might* be that simple, open-loop circuits, while having higher large-signal level distortion, potentially have less of these timing nonlinearities, which could explain their very good sound. I would need to get the sums together, but it just might be possible that this is one of the reasons why people prefer those simple, low-feedback amplifiers. Especially in transistor circuits, where the transistor parameters themselves are modulated by changing voltage and current.

So having simple circuits that minimize these changes and are designed to minimize power supply influences clearly helps. Of course, feedback can help in many ways and there is no fundamental reason that a feedback amplifier cannot exhibit exemplary results, providing care is taken to minimize modulation of the amplifier loop transfer function.

Another example: When Paul Mills was still at Essex, he was working on an amplifier design using a cascode stage (**Fig. 10A**) that had reasonably low distortion. Then I told him, "Look, Paul, I will make one modification to your circuit that lowers the nonlinearity by an order of magnitude!" What I did was re-locate the biasing for the cascode to its emitter rather than to the supply (**Fig. 10B**).

It doesn't look like much, but it is a very significant change, and I can explain it with **Fig. 10C**. Why is the Z_{out} of a cascode not infinitely high and its distortion zero? It has to do with transistor slope parameters and their modulation with signal level.

PHOTO 4: Professor Hawksford and PhD student Adam Hill in the university audio lab.

You can see that an error current that is the difference between the ideal output (collector) current and the actual one is a result of the non-infinite impedances between emitter-collector and base-collector of the cascode transistor, where in **Fig. 10C** these two impedances are modeled by Z_{ce} and Z_{cb}. The modulation of transistor slope parameters with signal level I mentioned can be described as modulation of Z_{ce} and Z_{cb}. So, if you could find a way to prevent these error currents from ending up in the output (collector) current, then their bad influence would be eliminated.

Now, what is the effect of re-locating the bias to the emitter instead of the supply? For example, the i_{cb} error current now no longer comes from the supply but from the emitter of the top transistor. It is subtracted from its emitter current, which is basically the same as the cascode collector output current. So when i_{cb} is added to the cascode output current, it is no longer an error but makes up for the current that was subtracted in the first place! For i_{ce} a similar reasoning can be made. So the error currents now circulate locally in the stage and don't contribute to the output. It doesn't work perfectly, because there are some minor errors due to base currents, but it is, nevertheless, a huge improvement. The output impedance goes up typically by a factor of 10, and the distortion goes down by a factor of 10!

Note that it does not matter whether these error currents have a nonlinear relationship to the signal, as they do not contribute to the output current. This technique therefore works well in large signal amplifiers. I just picture this process in my mind, and I "see" what's going on, and then the solution pops up.

JD: You need to make the mental leap to model this modulation as an error current, and then find a way to shunt that error current away.

MH: Yes, indeed. There are some issues involving stability, as there is some form of regeneration in the circuit, but that's the gist of it. Now, I often wonder whether I would have seen that if I had

plugged it into a simulator and run a distortion analysis. I like to think that I might *not* have made that connection. I also believe that you should lay out the PC board, build your designs, and think about the topology at the same time. The days of a light box and black tape were great and very intuitive, very human. You move the layout around, changing this and that and in some way that connects back to the circuit again and you may then end up improving the circuit. It's an iterative process that can give you just that extra bit of quality or performance that you don't get when doing a sim and then saying, well, that's it.

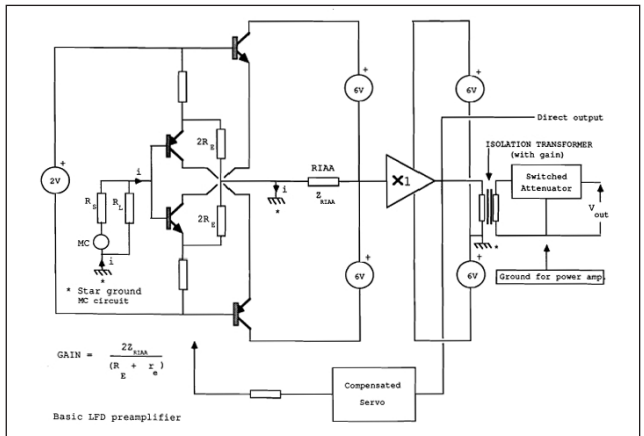


FIGURE 13: LFD preamp simplified diagram.

Anyway, this particular enhancement then appeared in my enhanced cascode paper[10]. Also, Richard Bews and I used this concept in the LFD preamp (**Fig. 13**), which, as previously mentioned, employed a true floating power supply system. And even if those batteries were to introduce some supply voltage nonlinearity, this doesn't show up in the output signal. There are no grounding problems because of the floating supplies. The floating-bias input pair is coupled to a cascode stage.

It's clear that any changes in that bias voltage do not have any influence on the output signal. So this will have high output impedance which drives current into the passive RIAA network to convert that current to voltage.

Now, if you look at which components determine the sound quality, it's only the input transistor emitter resistances and the components of the RIAA network. The cascodes don't do anything; the power supplies don't do anything, so it's an extremely linear circuit overall. And because it is only those few components, Richard was able to optimize component selection, ending up with a truly world-class preamp. Richard really is extraordinarily good at tuning and laying out circuits, and the battery-powered preamp worked extremely well. Also, this is why LFD Audio now enjoys almost cult status with its amplifier products.

JD: That **Fig. 13** circuit looks deceptively simple, but it is a very intricate circuit, isn't it?

MH: Yes, it is very simple, yet has a lot of interesting points: low noise, low distortion, almost no supply interaction, virtually no ground-rail current, very insensitive to transistor parameters, accurate RIAA correction, yet only a few active devices.

Often manufacturers have a good basic topology, but then they need to work in the power supply and grounding as well as the electrolytics and the other components in the signal path, and it all tends to blur the final sound. If you have many components in the signal chain, individual optimizations have relatively small impacts. But with this simple circuit, the components that determine the quality are few, and thus optimization has a relative large effect as well.

The absence of power supply interaction, however, is key to its performance. I find that at least as important as the topology itself, not only in preamps, but also in DACs and power amplifiers, for that matter. A lot of the differences between equipment in terms of clarity and cleanliness have to do with internal EMI issues and the power supply interactions and ground contamination.

JD: Well, we've already covered a lot of ground, but perhaps I can ask you about your views of switch-mode amplifiers.

MH: As you know, I've done a lot of work on Sigma-Delta (SD) modulation over the years. There is one proposal using an SD modulator driving an output stage with a pulse-density modulated signal. Now, the switching frequency would generally be higher than in the case of a PWM stage.

As you mentioned before, there is a basic problem with these types of circuits with EMI, and a higher switching frequency doesn't help. Do you remember our discussion with raised-cosine modulation in a DAC? Well, in this particular idea I used something similar. Instead of supplying the switching output stage with a stiff supply, you use a resonant supply synchronized to the switching frequency of the amplifier. The supply voltage would, in effect, be a raised cosine, so that at each switching instant the supply voltage would be zero, and would then smoothly rise toward the full value (**Fig. 14A**).

The result is that EMI problems are greatly reduced because the switching effectively occurs at zero voltage, and the harmonics are both lower in level as well as much lower in bandwidth. The output voltage of the amplifier is now no longer rectangular but somewhat sine-shaped (**Fig. 14B**). Switching efficiency of the output stage is im-

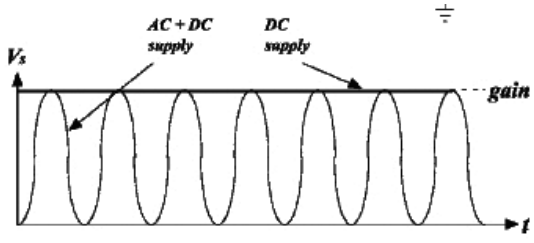


FIGURE 14A: Resonant power supply synchronized to sample rate outputs raised-cosine voltage.

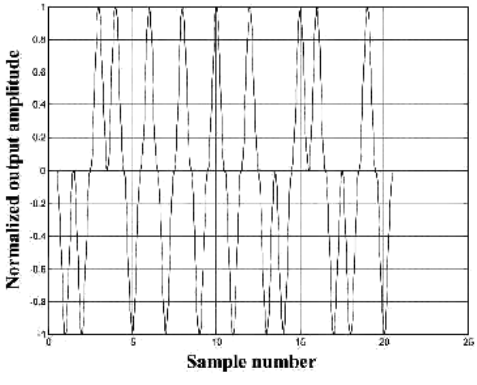


FIGURE 14B: Raised-cosine supply for switching amp dramatically reduces output signal bandwidth.

proved as well, and not only are those switches still either fully on or fully off, but because switching occurs with zero voltage across the device, power dissipation in the finite switching transition region is reduced. The average output level of this scheme is somewhat lower than a regular PWM amplifier, but that can be compensated for as described in the paper.

JD: Do you think that these switched-mode amplifiers can reach the quality levels of a good analog amplifier?
MH: Well, I've heard some commercial systems with B&O IcePower modules, which seemed to work really well, so I would say it's getting there, yes. It's an interesting technology, and even if the samples I've listened to were not always very low distortion, they did have a certain cleanliness and transparency to them. I'm not absolutely sure, but it may be related to the absence of low-level analog problems like dynamic modulation of device characteristics in an analog amplifier.

So, I'm fairly optimistic, also because it brings the digital signal closer and closer to the loudspeaker, skipping analog preamps and the like. Of course, you need to distinguish between "analog" switching amplifiers and "digital" switching amplifiers where the power amplifier is, in effect, the DAC. I have always been more interested in the latter class, especially the signal processing needed to achieve good linearity[11,12]. Just because an amplifier uses switching techniques does not necessarily make it a digital amplifier. This is an important distinction which is often misunderstood.

JD: Bruno Putzeys, a well-known designer of switching amplifiers, maintains that switching amps are analog amps: they work with voltage, current, and time—all analog quantities.
MH: Indeed. So, there are still a lot of problems to overcome, but they have a philosophical "rightness" about it.

JD: Not the least because of the high efficiency!
MH: Of course. And even if you want ultimate quality, running your amp in class-A with a 500W idle dissipation doesn't solve your quality issues either. There's much more to amplifier quality than just the choice between class-A or class-AB/B topology. An AB/B amplifier, properly implemented, with attention to all the often misunderstood issues of biasing, power distribution, grounding, and so forth, can sound so good that there is nothing to be gained by going to class-A. It's better to go for a simple system, with as few stages as possible because an additional stage cannot fully undo any damage done by a previous stage.

Now, a great-looking box with lots of dials and lights certainly may play music well, but for ultimate quality, get the best DAC you can afford (preferably a networked DAC linked to a NAS drive!), followed by a passive volume control and a great power amplifier and, of course, keep the cables short. Nothing can beat that, in my opinion.

JD: Professor Hawksford, thank you very much indeed for many hours of your time, for most interesting and illuminating discussions. In particular, I was intrigued by the correspondence between seemingly disparate phenomena, like noise shaping versus error correction and jitter versus analog phase modulation. I hope this will inspire readers to do their own experiments and come up with yet other interesting configurations. *aX*

REFERENCES

8. A MOSFET power amp with error correction, presented at the 72 AES Convention, Anaheim, 23-27 Oct. 1982, revised 25 Jul. and 27 Oct., 1985. Also on www.cordellaudio.com/papers/MOSFET_Power_Amp.pdf

9. Relationships between digital noise shaping and nested differentiating feedback loops, presented at the 93rd AES Convention, San Francisco, 1-4 Oct., 1992, revised 9 Nov. 1999.

10. "Reduction in Transistor Slope Impedance-dependent Distortion in Large Signal amplifiers," *J*AES, Vol. 36, no. 4, April 1988.

11. "Dynamic model-based linearization of quantized pulse-width modulation for applications in DA-converters and digital power amplifier systems," *J*AES, Vol. 40, no. 4, pp. 235-252, April 1992.

12. "Linearization of multi-level, multi-width digital PWM with applications in DA conversion," *J*AES, vol. 43, no. 10, pp. 787-798, October 1995.

Note: All papers referenced here and many more are available at www.essex.ac.uk/csee/research/audio_lab/malcolms_publications.html