ISSUES IN NONLINEAR CIRCUIT THEORY AND APPLICATION TO HIGH FREQUENCY LINEAR AMPLIFIER DESIGN

David G Haigh, Danny R Webster*, Reza Ataei* ,Tony E Parker♦ and Jonathan B Scott†*

*Department of Electronic & Electrical Engineering, University College, London WC1E 7JE, UK

♦ Department of Electronics, Macquarie University, Sydney 2109, Australia

† Agilent Technologies, 1400 Fountaingrove Parkway, Santa Rosa, 95403, USA

Abstract: This tutorial presents some concepts and techniques in nonlinear circuit theory which relate to the description of the nonlinear behaviour of high frequency field effect transistors and analyses the nonlinearity that occurs in circuits in which they are embedded. The concepts are sufficiently simple that they provide considerable engineering insight into nonlinear device and circuit behaviour and also allow hand analysis and design while providing sufficient relationship to real devices that practical high frequency power and medium signal amplifiers can be designed to meet stringent specifications that arise in communication systems.

1. INTRODUCTION

The need to apply nonlinear circuit theory to the solution of problems in the analysis and design of analogue circuits has never been greater [1]. The introduction of digital communications has not eliminated this need since the components operating at RF frequencies are necessarily analogue and must process carrier signals maintaining well defined amplitude and phase characteristics over a range of signal levels in order to obtain an acceptably low bit error rate. There is also a need to avoid spectral spreading of signals since this would require greater channel spacings and less efficient use of the available spectrum. These requirements impose a need for amplifiers which have to meet stringent linearity specifications and for nonlinear devices such as mixers in which the amplitudes of unwanted frequency components are strictly limited. Field effect transistors (FETs) used to realise communications circuits tend to have short channel lengths in order to obtain a high f_T (transition frequency or bandwidth) and this means that their characteristics tend to be quite unlike simple text book transistor models. Models with acceptable accuracy tend to be too complicated for hand analysis or design, let alone for the development of new circuit synthesis techniques required for modern systems. In this tutorial, we attempt to provide a bridge between the accurate description or modelling of FETs, some aspects of nonlinear circuit theory and the design of highly linear amplifiers for communications systems applications. The emphasis will be on presenting concepts which are powerful and yet sufficiently simple to provide insight and allow hand analysis.

We will assume the use of active devices which are FETs, which may be of any technology (eg Si JFET, Si MOSFET, SiGe FET, GaAs MESFET or HEMT, InP HEMT). Most of the nonlinear analysis work will be concerned with medium-signal linearity, assuming small signal excursions about a bias point. As we shall show, knowledge of medium-signal nonlinearity data over a wide range of bias voltages allows large signal design to be successfully achieved. The tutorial considers the problem for 'high frequency' amplifiers in the sense that the FETs are operating at frequencies where the parasitic device capacitances begin to affect linear and nonlinear performance, but the effect of the nonlinearity of device parasitic capacitances does not have to be considered.

2. DESCRIPTION OF FET NONLINEARITY USING DERIVATIVES

The symbol we will use for a FET is shown in Fig 1, together with a definition of its port voltages and currents. Our examples will be based on use of the depletion mode FET (eg JFET, MESFET or HEMT) although all the

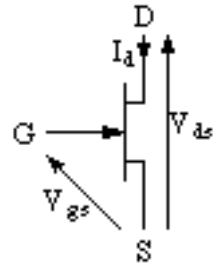


Fig 1 FET Symbol

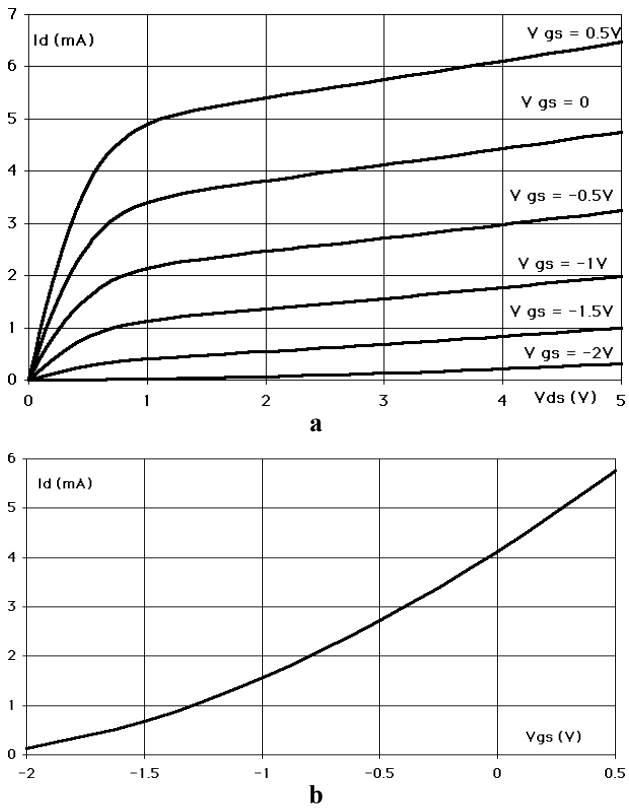


Fig 2 I-V curves for a FET (a) I_d versus V_{ds} for range of values of V_{gs} (b) I_d versus V_{gs} for a specific value of V_{ds}

results of this paper are equally applicable to enhancement mode FETs (eg MOSFETs). Typical characteristic curves for a FET are shown in Fig 2a, where FET drain current I_d is plotted against drain source voltage V_{ds} for fixed values of gate source voltage V_{gs}. Two regions can be identified in Fig 2a: for $V_{ds} > 1$ V (saturation region) the curves have a small uniform gradient and are widely spaced; for low V_{ds} (triode region) they have a highly variable gradient and highly variable spacing. For most linear applications, such as amplifiers, FETs are operated in the saturation region. In Fig 2b, we show a plot of I_d versus V_{gs} for a fixed value of V_{ds} in the saturation region. If we regard I_d as the output variable of the FET and V_{gs} as the input variable, then Fig 2b can be regarded as the FET transfer characteristic. In practice the gate of the FET is biased at a DC voltage V_{GS} and an AC voltage v_{gs} is superimposed (∴ Instantaneous total gate source voltage $V_{gs} = V_{GS} + v_{gs}$). If v_{gs} is sufficiently small, then we may say that

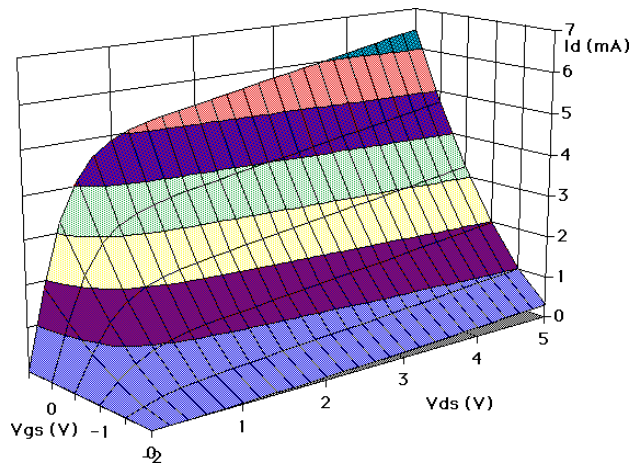


Fig 2(c) I_d versus V_{ds} and V_{gs}

the gain is given by the gradient of the curve in Fig 2b at the chosen bias point

$$i_d = \frac{\partial I_d}{\partial V_{gs}} v_{gs} = g_1 v_{gs} \qquad (1)$$

This is a small signal description and g_1 is referred to as the linear transconductance. If the signal variation v_{gs} is not very small, and we want to represent the nonlinear dependence of i_d on v_{gs}, we can use a power series in v_{gs}

$$i_d = g_1 v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3 + \ldots \qquad (2)$$

This equation can not predict the behaviour of the curve in Fig 2b near the threshold voltage (V_T = -2 V) (this problem will be discussed in section 5) and also can not efficiently describe the slight tendency towards saturation at high V_{gs}. It is therefore only suitable for a restricted range of signal variations about a bias point. It is referred to as a medium signal nonlinear transconductance description.

In Fig 2c, we show the information contained in Fig 2a as a 3-dimensional plot in which I_d is now represented by a surface. If a FET is biased in the saturation region by defining the gate and drain bias voltages V_{GS} and V_{DS}, then i_d shows a strong dependence on v_{gs} but a much weaker dependence on v_{ds}. For small signal variations, we may write the linearised relationship:

$$i_d = \frac{\partial I_d}{\partial V_{gs}} v_{gs} + \frac{\partial I_d}{\partial V_{ds}} v_{ds} = g_1 v_{gs} + g_{d1} v_{ds} \qquad (3)$$

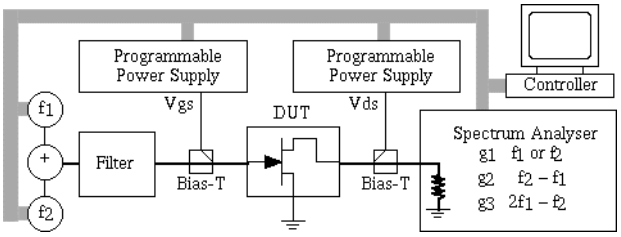


Fig 3 Block diagram of test set up for measurement of FET transconductance derivatives

This is equivalent to representing a small part of the I_d surface by a plane. If we want to represent the curvature of the surface at that bias point, we can use a power series in v_{gs} and v_{ds} to generate a medium signal (nonlinear) description [2,3]:

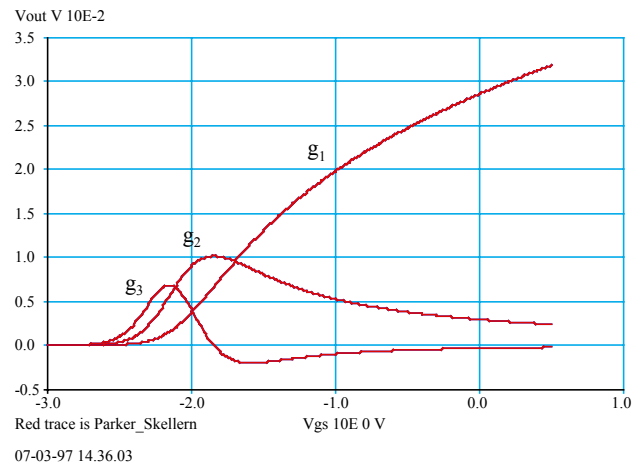
$$\begin{aligned} i_d &= g_1 v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3 + \ldots \\ &+ g_{d1} v_{ds} + g_{d2} v_{ds}^2 + g_{d3} v_{ds}^3 + \ldots \\ &+ m_{11} v_{ds} v_{gs} + m_{12} v_{ds} v_{gs}^2 + m_{21} v_{ds}^2 v_{gs} + \ldots \end{aligned} \tag{4}$$

The g_i coefficients describe the curvature of the surface along the v_{gs} axis (transconductance nonlinearity) and the g_{di} coefficients describe the curvature of the surface along the v_{ds} axis (output conductance nonlinearity); the m_{ij} coefficients describe the variation of the nonlinearity along the v_{gs} axis as v_{ds} is varied and vice versa (mixing terms). (4) is not valid for large signal swings and is a medium signal description. (4) will form the basis of our analysis in section 4 of the effect of varying load resistance on FET nonlinearity.

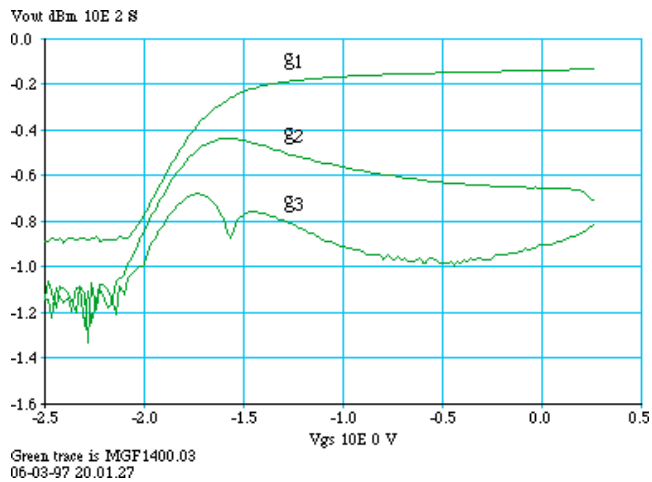
Since in (4), the g_i coefficients tend to dominate, the approximate form given in (2) which ignores v_{ds} dependence is often useful. Successive differentiation of i_d with respect to v_{gs} yields the following:

$$\begin{aligned} \frac{di_d}{dv_{gs}} &= g_1 + 2g_2 v_{gs} + 3g_3 v_{gs}^2 + \ldots \\ \frac{d^2 i_d}{dv_{gs}^2} &= 2g_2 + 6g_3 v_{gs} + \ldots \\ \frac{d^3 i_d}{dv_{gs}^3} &= 6g_3 + \ldots \end{aligned} \tag{5}$$

Since in practice $g_1 >> g_2 >> g_3 >>\ldots$ and we are restricted to medium signal amplitudes, the scaled coefficients, g_1, $2g_2$, $6g_3$, ... are in fact close approximations to successive derivatives of i_d with respect to v_{gs} and g_1, g_2, g_3, ... are therefore often referred to as derivatives. By assuming that the gate source port of a FET is excited by a sine wave, it may be shown, using (2) and (5), that the i'th derivative determines the magnitude of the i'th harmonic component. Since high frequency amplifiers tend to be narrow band, harmonic distortion is of little significance. In this case the i'th derivative determines the magnitude of the i'th intermodulation distortion component.



a



b

Fig 4 Typical form of transconductance derivatives for a FET (a) linear-linear scales (b) log-log scales

Techniques have been proposed for measuring the coefficients in (4) [4,5]. However, it is the transconductance derivatives (in (2)) which are largest and most significant and we show in Fig 3 a simple test set-up to measure these derivatives. Two signal

sources generate sinewaves of frequencies f_1 and f_2 which are combined and fed to the gate source port of the FET under test. The drain is connected to a load resistor and a spectrum analyser which measures the amplitude of the signals at f_1 (g_1), $f_2 - f_1$ (g_2) and $2f_1 - f_2$ (g_3). It is important that residual signals generated in the combiner at the test frequencies $f_2 - f_1$ and $2f_1 - f_2$ are negligible and a filter is provided to remove these components. The magnitudes of g_1, g_2 and g_3 are typically measured for a range of gate and drain bias voltages, but it is the variation with V_{gs} which is greater and of most interest. Fig 4a shows a typical plot of g_1, g_2 and g_3 for a GaAs MESFET. This general form of the curves has been observed in a range of FETs including JFETs, MOSFETs and MESFETs. HEMTs, to be discussed in section 6, have a slightly different form of curves due to their mode of operation. The form of the curves in Fig 4a can be easily explained since they are derivatives of the drain current curve, eg Fig 2b. As the gate source voltage is increased from the threshold voltage, I_d increases rapidly and then less rapidly for high V_{gs}. Since g_1 (transconductance) in Fig 4a is the derivative of I_d, it starts from zero, reaches a maximum gradient at around V_{gs} = -1.8 V and the gradient then reduces. g_2, the next derivative, must start from zero, reach a peak where g_1 has a maximum gradient and then fall. Finally g_3 reaches a positive peak, then falls to zero (where g_2 has its peak and g_1 its maximum gradient) and then exhibits a negative peak and decays to zero. Fig 4b shows the graph of Fig 4a but using logarithmic scales. It can be seen that the zero in g_3 has become a notch. Such curves, will form the basis of our linear circuit synthesis technique to be discussed in section 6.

2nd order distortion (ie g_2) is generally of little concern in narrowband communication systems because the frequency products which it generates (sum, difference and harmonic frequencies) are far removed from the carrier frequencies and are out of band. Although in general $|g_3| << |g_2|$, 3rd order distortion is much more serious because it generates intermodulation products which are close to the carrier frequencies.

The derivatives in Fig 4 are for medium signal variations about an operating point and are plotted against gate bias voltage V_{gs}. Since V_{gs} can represent a bias voltage plus signal variation ($V_{gs} = V_{GS} + v_{gs}$), Fig 4 shows us approximately how the derivatives vary over the instantaneous voltage values of a large signal. Since g_3 varies with V_{gs}, the effective g_3 for a large signal will be some kind of average of g_3 over the range covered by the instantaneous V_{gs} values. If a signal traverses parts of the g_3 derivative which have opposite signs, then cancellation can occur, as will be exploited in the latter part of section 6.

The derivatives in Fig 4 are measured for a constant value of V_{ds}. In practice, a FET will drive a load which will cause V_{ds} to vary with the signal and in many cases will be driving a resistive load which means that V_{ds} will be an inverted magnified version of V_{gs}. This corresponds to movement along a resistive load line of gradient $-1/R_L$ in Fig 2a. It is possible, in the measurement set up in Fig 3 to control V_{ds} and V_{gs} together to plot the derivatives along a load line in order to provide derivatives which can provide large signal distortion information for the resistive load case.

The concept of derivatives which can describe the nonlinearity of a device has been developed from the FET I-V curves in Fig 2c. High frequency FETs tend to suffer from frequency dispersion which causes the I-V curves measured at DC to differ from those measured using high frequency pulsed methods [6,7]. The method given for measuring derivatives does not use the I-V curves, but if we try to relate I-V curves to measured derivatives, this will only be valid if the I-V curves are measured under pulsed conditions.

The material in this section has been concerned with FET transconductance and output conductance nonlinearity which are frequency independent. In practice FETs have terminal capacitances. In the next section we consider this aspect of FET performance.

3. DEVICE-CIRCUIT INTERACTION AND FREQUENCY DEPENDENT DISTORTION

In a real device, the transconductance and output conductance nonlinearity we have described above may be considered to derive from the core of the device, the channel region under the gate. This region also gives rise to a distributed capacitance between the channel and the gate which is usually partitioned into gate-source and gate-drain capacitances. For a

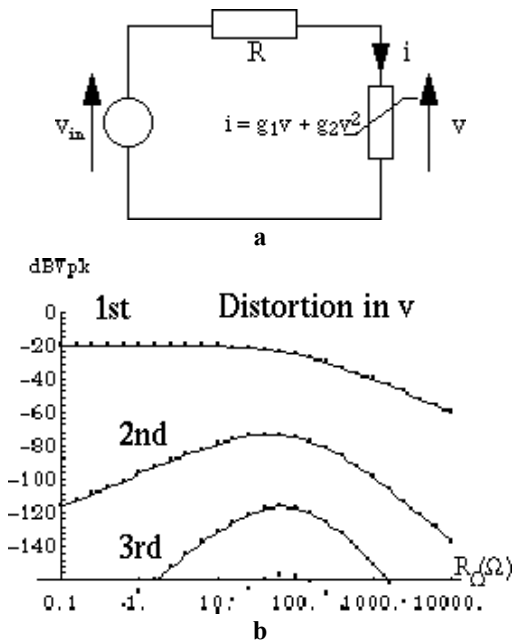


Fig 5 Example of device-circuit interaction (a) circuit (b) computed harmonics

Table 1 Expressions for harmonics for simple example circuit

Component	Amplitude	Phase
Fundamental	$\frac{1}{1+g_1R}$	**0**
2nd Harmonic	$\frac{1}{2}\frac{g_2R}{(1+g_1R)^3}V_{in}$	**π/2**
3rd Harmonic	$\frac{1}{2}\frac{(g_2R)^2}{(1+g_1R)^5}V_{in}^3$	**π**

given bias point, these capacitances consist of a linear (constant) capacitance in parallel with voltage dependent (nonlinear) capacitances. The linear capacitance (mainly the gate-source capacitance C_{gs}) determines the f_T or bandwidth of the FET. In this section we investigate a very important phenomenon called frequency-dependent distortion which is due to the interaction between the linear FET capacitance C_{gs} and the transconductance nonlinearity of the FET. It is the linear part of the device capacitance which most affects high frequency distortion, and we may lump the linear capacitances together with the linear elements of the circuit in which the FET is embedded, such as source and load impedances, feedback impedances, and consider the general interaction between a device (nonlinear transconductance) and linear circuit elements, which we call device-circuit interaction [3,8].

We begin our study of device-circuit interaction by considering some simple example circuits with passive elements. Consider the circuit in Fig 5a which contains a linear resistance R and a nonlinear conductance described as follows:

$$i = g_1 v + g_2 v^2 \tag{6}$$

where 3rd and higher order terms are neglected. Volterra analysis may be used to calculate the harmonics in the output voltage v and these are tabulated in Table 1 and plotted against the resistance R in Fig 5b. The first thing to notice is that 3rd harmonic exists even though the nonlinear conductance in (6) has $g_3 = 0$. The reason for this is that the 2nd harmonic in the current i due to g_2 flowing through R causes a 2nd harmonic component in v which is mixed with the fundamental by g_2 to cause a 3rd harmonic component in the current.

This is the simplest form of device circuit interaction and the form of the curves in Fig 5b can be seen in many situations. If the resistor R in Fig 5a is replaced by a capacitor then similar curves are obtained but the horizontal scale becomes 1/frequency. This is the simplest manifestation of frequency dependent distortion, in which, at the f_T frequency, the fundamental rolls off and the harmonic amplitudes reach a peak.

We now consider some basic FET circuits in which frequency dependent distortion occurs. Fig 6a shows a FET in common-gate configuration excited by a sinusoidal current at the source terminal. The admittance Y is given by $Y = G + j\omega C_{gs}$, where G is the self conductance of the current source and C_{gs} is the FET gate-source capacitance which defines its f_T. Assuming that the FET can be described by (6), the distortion in the drain current i as a function of Re(Y) and Im(Y) is plotted in Fig 7. We can consider the Im(Y) axis as the frequency axis and note that at the f_T frequency (Im(Y) = 1), the fundamental component begins to fall and both the 2nd and 3rd order distortion have significant peaks. For a real FET, $g_3 \neq 0$ and hence the 3rd harmonic in Fig 7 would approach a constant level as $Y \rightarrow 0$ dependent on the value of

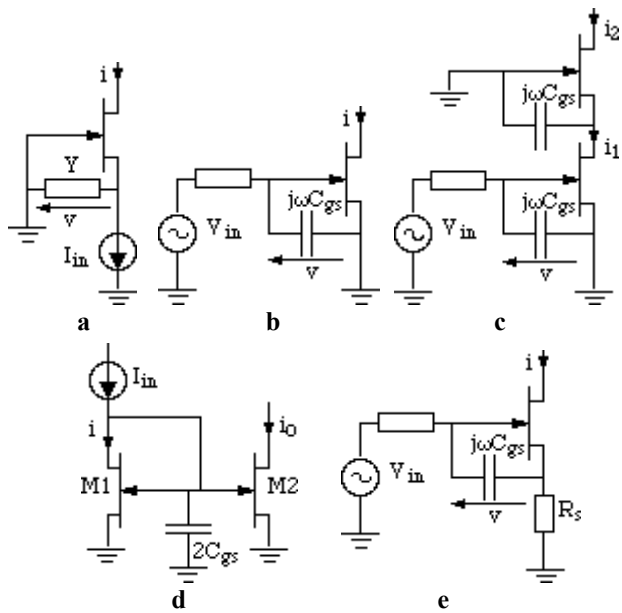


Fig 6 Simple FET circuits (a) common- gate FET (b) common- source FET (c) cascode FET amplifier (d) common-gate FET -to- current mirror conversion (e) practical common-source amplifier

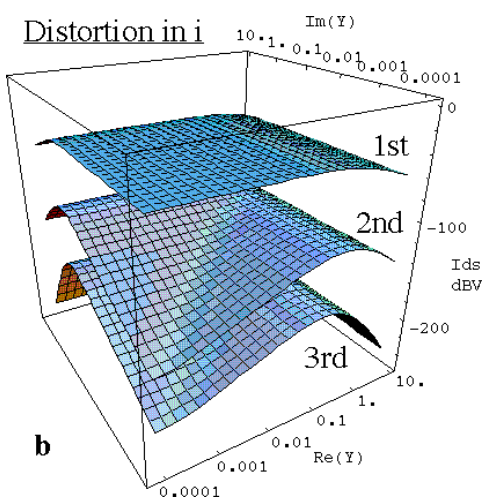


Fig 7 Harmonic distortion in the output current of the common-gate FET as a function of Re(Y) and Im(Y)

g_3. Nevertheless, the 3rd order frequency dependent distortion due to g_2 and C_{gs} will usually dominate the 3rd order distortion due to g_3 at frequencies approaching f_T.

The generation of 3rd harmonic by means of C_{gs} and g_2 in the common-gate FET of Fig 6a can be understood qualitatively as follows. Initially, we assume that Y is small. The sinusoidal current flows into the source and out of the drain, ie there is no distortion. But due to the FET characteristic in (6), the gate-source voltage v is given by the inverse of the characteristic in (6) which contains all harmonic components. As Y increases, then the nonlinear voltage v causes increasing nonlinear currents to flow in Y. By Kirchoff's current law, these nonlinear currents must also exist at the drain. Hence as frequency increases, all harmonic components at the drain increase in amplitude until Y is so large that the fundamental is diverted into Y and all harmonic components then start to decrease, as shown in Fig 7.

For comparison with the common-gate FET, Fig 6b shows a FET in common-source configuration excited by a sinusoidal voltage. In this case, the source is applied to the gate-source port of the FET via the source resistance and C_{gs}, the effect of which will be to linearly filter the signal before applying it to the gate-source port. Hence drain current will be determined only by the FET transconductance nonlinearity and there is no device circuit interaction or frequency dependent distortion apart from reduction of the fundamental and all frequency components around the f_T frequency. Thus the common-source FET will not exhibit the frequency dependent 3rd order distortion peaking due to C_{gs} and g_2 interaction and therefore distortion at high frequencies will be at a much lower level dependent on g_3.

Fig 6c shows a cascode FET pair which is a frequently used circuit configuration. Since the input signal is applied directly to the gate-source port of the lower FET, its drain current i_1 will be determined by its transconductance nonlinearity and will not exhibit frequency dependent distortion. However, when this current flows into the upper FET, harmonic components of the current will flow into its gate-source capacitance as frequency approaches f_T and these nonlinear currents will appear in the output current i_2. Hence the cascode FET pair will exhibit significant frequency dependent distortion.

If we ground the drain of the common-gate FET in Fig 6a, the FET becomes a 2-terminal device and may be interchanged with the current source as in Fig 6d. The current i in this circuit suffers from the same frequency-dependent distortion as does the circuit in

Fig 6a. When we add the 2nd FET M2 in Fig 6c, we recognise the current mirror circuit. The frequency dependent distortion in i, is mirrored in the output current i_o . Thus the current mirror circuit suffers from the same problem of frequency dependent distortion as the common-gate FET, which has been plotted in Fig 7. A novel current mirror circuit which overcomes this problem has been proposed in [9].

Fig 6e shows the common-source FET of Fig 6b but with the addition of resistance R_s which can represent the sum of the parasitic source resistance of the FET and any source degeneration resistance included in the circuit. If $R_s = 0$, then the circuit reduces to that in Fig 6b and there is no device circuit interaction and no frequency dependent distortion. If $R_s \neq 0$, the situation changes. The harmonics in the drain current i flow through R_s and cause voltage harmonics which act in series with the input source to determine the FET gate-source voltage. It has been shown that distortion rises with R_s and has some frequency dependence due to C_{gs}. If R_s is minimised to the parasitic resistance of the FET, this effect is not serious and frequency dependent distortion is practically negligible in this circuit.

From the results in this section, we conclude the following:-

- Frequency dependent distortion is insignificant in the common-source amplifier and it is a recommended choice for high frequency amplifier design. It may be important to consider the effect of feedback, particularly, source degeneration resistance, on distortion performance.
- The common-gate FET, the cascode FET, the current mirror and the common-source FET with significant feedback all manifest device-circuit interaction (between C_{gs} and g_2) which appears as frequency dependent distortion peaking at around the device f_T. Such circuit configurations are not recommended for use in high frequency linear circuit designs unless they are carefully analysed and distortion compensation is considered.
- For situations where the distortion perf-ormance of the common-source FET is not acceptable, this may be improved and suitable techniques will be discussed in Section 6.

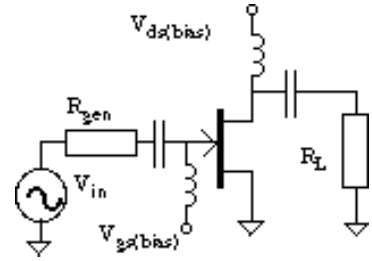


Fig 8 Common-source amplifier circuit

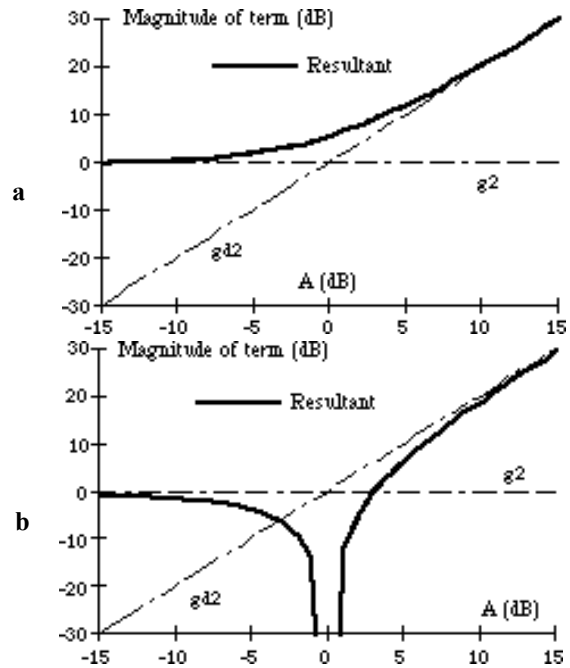


Fig 9 2nd order distortion scenarios for common-source FET amplifier

4. COMMON-SOURCE FET AMPLIFIER DISTORTION ANALYSIS

In section 3, we simplified the FET distortion analysis by making a number of assumptions, including the assumption of zero FET output conductance (ie assuming no dependence of i_d on v_{ds}) and the assumption of a simple FET transconductance model up to 2nd order (ie $g_3 = 0$). However, the analysis did serve to indicate that we should restrict ourselves to the common-source FET to avoid severe frequency dependent distortion. Having made this decision, we now develop a simple technique to predict the nonlinearity of a common-source FET amplifier where we do not ignore output conductance and higher order

transconductance nonlinearity [3]. We assume that the FET is connected in a circuit defining gate and drain bias as shown in Fig 8 with a load resistance R_L and we assume that we are working at a frequency up to say $f_T/5$ where nonlinear behaviour is still dominated by FET transconductance and output conductance nonlinearity. The analysis technique will make use of the FET derivatives introduced in section 2.

If the FET in Fig 8 is operated at well defined gate and drain bias voltages, it may be described by the medium signal nonlinear description obtained previously in (4). For the circuit in Fig 8, the output voltage is given by

$$v_{ds} = -i_d R_L \tag{7}$$

i_d contains all the terms given in (4), some of which depend on V_{gs} and some on V_{ds}, forming a recurrence relationship. However, we may take the dominant term in i_d, namely $i_d \approx g_1 v_{gs}$ in order to determine v_{ds} using (7). Hence we obtain $v_{ds} \approx -g_1 R_L v_{gs}$., which we may write

$$v_{ds} = A v_{gs} \tag{8}$$

where A is a parameter related to voltage gain. Substituting (8) into (4) and rearranging, we obtain

$$\begin{aligned} i_d &= g_1 v_{gs} + g_2 v_{gs}^2 + g_3 v_{gs}^3 + \ldots \\ &\quad + g_{d1} A v_{gs} + g_{d2} A^2 v_{gs}^2 + g_{d3} A^3 v_{gs}^3 + \ldots \\ &\quad + m_{11} A v_{gs}^2 + m_{12} A v_{gs}^3 + m_{21} A^2 v_{gs}^3 + \ldots \\ &= (g_1 + g_{d1} A) v_{gs} \\ &\quad + \left(g_2 + m_{11} A + g_{d2} A^2\right) v_{gs}^2 \\ &\quad + \left(g_3 + m_{12} A + m_{21} A^2 + g_{d3} A^3\right) v_{gs}^3 + \ldots \end{aligned} \tag{9}$$

where the terms responsible for 2nd and 3rd order distortion have been grouped together. First we consider 2nd order distortion (terms in v_{gs}^2). For a real FET, the effect of m_{11} is small. The effect of the remaining terms g_2 and g_{d2} depends on their relative signs. We may plot the 2nd order distortion terms against A using a log-log scale. The g_2 term has a gradient of zero and the g_{d2} term has a gradient of 2. Hence the lines corresponding to the individual terms will cross, as shown in Fig 9. If g_2 and g_{d2} have the same sign, then the resultant will move smoothly between the two lines (Fig 9a). If on the other hand they have opposite signs, then 2nd order distortion will show a sharp null with A (Fig 9b).

The 3rd order distortion mechanism is somewhat more complicated. For a typical FET, g_3, m_{12} and m_{21} are the most significant terms. g_3 and m_{21} have gradients of 0 and 2 and tend to have the same sign leading to an envelope which moves smoothly between the two (Fig 10a). The term m_{12} has a gradient of 1 and an opposite sign and different scenarios are possible depending on the magnitude of m_{12}. If m_{12} does not cross the resultant (Fig 10a), then the overall resultant has a broad shallow null. If it just touches the resultant, then there is a broad deep null (Fig 10b). If m_{12} crosses the resultant, then there are two deep narrower nulls (Fig 10c). Thus a wide variety of behaviour is possible depending on the magnitude of the coefficients.

In Fig 11 we show the measured distortion versus load resistance (ie gain) for a typical GaAs MESFET in common-source configuration [10]. It can be seen that the 2nd order distortion exhibits a deep null, manifesting the behaviour shown in Fig 9b and that the 3rd order distortion shows a shallow null according to the behaviour shown in Fig 10a. In Fig 12, we show simulations of the distortion of this amplifier using a number of standard FET models. It can be seen that the models vary widely in their predictions showing all behaviours characterised in Figs 9 and 10 (shallow null, single null and double null). Only the Parker Skellern model [11] comes close to predicting the actual measured behaviour in Fig 11.

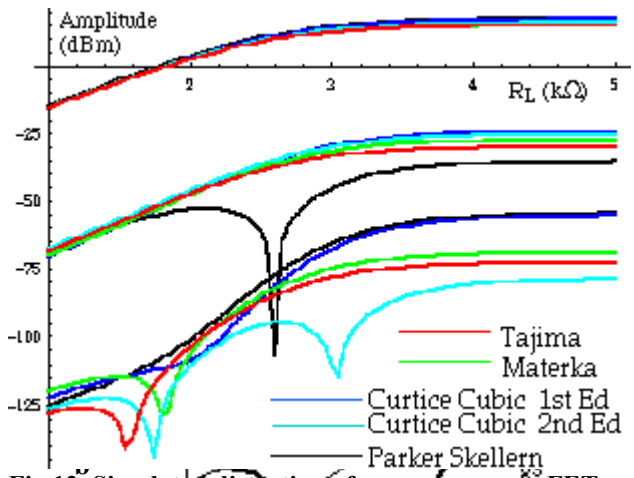


Fig 12 Simulated distortion of common-source FET amplifier using various FET models

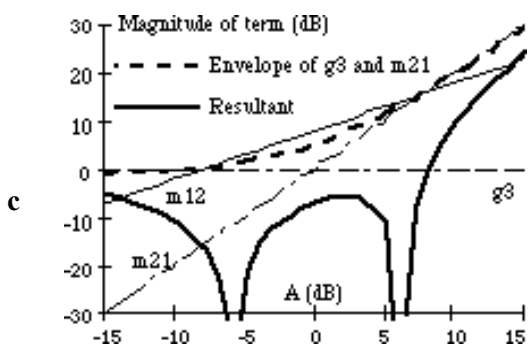


Fig 10 3rd order distortion scenarios for common-source FET amplifier

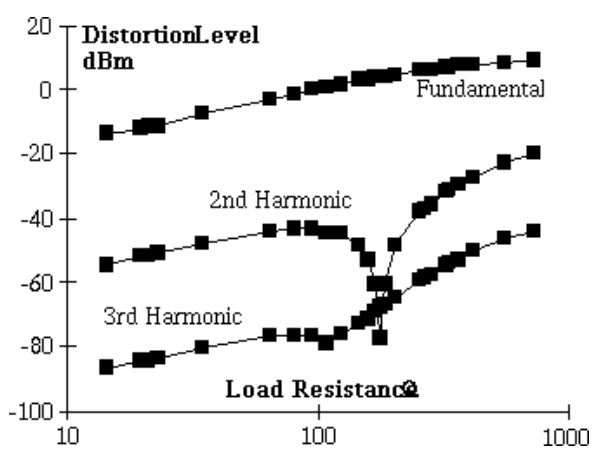


Fig 11 Measured distortion versus R_L for a common-source FET amplifier

This exercise teaches us a number of important lessons. The first is the over-riding importance of derivatives. If accurate derivatives are known for a FET at a given bias point, dependence of distortion on load resistance can easily be predicted as we have shown. Secondly, the success a model achieves in accurate prediction of distortion depends on the accuracy of its implied derivatives at the chosen bias point. Distortion is a very sensitive function of the derivative magnitudes and signs, as we have seen, and even small errors in implied derivatives can lead to totally erroneous distortion prediction. In extreme cases, it may be better not to use a FET model at all, but instead measure the derivatives of the FET and use them as a basis for design and simulation. Finally, the Parker Skellern model tends to give good predictions of the form of the derivatives with bias and reasonable predictions of their magnitudes in most cases and is therefore recommended for nonlinear simulation work.

5. SIMPLE FET MODELS WITH REALISTIC DERIVATIVES USING THE SOFT PINCH-OFF FUNCTION

We have concluded that only very accurate and hence complex FET models, such as the Parker Skellern model, can provide accurate prediction of nonlinear behaviour over the full range of bias points and load conditions. In this section, we show that, provided that certain restrictions are accepted and certain principles are applied, it is possible to use surprisingly simple models to give surprisingly good results. The importance of this is that it makes hand analysis possible and therefore facilitates the initial stages of nonlinear circuit design and also could allow the development of circuit synthesis techniques to meet linear, or nonlinear, requirements.

We begin with the simplest large signal model for the saturation region known as the square law model

$$I_d = \beta\left(V_{gs} - V_T\right)^2 \qquad (10)$$

where β is the transconductance factor and v_T is the threshold voltage. For simplicity, we neglect output conductance, ie dependence of I_d on V_{ds}. I_d is shown plotted against V_{gs} in Fig 13a for V_T = -2 V and β = 1 AV^{-2}. We also show in Fig 13a, the first, second and third derivative of I_d with respect to V_{gs}, with scaling factors of 1, 1/2 and 1/6, as in (5), to yield the

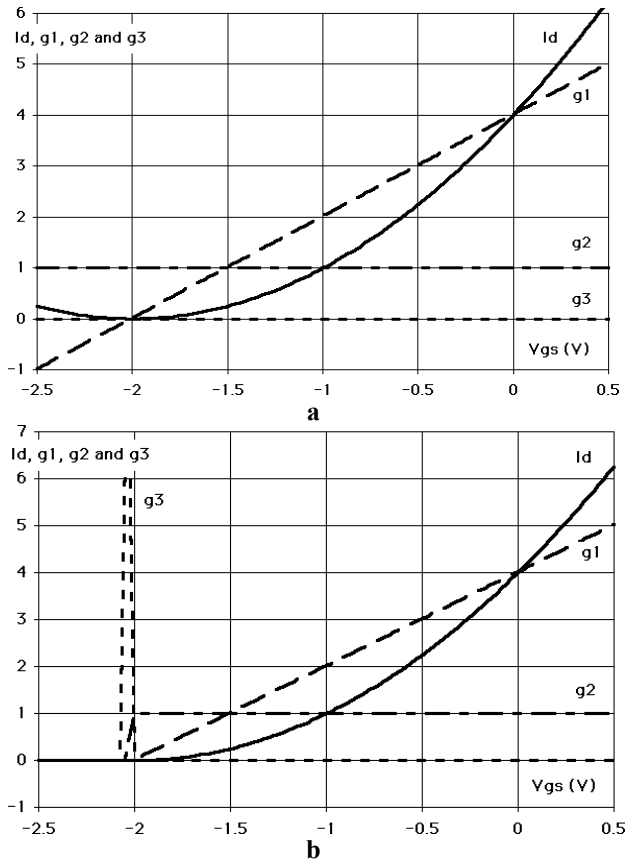


Fig 13 Plots of I_D, g_1, g_2 and g_3 versus V_{GS} for various FET models (a) square law (b) truncated square law

coefficients g_1, g_2 and g_3 in (2),. These derivatives can be compared with those for a real FET shown in Fig 4a, and it can be seen that the derivatives in Fig 13a are totally unrealistic; g_1 is predicted to be a linear function, g_2 a constant and $g_3 = 0$, implying that the FET has no 3rd order distortion. Another problem is that below the threshold voltage (V_T = -2 V), the drain current rises and g_1 continues its linear descent, becoming negative. The non-physical behaviour of I_D and g_1 for $V_{gs} < V_T$ can be overcome by introducing a switching function.

$$I_d = \beta\left(V_{gs} - V_T\right)^2 \quad \text{for} \quad V_{gs} \geq V_T$$
$$I_d = 0 \quad \text{for} \quad V_{gs} < V_T \qquad (11)$$

I_d and the derivatives for this case are shown plotted in Fig 13b. Although I_d and the derivatives are now zero for $V_{gs} < V_T$, the effect has been to introduce a gradient discontinuity in g_1, a step discontinuity in g_2 and a spike discontinuity in g_3. Since (11) is not differentiable, the height of the spike discontinuity in g_3 depends on the step size used in the numerical differentiation. Also, the form of the derivatives for $V_{gs} > V_T$ is the same as in Fig 13a and therefore an equally bad match to the real derivatives in Fig 4a. So the idea of a switching function, as implemented in (11) is not helpful.

The region of the FET characteristic curves for which $V_{gs} < V_T$ where I_d is close to zero is called the pinch-off region, since the depletion region under the FET gate is filling the whole channel region preventing any current flow. In the pinch-off region the current of a real device is not zero but only very small and measurements can be made to show that they follow an exponential dependence on V_{gs}. Thus in a real device, the transition between normal conduction and pinch-off is not sudden, as implemented in the switching function in (11), but smooth and gradual. This gradual pinch-off, or soft pinch-off, can be implemented using the soft pinch-off function (which is part of the Parker Skellern FET model [11])

$$V_{gs}' - V_T = V_{ST} \ln\left| e^{\left(V_{gs} - V_T\right)/V_{ST}} + 1 \right| \qquad (12)$$

V_{gs} is the gate source voltage of the FET and V_{gs}' replaces V_{gs} in (10). Equation (12) is plotted in Fig 14. It can be seen that, when $V_{gs} - V_T > 0$, $V_{gs}' - V_T$ follows $V_{gs} - V_T$ closely. However, when $V_{gs} - V_T < 0$, then $V_{gs}' - V_T$ converges smoothly to zero. The parameter V_{ST} determines the softness of the soft pinch-off effect; for a FET a typical value is 0.07. Unlike the switching function in (11) the soft pinch-off function in (12) is differentiable.

In Fig 15a, we show a plot of I_d and the derivatives for the square law model of (10) with V_{gs} replaced by V_{gs}' and V_{gs}' determined by the soft pinch-off expression of (12). It can be seen that as for the real device, I_d and all the derivatives now fall to zero around the threshold voltage. Comparison with the derivatives of the real device in Fig 4a shows that the model derivatives are now much more realistic. g_2 now has a slight peak and g_3 becomes negative, although these effects are less pronounced than for the real device.

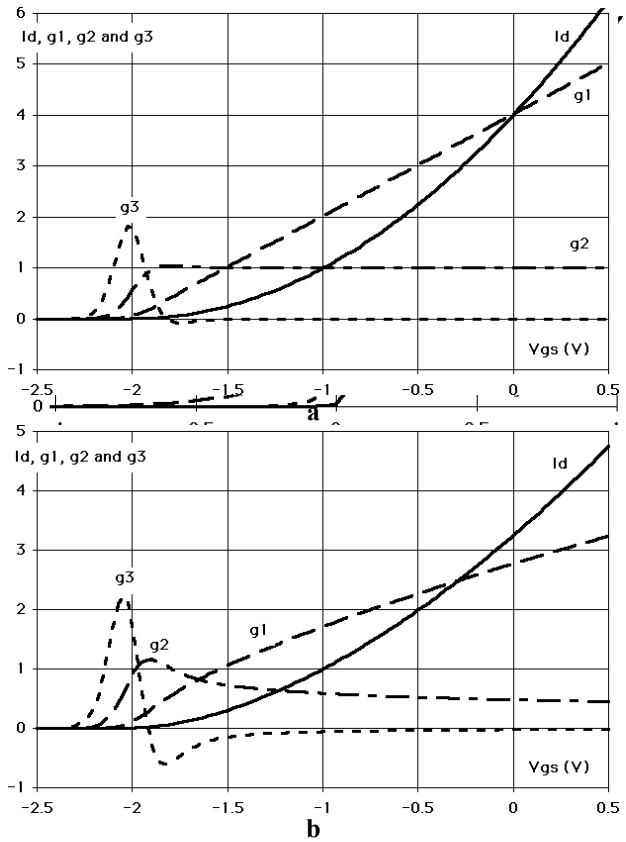


Fig 15 Plots of I_D, g_1, g_2 and g_3 versus V_{gs} for FET models with soft pinch-off (a) square law (b) Q = 1.7 law

The final stage in this study of simple models is to replace the power of 2 in (10) by a general parameter Q

$$I_d = \beta\left(V_{gs}^{'} - V_T\right)^Q \qquad (13)$$

In Fig 15b, we show a plot of I_d and the derivatives for the model of (13) with Q = 1.7 and V_{gs}' determined by the soft pinch-off expression of (12). It can be seen that the effect of changing Q from 2 to 1.7 has been to emphasise the peak in g_2 and hence increase the negative peak in g_3 to the point where the derivatives in Fig 15b are a good match to the derivatives of the real device in Fig 4a. Since (12) is differentiable, these model derivatives are relatively insensitive to the step size used in the numerical differentiation.

In this section, we have identified some key principles of nonlinear FET modelling:-

- Although a device is not normally used in the pinch-off region, modelling of the pinch-off region affects nonlinear behaviour significantly in the normal conduction region and it is vital to implement a realistic soft-pinch-off using an expression such as (12).
- The square law model is usually not accurate and a power value other than 2 will be necessary, especially for high frequency short-channel FETs.

We have ignored dependence of I_d on V_{ds}, ie we have assumed that FET output conductance is zero. This is generally a reasonable assumption for the initial design of a circuit because the transconductance derivatives play the major role in determining distortion behaviour. If desired, non-zero FET output conductance can be taken into account to a first order approximation by replacing V_{gs} in (12) by $V_{gs} + \gamma V_{ds}$.

The model we have developed in (13) and (12) is sufficiently simple that it can be used for hand analysis of circuits. For demanding design situations where the distortion of the common-source FET is too high the model could make it possible to synthesise new circuits or design techniques with reduced distortion. One example of such a design technique is derivative superposition, and we consider this approach in the next section.

6. IMPROVING THE LINEARITY OF THE COMMON-SOURCE FET USING DERIVATIVE SUPERPOSITION

We have seen that, as an architecture for high frequency linear amplifiers, the common-source amplifier is attractive because significant frequency dependent distortion does not occur and the distortion behaviour up to high frequencies is governed by the FET derivatives. For some applications, where very low levels of distortion are specified or where large signal levels have to be used, as in a power amplifier, the distortion produced by the common-source FET may be too high and design techniques are required to reduce it. Derivative superposition is one such technique and will be described in this section. In most communications applications, the 3rd order distortion is much more of a problem than 2nd order distortion because 2nd order distortion has the effect of introducing additional frequency components far removed from the carriers and therefore out side the system bandwidth. 3rd order distortion, on the other

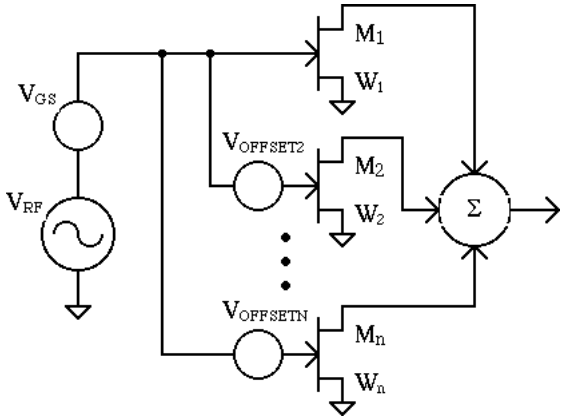


Fig 16 Derivative superposition amplifier architecture

hand, causes distortion components close to the carriers making filtering impossible. Thus a very important need is to find a way to reduce the 3rd order distortion of the common-source FET amplifier.

Consider the 3rd order derivative (g_3) for a typical MESFET shown plotted in Fig 4a. The form of the curve has a positive peak followed by a zero and then a negative peak. Derivative superposition, in its most general sense involves taking a number of such derivative curves, shifting them along the V_{gs} axis by arbitrary amounts and applying scaling to the amplitude of the derivatives in order that the composite derivative obtained by adding or superimposing the derivatives meets some linearity (or nonlinearity) requirement [12]. In the most common applications, we start with a main device (say the one having the derivative structure in Fig 4a) and add a 2nd set of derivatives shifted and scaled so that the positive peak of one g_3 derivative lies on top of the negative region of the original g_3 derivative, causing a cancellation of the derivative. Several curves may be added to obtain low g_3 over a wide range of V_{gs}, ie for a range of small and large signal levels.

A generic circuit structure for a derivative superposition (DS) amplifier is shown in Fig 16. A number of common-source FETs share the same RF input signal V_{RF} and overall bias V_{GS} applied to their gates but each FET has a separate gate bias offset $V_{OFFSETi}$ which effectively shifts its derivatives along the V_{gs} axis as required. Each FET has a specified gate width, to scale the amplitude of its derivatives appropriately, and therefore the technique is most appropriate for integrated circuit implementation. The drain currents are added to form the overall output signal. This addition of the FET output currents implies addition of their individual derivatives to obtain an overall set of derivatives for the composite device. The output summing may be by direct connection, by use of a tapped transmission line, a hybrid transformer or any other suitable method. Since all the FETs are in common-source configuration, frequency dependent distortion is not a problem.

We shall briefly consider two examples of DS amplifiers. The first circuit consists of 4 nominally identical discrete HEMTs on a printed circuit with their drains connected to the output via attenuators to represent device width scaling [12]. The gates are connected to the common signal source via capacitors and to their respective bias voltages via resistors. The starting point was to measure the derivatives of one of the discrete HEMTs alone and the result is shown in Fig 17 (white symbols). Comparison with Fig 4b shows that the derivatives for the HEMT are more complex than those for the MESFET since, for high V_{gs}, g_1 falls, causing a null in g_2 and two nulls in g_3. Using these measured derivatives, offset voltages and scaling factors for the 4 HEMTs of the DS amplifier were chosen in order to obtain a low combined value for g_3 over a range of V_{gs} values. The measured derivatives for the 4-HEMT DS amplifier are also shown in Fig 17 (black symbols). It can be seen that it has been possible to obtain a significant reduction in 3rd order distortion for gate voltages between about -0.3 and 0 V. Notice that since the 2nd order derivative is always positive this form of derivative superposition tends to increase 2nd order distortion, but this is usually not a problem because its effects are out of band. Notice that the superposition operation has significantly increased the fundamental. If the DS amplifier is biased in the centre of the low distortion range, about V_{GS} = - 0.15 V, then for small signal amplitudes (small excursions about the bias point), the benefits of the low distortion design will be obtained. However, as the amplitude of the input signal is increased, the part of the derivative beyond the low distortion region will start to come into play and distortion will increase. This can be observed in Fig 18 where we have compared distortion against signal

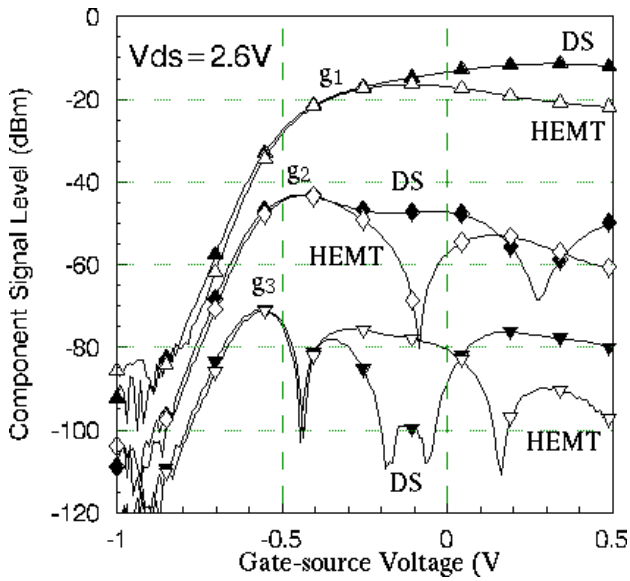


Fig 17 Comparison of measured derivatives g_1, g_2 and g_3 for a 4-HEMT DS amplifier and for a single HEMT

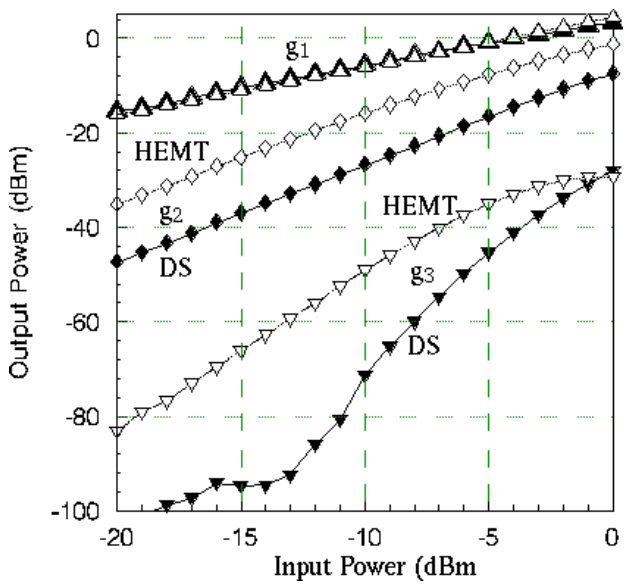


Fig 18 Comparison of fundamental and 2nd and 3rd order intermodulation distortion against input power for 4-HEMT DS amplifier and single HEMT

level for the DS amplifier and for the single HEMT. This shows clearly the considerable improvement at low and medium signal levels being eroded as the signal level is increased towards the compression point. This approach is fine for small and medium signal amplifiers but is not very attractive for power amplifiers because power amplifiers have to be operated with large signal levels close to compression

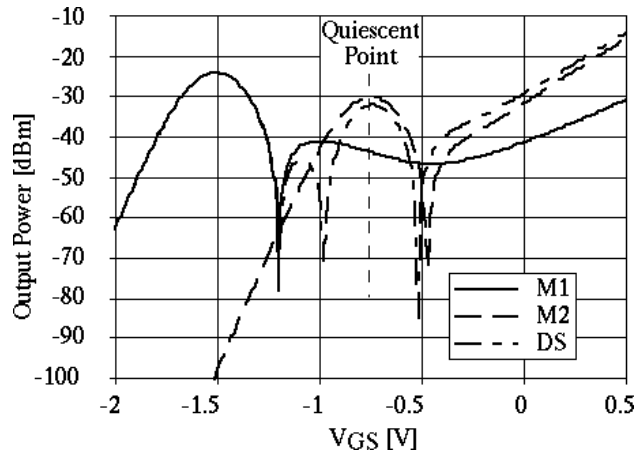


Fig 19 3rd order derivatives for phase reversal form of derivative superposition

to obtain reasonable efficiency and it is at these high power levels that low distortion is needed. To meet such a requirement would need many devices and would represent very inefficient use of the large total gate width.

In order to meet the need for power amplifiers with low 3rd order distortion at high power levels, an alternative form of derivative superposition has been developed [13]. Rather that reducing the amplitude of the overall 3rd order derivative around the bias point, as illustrated in Fig 17, in the alternative technique the aim is not to reduce the amplitude of the derivative but to introduce 180° phase shifts in the derivative each side of the bias point. The way in which this works in the derivative domain is illustrated in Fig 19. In this case there are just two devices, a main device (M1) and an auxiliary device (M2). The auxiliary device has a greater width than the main device and this causes a change in sign of the overall derivative (labelled DS) at the quiescent point of Vgs = -0.75 V and 180° sign changes each side, shown by nulls in the DS curve around Vgs = -0.5 V and -1 V. As the signal level is increased, the derivative will change sign for parts of the cycle as the peaks traverse the curves beyond the sign changes. It is possible to design the circuit so that a net zero g_3 derivative is obtained for a specified large signal level. Measured results for a MMIC chip are shown in Fig 20 where we have plotted carrier-to-interference ratio (C/I) against input power for the DS amplifier and for a single device biased at different quiescent points, corresponding to operation in classes A, AB and B.

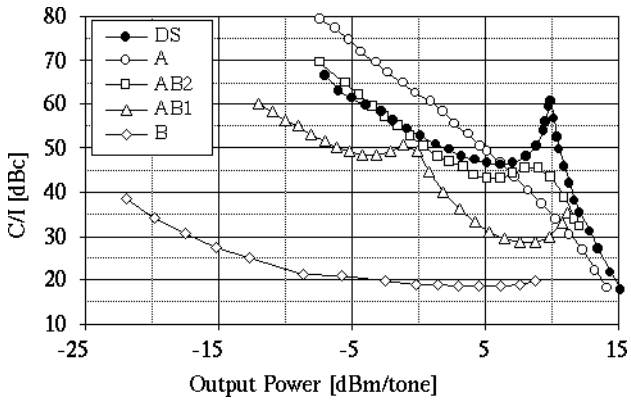


Fig 20 Measured 2-tone C/I ration for DS amplifier and single FET biased at different points

The peak in C/I for the DS amplifier at an output power of 10 dBm corresponds to the null in 3rd order distortion caused by the phase reversal technique. It was shown that the DS amplifier provides the best compromise between distortion and efficiency compared to all the single FET modes considered at the cost of an increase in total gate width. In many communications system scenarios, a combination of high efficiency and low distortion is a very attractive design option.

7. CONCLUSIONS

The authors have presented some concepts which have been found to be useful for the nonlinear analysis and design of FET circuits. The techniques are sufficiently realistic that they allow the design of circuits needed for today's and future communication systems, providing for example potentially attractive solutions to specifications on high linearity and high efficiency for power amplifiers. On the other hand, the techniques are sufficiently simple that they provide considerable engineering insight to problems in linear and nonlinear circuits and, in many cases, allow hand calculations to be performed. The possibility of hand calculations is important for two reasons. Firstly, in the early stages of the circuit design process, specifications on nonlinearity can be included at the outset and designed for, rather than adopting a trial and error approach. Secondly, the availability of simple expressions providing a realistic description of the nonlinearity of high frequency FETs could potentially lead to the development of new formal synthesis techniques for circuit design taking nonlinearity into account.

REFERENCES

[1] S.A.Maas, Nonlinear Microwave circuits, Artech House 1988

[2] J. C. Pedro and J. Perez, 'Accurate simulation of GaAs MESFETs intermodulation distortion using a new drain-source current model', *IEEE MTT Trans*, 1994, Vol. 42, No. 1, pp. 25-33

[3] D.R.Webster, D.G.Haigh, G.Passiopoulos, A.E.Parker, "Distortion Generation and Reduction in Short Channel FET Circuits", Chap 24 of *Low-power HF microelectronics*, Ed G. Machado, IEE Books, Feb 1996

[4] J B Scott, D R Webster, A E Parker, D G Haigh, J P M Proctor, and O J Ridler, Extraction Of MESFET Distortion Behaviour, 1994 Asia Pacific Microwave Conference, Tokyo, Dec 1994

[5] I. Angelov, H. Zirath and N. Rorsmann, "Validation of a Nonlinear Transistor Model by Power Spectrum Characteristics of HEMTs and MESFETs", *IEEE MTT Trans* Vol. 43, No. 5, May, 1995, pp 1046-1052

[6] J B Scott, J G Rathmell, A E Parker and M Sayed, Pulsed device measurements and applications, IEEE *MTT Trans* ,vol 44, pp 2718-2723, Dec 1996

[7] D.R.Webster, A.E.Parker, D.G.Haigh, "Observations on the Effect of Bias on Pulsed I-V Characteristics of MESFETs and HEMTs", IEEE Workshop on High Performance Electron Devices for Microwave and Optoelectronic Applications, 23th-24th November 1998

[8] D R Webster, D G Haigh, A E Parker and J B Scott, Device Circuit Interaction In The Common Source GaAs MESFET Amplifier , 1994 IEEE International Symposium on Circuits and Systems, London, May 30th-June 2nd 1994, pp 5-241 - 244

[9] R Balmford, W Redman-White, A New CMOS Current Mirror Circuit with Reduced Harmonic Distortion for High Bandwidth Applications, 1993 ECCTD, Davos, Switzerland, Aug 30th-Sep 3rd, Ed H Dedieu, Elsevier

[10] A E Parker and J B Scott, "Intermodulation Nulling in GaAs MESFETs", *IEE Electronics Letters*, vol. 29, no. 22, pp. 1961-1962, October 28, 1993

[11] A.E.Parker, D. J. Skellern, "A Realistic Large-Signal MESFET Model for SPICE", *IEEE MTT Trans.* Vol. 45, No. 9, September 1997 pp. 1563-1571

[12] D. Webster, J. Scott, D. Haigh, "Control of Circuit Distortion by the Derivative Superposition Method", *IEEE Microwave and Guided Wave Letters*, Vol. 6, No. 3, March 1996, pp. 123-125

[13] D R Webster, G Ataei and D G Haigh, Low distortion MMIC power amplifier using a new form of

derivative superposition, IEEE *MTT Trans*, vol 49, no 2, Feb 2001, pp 328 - 332