A Capacitance-Compensation Technique for Improved Linearity in CMOS Class-AB Power Amplifiers

Chengzhou Wang, *Member, IEEE*, Mani Vaidyanathan, *Member, IEEE*, and Lawrence E. Larson, *Fellow, IEEE*

***Abstract*—A nonlinear capacitance-compensation technique is developed to help improve the linearity of CMOS class-AB power amplifiers. The method involves placing a PMOS device alongside the NMOS device that works as the amplifying unit, such that the overall capacitance seen at the amplifier input is a constant, thus improving linearity. The technique is developed with the help of computer simulations and Volterra analysis. A prototype two-stage amplifier employing the scheme is fabricated using a 0.5-μm CMOS process, and the measurements show that an improvement of approximately 8 dB in both two-tone intermodulation distortion (IM3) and adjacent-channel leakage power (ACP1) is obtained for a wide range of output power. The linearized amplifier exhibits an ACP1 of −35 dBc at the designed output power of 24 dBm, with a power-added efficiency of 29% and a gain of 23.9 dB, demonstrating the potential utility of the design approach for 3GPP WCDMA applications.**

***Index Terms*—Adjacent-channel power ratio (ACPR), class-AB power amplifiers, CMOS, intermodulation distortion, linearity, radio-frequency (RF) circuits, Volterra series, WCDMA.**

I. Introduction

PRESENTLY, there is widespread interest in pursuing a single-chip, handheld, wireless transceiver implemented in complementary, metal-oxide-semiconductor (CMOS) technology. A key component of such a system would be the power amplifier (PA), and several workers have recently described implementations of CMOS PAs. However, most of these designs, such as those described in [1]–[4], were intended for constant-envelope modulation schemes, and are hence intrinsically very nonlinear. For nonconstant-envelope modulation schemes, nonlinearity can cause severe regrowth in the spectral sidebands and an increase in the transmitted error-vector magnitude. In such cases, stringent requirements are placed on amplifier linearity. At the same time, to prolong battery life, the power amplifier must also operate at reasonable levels of efficiency.

To meet the simultaneous requirements of high linearity and reasonable efficiency, power amplifiers in nonconstant-envelope systems are often operated in a class-AB mode; the linearity can be superior to that in class-B or higher operation and the efficiency is superior to that in class-A operation. Of particular importance is the nonlinearity of the class-AB amplifier; while more linear than a class-B or higher amplifier, the intrinsic linearity obtained in class-AB operation is often still insufficient to meet required specifications. While many external linearization techniques are known [5, ch. 9], they are complex and inconvenient for handset applications, and it is thus important that the intrinsic amplifier linearity be made as high as possible.

In this work, it is shown that the gate-source capacitance of a MOS device is a major source of nonlinearity that can limit the performance of a CMOS class-AB power amplifier. A simple technique to compensate the nonlinearity is suggested, and simulations and experiments on a prototype amplifier are used to demonstrate its effectiveness. Although the idea has been discussed in [6], this work presents a more detailed and rigorous analysis, along with the design, implementation, and measurement details.

In Section II, computer simulations are used to identify the role of the gate-source capacitance in limiting the linearity. In Section III, a scheme to compensate this nonlinearity, and hence improve overall amplifier linearity, is developed. In Section IV, the effectiveness of the scheme is demonstrated through experiments. Section V summarizes the conclusions.

II. Distortion Effects of the Gate-Source Capacitance

A. Simplified Model

Fig. 1(a) shows a *highly* simplified model for an NMOS device working as a class-AB amplifier; only signal quantities are shown. The input signal current is i_s, the input-matching network (which includes the source admittance) is I, the output-matching network is O, and the load resistance is R_L. The transistor itself is modeled using only the quasistatic, drain-source signal current $i_{\rm dsn}(v_{\rm gs}, v_{\rm ds})$, which is a function of both the gate-source and drain-source signal voltages, $v_{\rm gs}$ and $v_{\rm ds}$, and the following device capacitances: the gate-body capacitance, $C_{\rm gbn}$; the gate-source capacitance, $C_{\rm gsn}$; and the gate-drain capacitance, $C_{\rm gdn}$. This model assumes that the intrinsic source and body (substrate) are connected together, and omits a number of elements, including the gate, drain, and

Manuscript received October 21, 2003; revised July 15, 2004. This work was supported by the UCSD Center for Wireless Communications, its member companies, and the State of California on a UC Discovery Grant.

C. Wang was with the Center for Wireless Communications, Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA. He is now with Marvell Semiconductor, Sunnyvale, CA 94089 USA (e-mail: czwang@marvell.com).

M. Vaidyanathan was with the Center for Wireless Communications, Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA. He is now with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4 Canada (e-mail: maniv@ece.ualberta.ca).

L. E. Larson is with the Center for Wireless Communications, Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail: larson@ece.ucsd.edu).

Digital Object Identifier 10.1109/JSSC.2004.835834

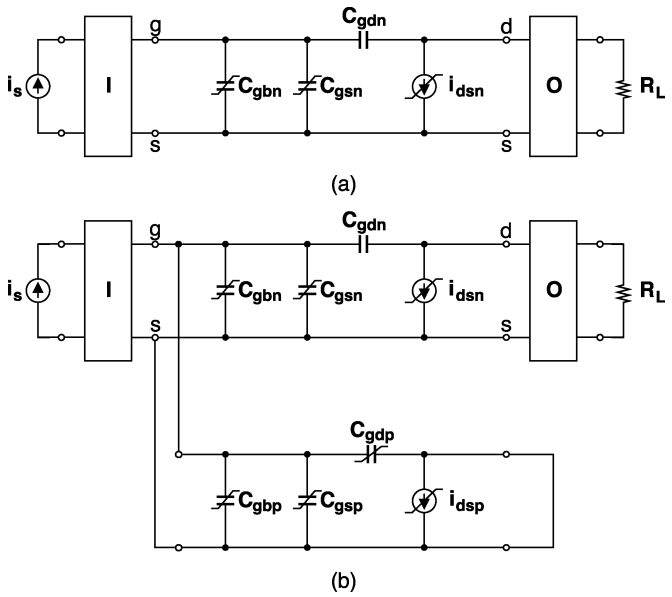


Fig. 1. Simplified models of CMOS class-AB power amplifiers. (a) NMOS device working alone. (b) NMOS device along with a PMOS device used to provide a compensating input capacitance. Nonlinear elements are marked in the usual fashion.

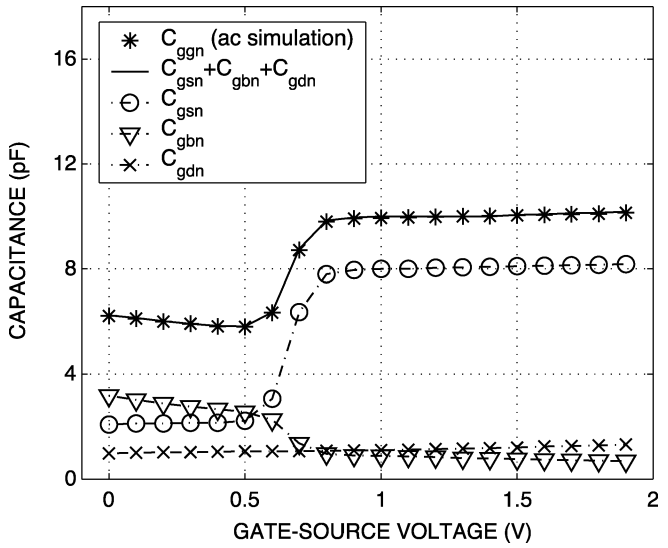


Fig. 2. Plots of the simulated NMOS device capacitances as a function of gate-source voltage, for a fixed drain-source voltage of 3.3 V. The device length and width are 0.5-μm and 3 mm, respectively, and the device threshold voltage is $V_{\mathrm{Tn}} = 0.66$ V.

source resistances, a substrate network, and the capacitance between drain and source (although the linear parts of some of these elements could be absorbed into I and O). These simplifications are justified, since the purpose of the model is merely to illustrate the main sources of nonlinearity under class-AB operation. For accurate simulation results needed in final designs, however, it should be noted that radio-frequency (RF) MOS models should include the omitted elements [7]–[11]. Fig. 1(b) will be discussed in Section III-A.

B. Capacitance Components

Shown in Fig. 2 are plots of the simulated NMOS device capacitances as a function of gate-source voltage, for a fixed drain-source voltage. The variation of the capacitances with drain-source voltage can be neglected as long as the device remains in saturation [12, ch. 8]; this is typically ensured in power-amplifier design, since appreciable distortion would otherwise occur when the device transits across the knee that exists in the current-voltage characteristics between the saturation and triode regions. The device is from IBM's SiGe5AM technology, and the plots were obtained using the well-known SPECTRE circuit simulator and the associated commercial MOS model released by IBM; the model employs BSIM3v3.2 as an intrinsic subcircuit, along with extrinsic parasitics to account for RF effects [13, p. 53].

Fig. 2 confirms that the total capacitance seen looking into the gate, as found from an ac simulation at each gate-source voltage, $C_{\mathrm{ggn}} \equiv \mathrm{Im}\{y_{11}\}/\omega$, where y_{11} is the short-circuit, common-source input admittance and $\omega = 2\pi$ (1.95 GHz) is the radian frequency, is equal to the sum of the individual capacitance components mentioned earlier: $C_{\mathrm{ggn}} = C_{\mathrm{gsn}} + C_{\mathrm{gbn}} + C_{\mathrm{gdn}}$. This is to be expected when the device's parasitic resistances are negligible [9, eq. (9)], and helps to validate the simplified model of Fig. 1(a). More importantly, Fig. 2 shows that

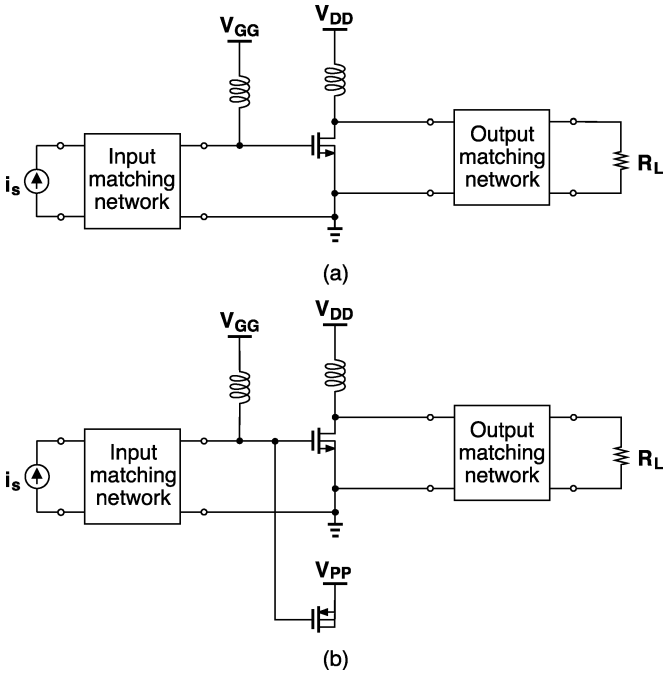


Fig. 3. Simplified schematics of class-AB amplifiers used to illustrate the impact of the gate-source capacitance on linearity. The basic amplifier is in (a), and the linearized version is in (b). The NMOS and PMOS devices are the same as those in Figs. 2 and 6, respectively.

while C_{gdn} and C_{gbn} are relatively constant, C_{gsn} varies substantially as the device transits from an "off" (below threshold) to an "on" (above threshold) state. While C_{gsn} as plotted includes both intrinsic and extrinsic parts, almost all of this variation can be traced to a change in the intrinsic part [9, Fig. 3(a)]. This variation is particularly germane for class-AB operation, because the transition in the capacitance occurs at the device's threshold voltage, close to where it is typically biased. As will

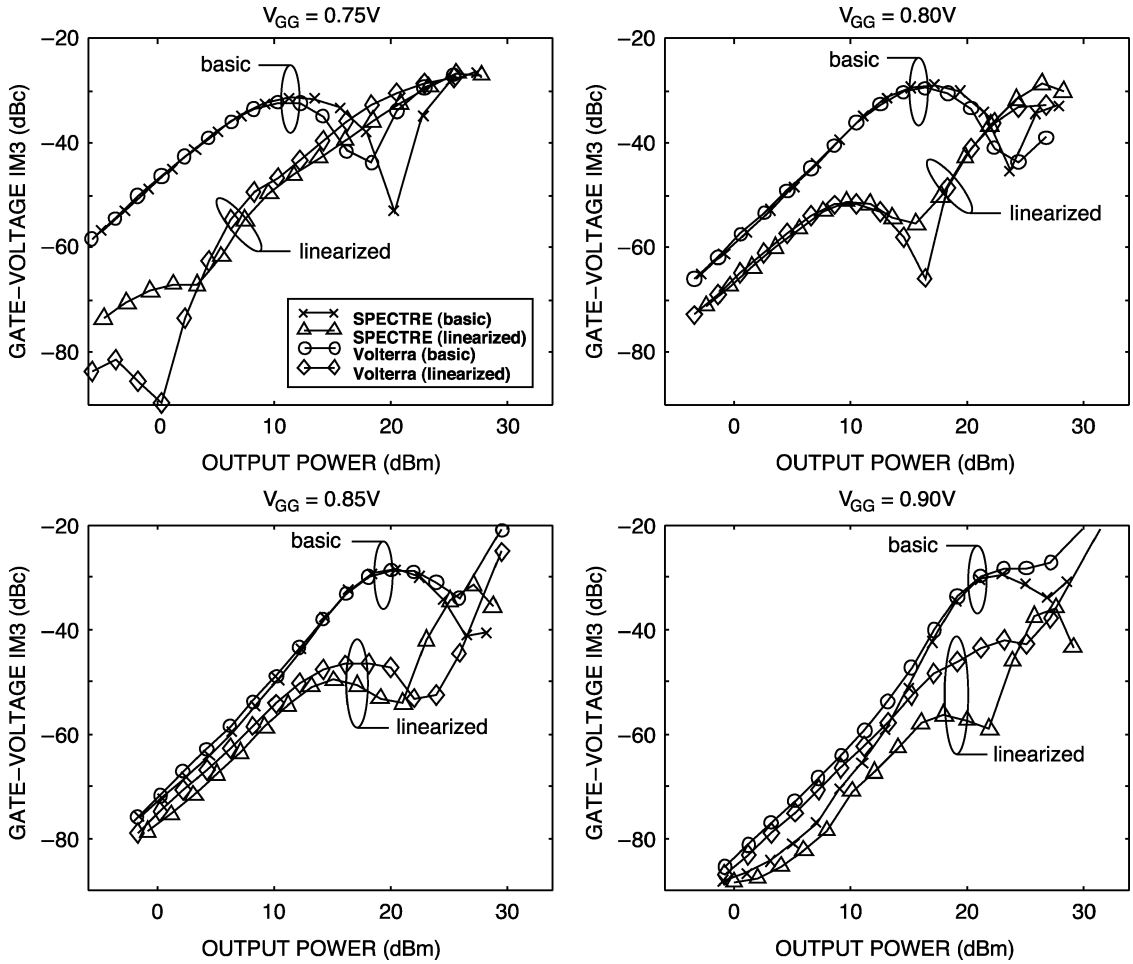


Fig. 4. Third-order intermodulation distortion at $2\omega_1 - \omega_2$ versus peak-envelope output power, at various gate bias voltages. The circuits are the basic and linearized class-AB amplifiers in Figs. 3(a) and 3(b), respectively. These plots are for the distortion *in the gate voltage*. Values from both simulation (using SPECTRE) and Volterra theory [using (10)–(16)] are shown. In each case, $V_{\rm DD} = 3.3$ V.

be shown, the change in capacitance leads to substantial distortion at the gate, and subsequently at the drain, and this can limit overall amplifier linearity.

C. Impact on Linearity

In order to illustrate the impact of the gate-source capacitance on the linearity of a class-AB amplifier, the simplified circuits of Fig. 3 will be used; the circuit in Fig. 3(a) is a basic class-AB amplifier, and the circuit in Fig. 3(b) includes additional circuitry to "compensate" or "linearize" the nonlinear capacitance between the gate and source that will be explained in Section III-A. In addition to providing appropriate matches at the fundamental frequency, the input and output matching networks include short-circuit terminations at the harmonic frequencies, which we found helped overall linearity; they also helped to boost the fundamental output power [14, p. 384]. The input network includes the source admittance, chosen in this case to represent the output admittance of a driving class-A stage. In fact, the circuits in Fig. 3 are simplified versions of actual two-stage, class-AB amplifiers that were built and tested, and which will be described in Section IV.

Figs. 4 and 5 show SPECTRE simulations of the third-order intermodulation distortion (IM3) at $2\omega_1 - \omega_2$ for a two-tone input at frequencies $\omega_1 = 2\pi$ (1.96 GHz) and $\omega_2 = 2\pi$ (1.94 GHz), at the gate and drain, respectively; note that the drain IM3 is equivalent to the load IM3, since O and R_L are linear and $2\omega_1 - \omega_2 \approx \omega_1$. As shown, the basic amplifier of Fig. 3(a) incurs substantial distortion at both the gate and drain; it will be proven in Section III-B that most of this distortion is due to the change in gate-source capacitance as the device turns on and off during class-AB operation. On the other hand, Figs. 4 and 5 show that much better performance can be obtained by employing the scheme illustrated in Fig. 3(b), where a compensating nonlinear capacitance is added at the input. The details of this compensation scheme will be discussed next.

III. Compensation Technique

A. Basic Idea

Shown in Fig. 6 are plots of the simulated device capacitances of a PMOS transistor as a function of its gate-source voltage, with the drain-source voltage held at zero. As shown, while $C_{\rm gbp}$ is relatively constant, $C_{\rm gdp}$ and $C_{\rm gsp}$ change[1] from a high to a low value as the device transits from an "on" to an "off" state. This behavior is exactly complementary to that of $C_{\rm gsn}$ in

[1]Since the drain-source voltage is zero, $C_{\rm gdp}$ should equal $C_{\rm gsp}$; the small discrepancy occurs due to an implementation limit in BSIM3v3 [15, ch. 4].

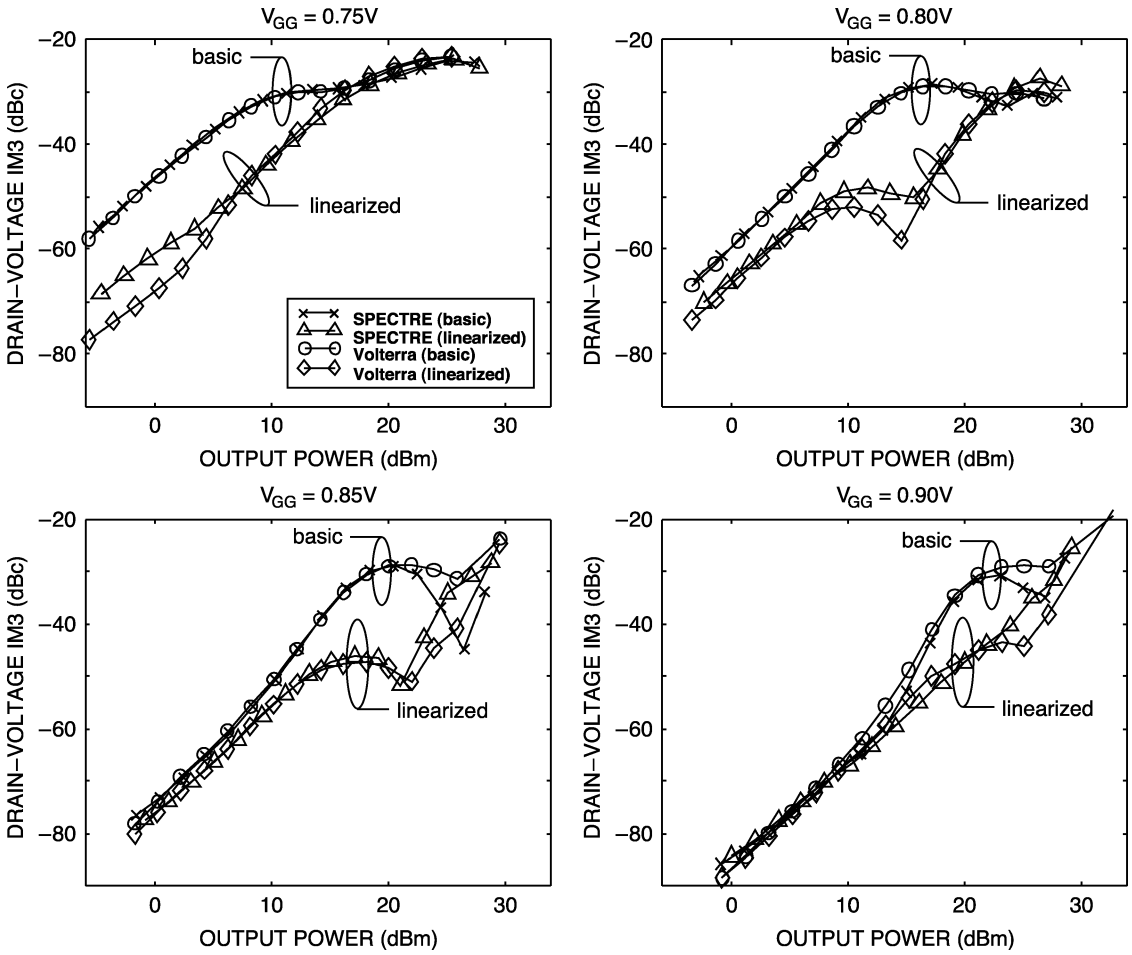


Fig. 5. Third-order intermodulation distortion at $2\omega_1 - \omega_2$ versus peak-envelope output power, at various gate bias voltages. The circuits are the basic and linearized class-AB amplifiers in Figs. 3(a) and 3(b), respectively. These plots are for the distortion *in the drain voltage*. Values from both simulation (using SPECTRE) and Volterra theory [using (10)–(16)] are shown. In each case, $V_{\rm DD} = 3.3$ V.

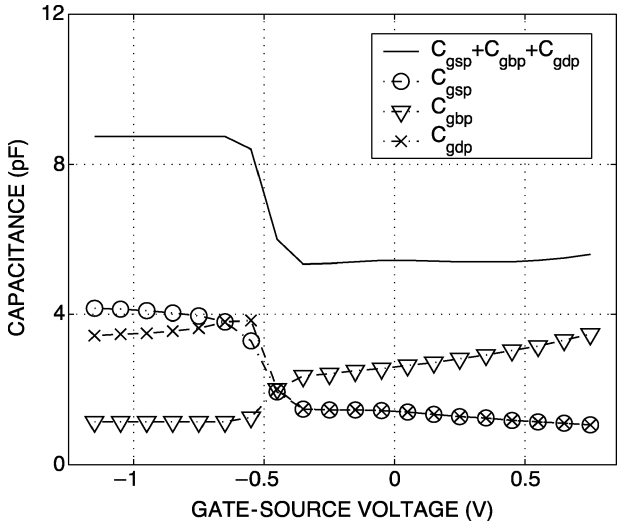


Fig. 6. Plots of the simulated device capacitances of a PMOS transistor as a function of its gate-source voltage, with its drain-source voltage held at zero. The device length and width are 0.5-μm and 2 mm, respectively, and the device threshold voltage is $V_{\rm Tp} = -0.49$ V.

Fig. 2. Therefore, it should be possible to "linearize" or "compensate" $C_{\rm gsn}$ with the aid of a PMOS device. The basic idea is simply to place a PMOS device alongside the NMOS device as illustrated in Fig. 3(b); the model for the situation is shown in Fig. 1(b). When the PMOS device is properly biased and sized, the total capacitance $C_{\rm ggn} + C_{\rm ggp}$ seen at the NMOS gate will be a constant, which reduces the distortion generated at the gate, and subsequently at the drain.

Since the change in the NMOS and PMOS capacitances occurs at their respective threshold voltages, it is clear that the PMOS bias voltage $V_{\rm PP}$ in Fig. 3(b) should be

$$V_{\rm PP} = V_{\rm Tn} - V_{\rm Tp}. \tag{1}$$

Neglecting $C_{\rm gbn}$ and $C_{\rm gbp}$ and extrinsic contributions to the capacitances, an appropriate figure for the sizing of the PMOS device can be obtained by noting that the NMOS device switches between weak and strong inversion, and the PMOS device works in the triode region. Therefore [12, sec. 8.3.2], the changes in NMOS and PMOS capacitances are approximately

$$\Delta C_{\rm ggn} \sim \Delta C_{\rm gsn} \approx \frac{2}{3} W_n L_n C_{\rm ox\,n} \tag{2}$$

and

$$\Delta C_{\rm ggp} \sim \Delta (C_{\rm gsp} + C_{\rm gdp}) \approx 2 \left[\frac{W_p L_p C_{\rm ox\,p}}{2} \right] = W_p L_p C_{\rm ox\,p} \tag{3}$$

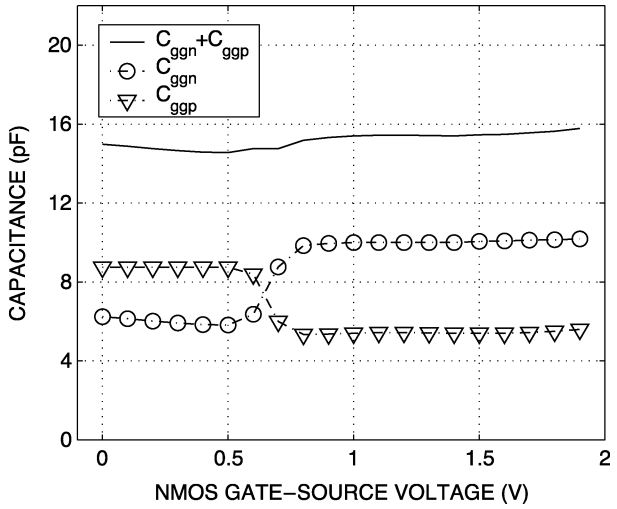


Fig. 7. Plots of simulated $C_{\rm ggn}$, $C_{\rm ggp}$, and the sum $C_{\rm ggn} + C_{\rm ggp}$ for the NMOS and PMOS devices of Figs. 2 and 6.

where W_n and L_n, and W_p and L_p, are the widths and lengths of the NMOS and PMOS devices, and $C_{\rm ox\ n}$ and $C_{\rm ox\ p}$ are their oxide capacitances, respectively. Assuming the changes in the capacitances are abrupt, we then require

$$\frac{\Delta C_{\rm ggn}}{\Delta C_{\rm ggp}} \sim \frac{2}{3}\frac{W_n L_n}{W_p L_p}\frac{C_{\rm ox\ n}}{C_{\rm ox\ p}} \sim 1 \tag{4}$$

which can be used as a guide to size the PMOS device.

Fig. 7 shows plots of $C_{\rm ggn}$ and $C_{\rm ggp}$, found from $\mathrm{Im}\{y_{11}\}/\omega$, and of the sum $C_{\rm ggn} + C_{\rm ggp}$, for the NMOS and PMOS devices of Figs. 2 and 6. As shown, while both $C_{\rm ggn}$ and $C_{\rm ggp}$ vary with the NMOS gate-source voltage, the sum $C_{\rm ggn} + C_{\rm ggp}$ remains roughly constant. The small ripple that occurs in the sum at the transition point arises because the capacitances do not change abruptly; the slope of the $C_{\rm ggn}$ curve is not exactly equal (in magnitude) to that of the $C_{\rm ggp}$ curve. The ripple can be minimized by adjusting the bias and size of the PMOS device from the nominal values given by (1) and (4). Additionally, we should mention that while the use of the PMOS device does help to linearize the total gate capacitance, it also doubles its value, which will cause a decrease in overall PA efficiency since the NMOS–PMOS combination will need to be driven with a higher input power, for example, by a driver stage, which would then consume more dc power. With this tradeoff borne in mind, the key point is that the technique does improve linearity over a wide power range, which is important for nonconstant-envelope modulation schemes. The impact on the linearity can be understood with a simple Volterra analysis.

B. Volterra Analysis

Usually, Volterra analysis assumes each nonlinear element in a circuit can be described by a third-order power-series expansion in which the series coefficients depend only on the circuit's bias point. Such analysis cannot be used to describe a highly nonlinear circuit, such as a class-AB power amplifier. However, we will attempt to alleviate this problem by employing power-series expansions of order greater than three, and by allowing the series coefficients to depend on *both* the bias point *and* the RF signal power.

Defining an effective gate-source capacitance $C_{\rm eff}$, and referring to Fig. 1(a) and (b), the values of $C_{\rm eff}$ in the uncompensated and compensated cases are, respectively, as follows:

$$C_{\rm eff} = C_{\rm gbn} + C_{\rm gsn} \tag{5}$$

and

$$C_{\rm eff} = C_{\rm gbn} + C_{\rm gsn} + C_{\rm gbp} + C_{\rm gsp} + C_{\rm gdp}. \tag{6}$$

At each bias point, the RF signal power determines the range of excursion of the NMOS gate-source voltage; for simplicity, this range can be approximated to be the peak-to-peak excursion of the two-tone envelope (i.e., the envelope arising from the fundamental signal components at ω_1 and ω_2, and neglecting the much smaller harmonic and intermodulation components). With knowledge from SPECTRE of the behavior of the individual components of $C_{\rm eff}$ versus this voltage, $C_{\rm eff}$ can then be modeled as a power series. We found that a fifth-order power series would work well for all bias points and for all RF signal powers considered,[2] i.e., $C_{\rm eff}$ could always be written as follows:

$$C_{\rm eff} = c_1 + c_2 v_{\rm gs} + c_3 v_{\rm gs}^2 + c_4 v_{\rm gs}^3 + c_5 v_{\rm gs}^4 + c_6 v_{\rm gs}^5. \tag{7}$$

It is important to emphasize that when the bias point *or* RF signal power changes, the coefficients c_1 through c_6 also change, such that the expansion in (7) always traces out the appropriate $C_{\rm eff}$ versus $v_{\rm gs}$ curve.

The behavior of the large-signal, quasistatic, drain-source current $i_{\rm DSN}(v_{\rm GS}, v_{\rm DS})$ for the NMOS transistor as a function of $v_{\rm GS}$ and $v_{\rm DS}$ can be simulated with SPECTRE, and the results can be used to expand the corresponding signal current $i_{\rm dsn}$ in Fig. 1(a) and (b) as a power series. In performing the expansion, for simplicity, the dependence on the drain-source voltage is first eliminated. Referring to Fig. 3(a) and (b), this is done by *approximating* $v_{\rm DS}$ to be a superposition of the dc bias and the purely linear part of the output signal:

$$v_{\rm DS} \approx V_{\rm DD} - g_m v_{\rm gs} R_O \tag{8}$$

where g_m is the short-circuit transconductance, given by $g_m \equiv \partial i_{\rm DSN}/\partial v_{\rm GS}$ with $v_{\rm DS} \equiv V_{\rm DD}$, and R_O is the equivalent resistance (at the fundamental frequency) seen looking into the output matching network from the NMOS drain. This approximation is used *solely* for the purpose of simplifying the power-series expansion of $i_{\rm dsn}$; once the expansion is established, the true nonlinear relationship between the drain and gate voltages will be taken into account by the Volterra analysis. At each NMOS bias point $(V_{\rm GG}, V_{\rm DD})$, a given RF signal power defines the range of excursion of $v_{\rm gs}$, which is again approximated to be the peak-to-peak excursion of the two-tone envelope, and for each such excursion, the locus of points traced out by $i_{\rm DSN}(V_{\rm GG} + v_{\rm gs}, V_{\rm DD} - g_m v_{\rm gs} R_O)$ can be used to find a power series for $i_{\rm dsn}$ in terms of $v_{\rm gs}$. In this case, we found a series of order three sufficed, i.e., $i_{\rm dsn}$ could be written as follows:

$$i_{\rm dsn} = g_1 v_{\rm gs} + g_2 v_{\rm gs}^2 + g_3 v_{\rm gs}^3 \tag{9}$$

[2]As the RF signal power increases, the precision of the fifth-order polynomial gets worse. However, for the power levels considered in our work, the precision was always sufficient; this is borne out by the ultimate agreement (to be discussed later) between the Volterra analysis and SPECTRE simulations in Figs. 4 and 5.

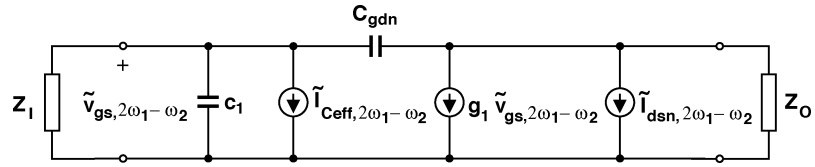


Fig. 8. Circuit for the Volterra calculation.

where, as before, the coefficients g_1 through g_3 change with *both* the bias point *and* the RF signal power, such that (9) always traces out the appropriate $i_{\rm dsn}$ versus $v_{\rm gs}$ curve.

With the power series in (7) and (9) established, the circuit for the Volterra calculation, based on the "method of nonlinear currents" [14, pp. 190–207], is shown in Fig. 8. Here, Z_I represents the impedance seen looking into the input matching network from the NMOS gate when $i_s = 0$, and Z_O represents the impedance seen looking into the output matching network from the NMOS drain. Since Z_I presents a short circuit at even-order frequencies (see Section II-C), the distortion currents generated by $i_{\rm dsn}$ and $C_{\rm eff}$ have the following phasor amplitudes:

$$\tilde{i}_{\rm dsn,2\omega_1-\omega_2} = \frac{3}{4} g_3 \tilde{v}^2_{\rm gs,\omega_1} \tilde{v}^*_{\rm gs,\omega_2} \tag{10}$$

and

$$\tilde{i}_{C_{\rm eff},2\omega_1-\omega_2} = j(2\omega_1-\omega_2)\left[\frac{1}{4}c_3\tilde{v}^2_{\rm gs,\omega_1}\tilde{v}^*_{\rm gs,\omega_2} + \frac{1}{8}c_5\left(2\tilde{v}^3_{\rm gs,\omega_1}\tilde{v}^*_{\rm gs,\omega_1}\tilde{v}^*_{\rm gs,\omega_2} + 3\tilde{v}^2_{\rm gs,\omega_1}\tilde{v}_{\rm gs,\omega_2}\tilde{v}^{*2}_{\rm gs,\omega_2}\right)\right] \tag{11}$$

where $\tilde{v}_{\rm gs,\omega_1}$ and $\tilde{v}_{\rm gs,\omega_2}$ are the phasor amplitudes of the gate-source voltage at the fundamental frequencies, and * denotes complex conjugation. The distortion voltages that result at the gate and drain can then be computed using the circuit of Fig. 8 as in (12) and (13) shown at the bottom of the page, where $Z_I' \equiv Z_I \| c_1$, and the impedances Z_I' and Z_O should be evaluated at the intermodulation frequency $2\omega_1 - \omega_2$. The drain voltage at the fundamental frequency is also easily found to be

$$\tilde{v}_{\rm ds,\omega_1} = \frac{-g_1 Z_O + jw_1 C_{\rm gdn} Z_O}{1 + jw_1 C_{\rm gdn} Z_O}\tilde{v}_{\rm gs,\omega_1} \tag{14}$$

where, in this case, Z_O should be evaluated at the fundamental frequency ω_1. The IM3 at the gate and drain are then simply

$$\mathrm{IM3}_G = 20\log\left|\frac{\tilde{v}_{\rm gs,2\omega_1-\omega_2}}{\tilde{v}_{\rm gs,\omega_1}}\right| \tag{15}$$

and

$$\mathrm{IM3}_D = 20\log\left|\frac{\tilde{v}_{\rm ds,2\omega_1-\omega_2}}{\tilde{v}_{\rm ds,\omega_1}}\right|. \tag{16}$$

Superimposed on the SPECTRE simulation results in Figs. 4 and 5 are values for the gate and drain IM3 found from (10)–(16), with $\tilde{v}_{\rm gs,\omega_1} \approx \tilde{v}_{\rm gs,\omega_2}$ obtained from the terminal gate-source voltage of the NMOS device in SPECTRE. As shown, the Volterra expressions are able to predict the main trends in IM3 as a function of both bias and power level. Of course, since the power-series coefficients in (7) and (9), and the values of $\tilde{v}_{\rm gs,\omega_1} \approx \tilde{v}_{\rm gs,\omega_2}$, were all found using information from SPECTRE, this agreement may not be too surprising. However, the real utility of the Volterra expressions lies in their ability to isolate the impact of the individual nonlinearities.

Fig. 9 shows the contributions to the drain IM3 arising from the $C_{\rm eff}$ and $i_{\rm dsn}$ nonlinearities, as computed from (13), (14), and (16). The contribution from $C_{\rm eff}$ is found by setting $\tilde{i}_{\rm dsn,2\omega_1-\omega_2} \equiv 0$ in the expressions, and the contribution from $i_{\rm dsn}$ is found by setting $\tilde{i}_{c_{\rm eff},2\omega_1-\omega_2} \equiv 0$. The $C_{\rm eff}$ contributions are shown for both the basic and linearized amplifiers; the $i_{\rm dsn}$ contributions do not change, so only one curve is shown. As illustrated, in the basic amplifier, the $C_{\rm eff}$ nonlinearity limits the drain IM3 over most power levels; only at very high power levels does the $i_{\rm dsn}$ nonlinearity become important, which is simply a result of increased clipping in class-AB mode. On the other hand, in the linearized amplifier, the impact of the $C_{\rm eff}$ nonlinearity is greatly reduced, and correspondingly, except at high power levels where the $i_{\rm dsn}$ nonlinearity dominates, the compensation scheme leads to the improved performance originally seen in Fig. 5. Similar analysis could be undertaken and comments made for the gate IM3 in Fig. 4. (Again, there is no improvement at very high power levels due to the $i_{\rm dsn}$ nonlinearity, which can impact the gate IM3 by way of feedback through $C_{\rm gdn}$.)

IV. Experimental Results

A. IC Implementation

Fig. 10 shows a simplified schematic of a fully matched two-stage CMOS class-AB power amplifier that was designed and implemented. A single-ended configuration, which avoids the use of baluns, was employed to make the amplifier more cost-effective and easier to integrate. Meanwhile, a two-stage topology was utilized to achieve a gain higher than 20 dB. In order to make the gain and stability less sensitive to parasitic

$$\tilde{v}_{\rm gs,2\omega_1-\omega_2} = -\frac{Z_I'\left\{\tilde{i}_{\rm dsn,2\omega_1-\omega_2}\left[j(2\omega_1-\omega_2)C_{\rm gdn}Z_O\right] + \tilde{i}_{C_{\rm eff},2\omega_1-\omega_2}\left[1+j(2\omega_1-\omega_2)C_{\rm gdn}Z_O\right]\right\}}{1+j(2\omega_1-\omega_2)C_{\rm gdn}\left(Z_I'+Z_O+g_1Z_I'Z_O\right)} \tag{12}$$

$$\tilde{v}_{\rm ds,2\omega_1-\omega_2} = -\frac{Z_O\left\{\tilde{i}_{\rm dsn,2\omega_1-\omega_2}\left[1+j(2\omega_1-\omega_2)C_{\rm gdn}Z_I'\right] - \tilde{i}_{C_{\rm eff},2\omega_1-\omega_2}\left[g_1-j(2\omega_1-\omega_2)C_{\rm gdn}Z_I'\right]\right\}}{1+j(2\omega_1-\omega_2)C_{\rm gdn}\left(Z_I'+Z_O+g_1Z_I'Z_O\right)} \tag{13}$$

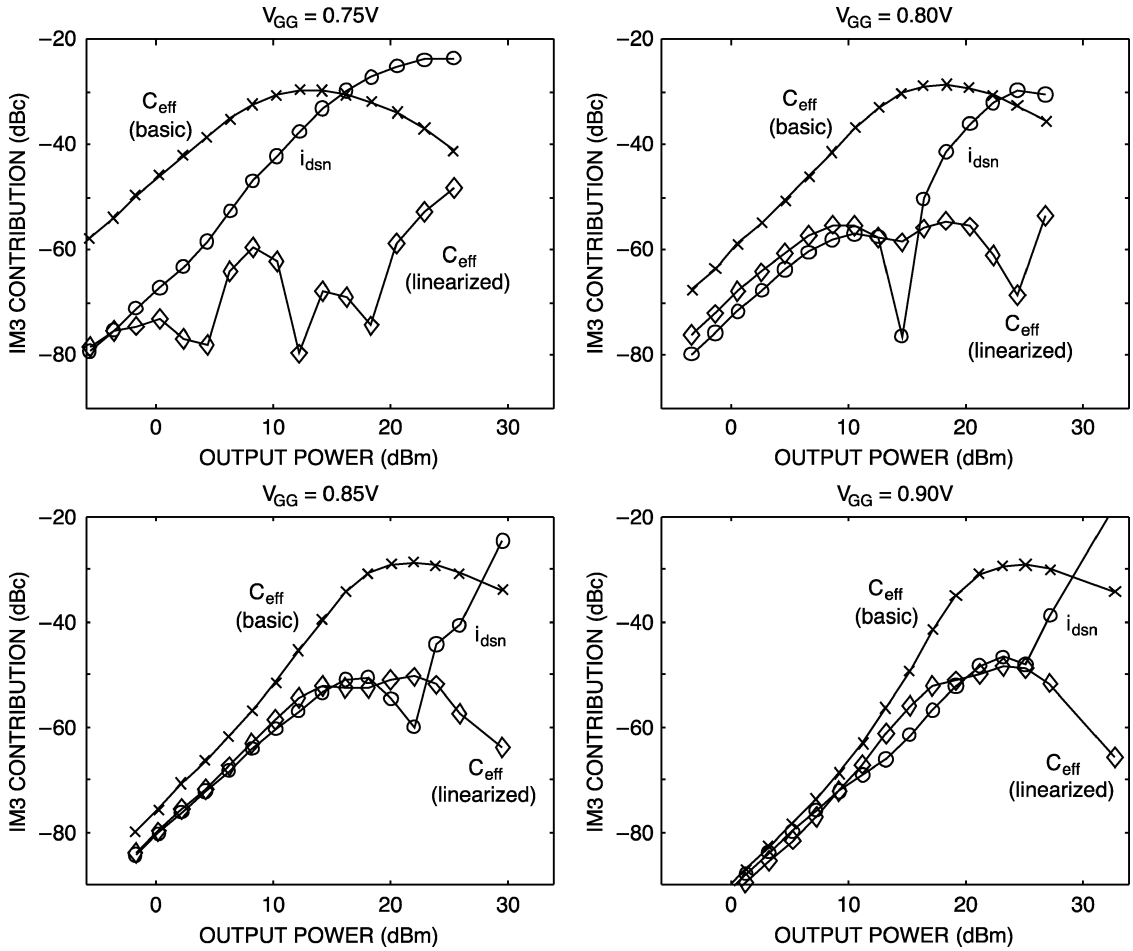


Fig. 9. Calculated contributions to the drain IM3 from the $C_{\rm eff}$ and $i_{\rm dsn}$ nonlinearities for both the basic and linearized amplifiers in Fig. 3(a) and (b), respectively. The values are computed from the Volterra expressions (10)–(16), as described in the text. In each case, $V_{\rm DD} = 3.3$ V.

bondwire inductance, our analysis, which is in excellent agreement with full-chip SPECTRE simulations, revealed that all ground connections should be made through a single node s_0, as shown in Fig. 10. Further details on PA stability and design strategies, such as the choice of device widths and the design of matching networks, can be found in [16, ch. III].

For comparison purposes, three PAs were fabricated: PA1 is the uncompensated and fully integrated version, which means that all the matching (input, interstage, and output) is on-chip; PA2 is also fully integrated but with the compensation circuitry applied; PA3 is the same as PA2 except that its output matching was off-chip.

The circuits were fabricated in a 0.5-μm four-metal-layer IBM Silicon Germanium BiCMOS process (SiGe5AM), in which only the CMOS devices were used. The fully integrated and compensated chip (PA2) occupies an area of 2.0×1.6 mm^2 including bonding pads. The dies were assembled using Amkor MicroLeadFrame (MLF) packages and tested on standard two-layer RO4350 20-mil printed circuit boards (PCBs). Figs. 11 and 12 show the die microphotograph of PA2 and the prototype PCB of PA3, respectively.

A note should be made regarding the impact of interstage matching in two-stage CMOS PAs on the intended frequency of operation. The interstage matching of two-stage CMOS PAs is generally difficult because of the large gate capacitance exhibited by the active device of the output stage. In our case, the total gate capacitance of the output stage of PA2 is approximately 22 pF including the layout parasitics. This results in a value of only 0.3 nH for the interstage matching inductor L_1, while the parasitic inductance of the matching network itself is roughly 0.1 nH. As a result, it is difficult to tune the interstage matching network to a precise frequency of operation, which can impact the gain and efficiency. In our case, we found that the two-stage PAs exhibited higher gains and better efficiency at frequencies slightly below the design value of 1.95 GHz. As a result, to acquire the required gain and efficiency performance, measurements were carried out at 1.75 GHz instead of 1.95 GHz. Additional off-chip input and output matching circuitry, which is not shown in Fig. 10, was necessary to modify the input and output matching to 1.75 GHz. However, this slight modification does not impact our conclusions or the generality of our results.

Each of the amplifiers was operated at a $V_{\rm DD}$ of 3.3 V and drew a total quiescent current of 97 mA (46 mA for the driver stage and 51 mA for the output stage) when the output stage was biased at $V_{\rm GG} = 0.8$ V.

B. Measurement Results

1) Gain and Efficiency: Fig. 13 shows the measured gain and power-added efficiency (PAE) of the three PAs. As can be seen, the uncompensated and fully integrated PA (PA1) achieves a small-signal gain of 24.3 dB and a peak PAE of 23% at the designed output power of 24 dBm; it is worth noting that these are

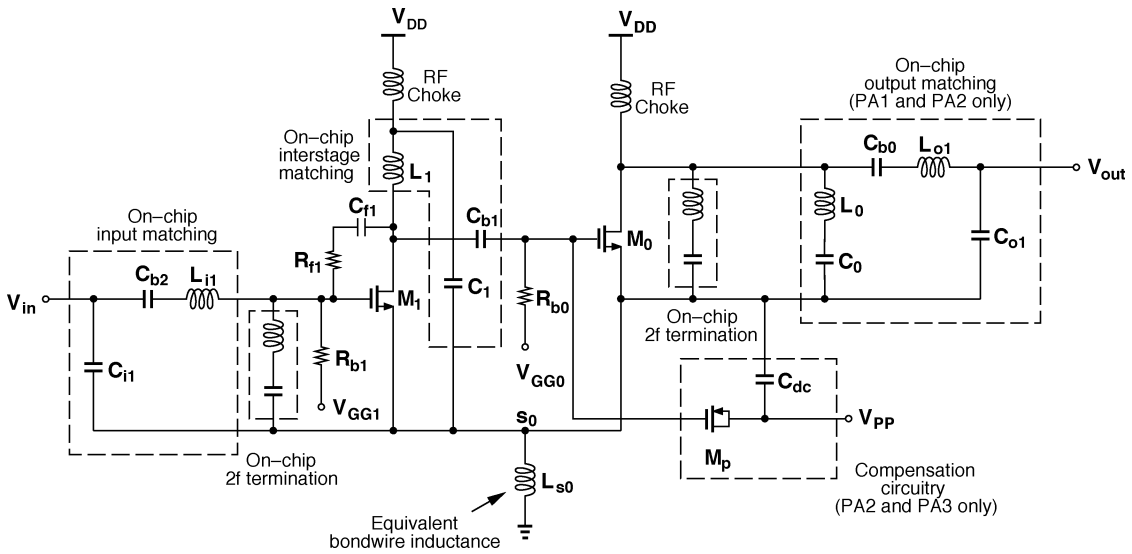


Fig. 10. Simplified schematic of the fully matched two-stage CMOS class-AB power amplifier. L_{s0} represents the equivalent inductance of multiple bondwires in parallel from s_0 to ground.

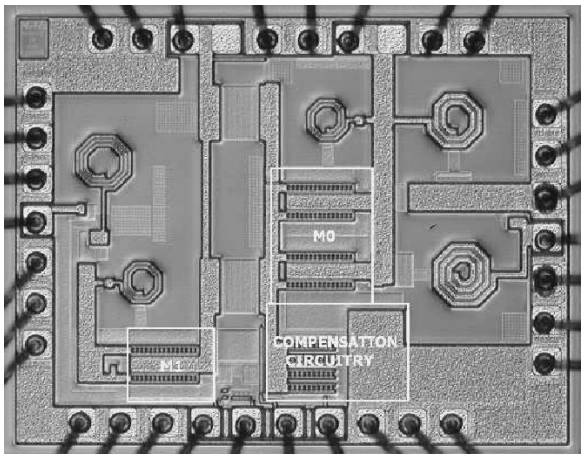


Fig. 11. Die microphotograph of the fully integrated and compensated two-stage CMOS PA (PA2).

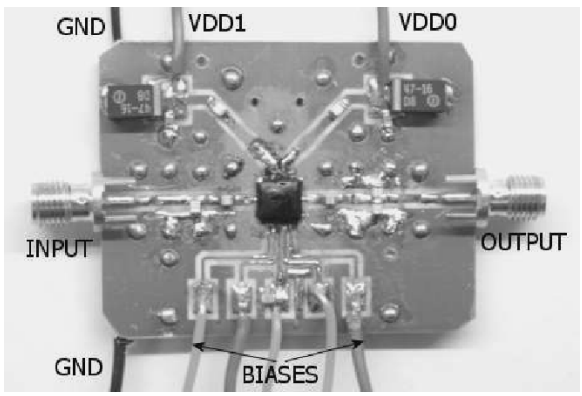


Fig. 12. Printed circuit board implementation of PA3.

close to the values of 25 dB and 25%, respectively, predicted by full-chip simulations during the design phase. PA2 achieves similar PAE performance but with a gain that is 3 dB lower than PA1; this reduced gain is attributed to the increased input

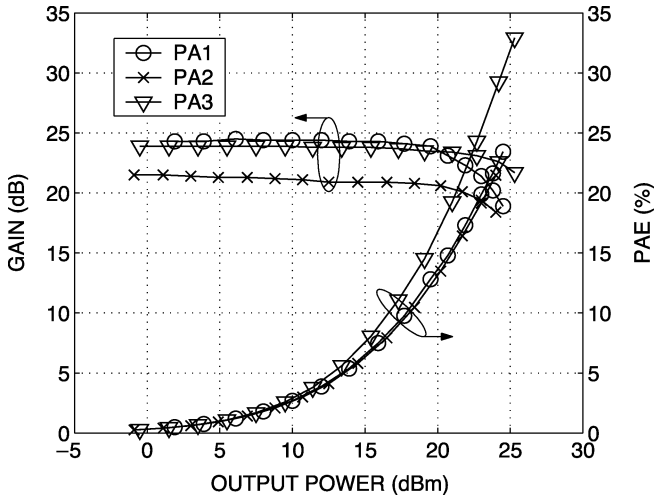


Fig. 13. Measured gain and power-added efficiency versus output power of the three PAs. The input is a real-time 3GPP WCDMA signal generated by an Agilent E4438C vector signal generator. The output stages of the PAs are all biased at $V_{\rm GG} = 0.8$ V, $V_{\rm DD} = 3.3$ V.

capacitance associated with the compensation scheme, as discussed previously. PA3 has better gain and efficiency than PA2 because of its low-loss, off-chip output matching; it achieves a small-signal gain of 23.9 dB and a PAE of 29% at an output power of 24 dBm, and the peak efficiency is 33% at an output power of 25 dBm. Curves of gain and output power versus input power were also constructed to obtain the 1-dB compression point. The output power and PAE at the 1-dB compression point for PA1, PA2, and PA3 are 20.5 dBm and 15%, 20.2 dBm and 13.5%, and 24 dBm and 29%, respectively.

2) Linearity: To verify their linearity performances, the PAs were tested under various bias and power levels using both two-tone and real-time 3GPP WCDMA signals generated by an Agilent E4438C ESG vector signal generator. Figs. 14, 15, and 16 show the measured third-order intermodulation, adjacent-channel leakage power (ACP1), and alternate-channel power (ACP2) for the three PAs, respectively. Again, the output stages of all the PAs were biased at 0.8 V. The measurements show

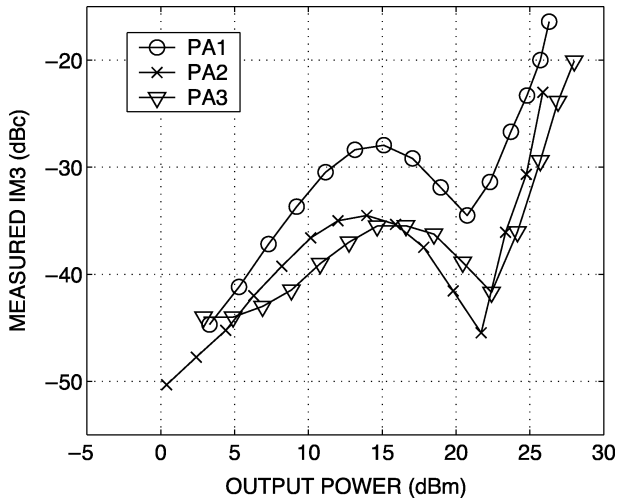


Fig. 14. Measured IM3 versus peak-envelope output power of the three PAs. The output stages of the PAs are all biased at $V_{\rm GG} = 0.8$ V, $V_{\rm DD} = 3.3$ V.

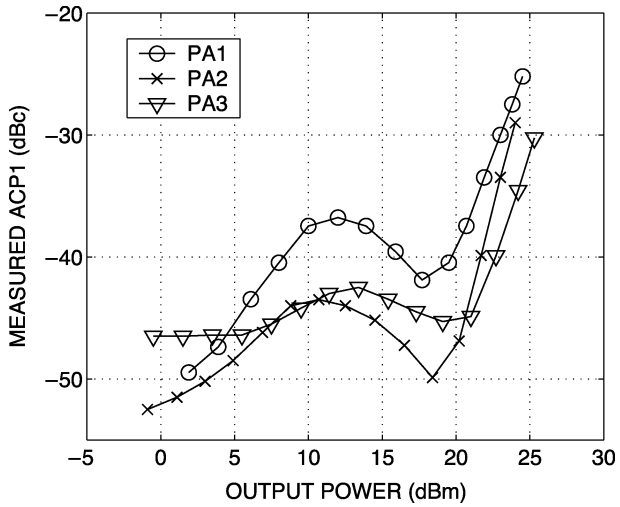


Fig. 15. Measured adjacent-channel leakage power versus carrier output power of the three PAs. The output stages of the PAs are all biased at $V_{\rm GG} = 0.8$ V, $V_{\rm DD} = 3.3$ V.

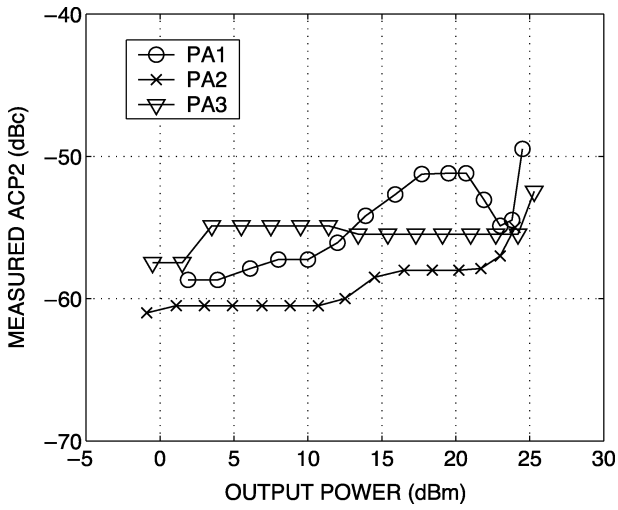


Fig. 16. Measured alternate-channel power versus carrier output power of the three PAs. The output stages of the PAs are all biased at $V_{\rm GG} = 0.8$ V, $V_{\rm DD} = 3.3$ V.

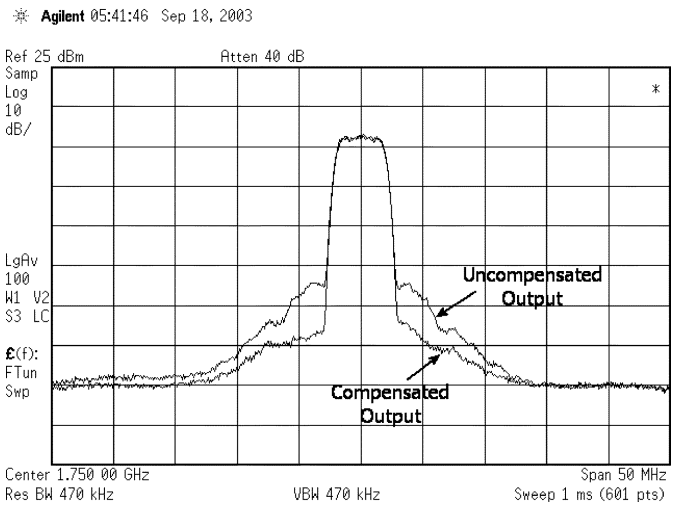


Fig. 17. Measured WCDMA spectra of PA1 and PA2 at a carrier output power of nearly 20 dBm. The output stages of the PAs are both biased at $V_{\rm GG} = 0.8$ V, $V_{\rm DD} = 3.3$ V.

that the compensated PAs (PA2 and PA3) have much better linearity than the uncompensated PA (PA1) for various gate biases and a wide range of output power; in addition, the IM3 measurements show similar trends as those shown in Fig. 5 of Section II-C. As can be seen, PA3 achieves an ACP1 of -35 dBc and ACP2 of -55 dBc at a carrier output power of 24 dBm, which is compliant with the 3GPP WCDMA ACP requirements of -33 dBc and -43 dBc [17], respectively. Due to the loss of on-chip output matching, PA1 and PA2 can only meet the WCDMA ACP requirements at output powers of 22 and 23 dBm, respectively. Fig. 17 shows the measured WCDMA spectra of PA1 and PA2 at a carrier output power of nearly 20 dBm.

We should point out here that while we used the published 3GPP WCDMA specifications from [17] as a guideline for our design, these specifications are actually for the entire cellphone, and the requirements for the PA itself will be more stringent. In a design for any commercial mass-produced product, one should account for isolator, duplexor, switchplexor, and filter losses in arriving at the PA requirements, and also test functionality in worst-case process and temperature corners.

It is also worth mentioning that all the bias voltages utilized in our measurements are almost exactly the designed values; in addition, no oscillation was observed during the entire measurement procedure, even when both the source and load were disconnected.

Table I compares the performance of recently reported linear power amplifiers for handset applications. As can be seen, although a CMOS PA's peak efficiency is generally lower than its GaAs HBT (FET) counterpart, if properly linearized, it can effectively be used as a low-cost alternative, especially for low-supply voltage and medium-power applications.

V. Conclusion

The following conclusions can be drawn from this study of CMOS class-AB power amplifiers.

TABLE I
PERFORMANCE COMPARISON OF RECENTLY REPORTED LINEAR POWER AMPLIFIERS FOR HANDSET APPLICATIONS

Ref.	Technology	Pout (dBm)	PAE	Gain (dB)	[Signal] ACPR @ Pout	VDD (V)	Freq. (MHz)	Operating class
Su 98 [18]	CMOS 0.8 μm	28	33 %	N/A	[NADC] -30 dBc @ 28 dBm	3	836	AB (linearized)
Giry 00 [19]	CMOS 0.35 μm	23.5	35 %	24.6	[PDC] -55 dBc @ 21.5 dBm	2.5	1910	AB
Yen 03 [20]	CMOS 0.25 μm	20	28 %	11.2	[$\pi/4$ DQPSK] -28 dBc @ 18 dBm	2.5	2450	AB (linearized)
This work (PA3)	CMOS 0.5 μm	24	29 %	23.9	[WCDMA] -35 dBc @ 24 dBm	3.3	1750	AB (linearized)
Vintola 01 [21]	AlGaAs/GaAs HBT	>24	>27 %	>30	[WCDMA] -36 dBc @ 26 dBm	3.5	1950	AB
Jager 02 [22]	InGaP/GaAs HBT	27	38 %	22.6	[WCDMA] -37 dBc @ 27 dBm	N/A	1950	AB
Srirattana 03 [23]	GaAs FET	29.7	46 %	8.5	[WCDMA] -38 dBc @ 28.6 dBm	N/A	1950	Doherty 3-stage

1) The nonlinear gate-source capacitance is a dominant source of distortion that may limit the linearity of CMOS class-AB power amplifiers.
2) Improved performance can be obtained by using a compensating nonlinearity, provided by the gate-source capacitance of an appropriately biased and sized PMOS device placed alongside the NMOS device that provides the class-AB amplification.
3) Simulations and experiments show that the method can improve both the two-tone IM3 and adjacent-channel leakage power by approximately 8 dB over a wide range of output power.
4) The linearized two-stage amplifier is capable of delivering an output power of 24 dBm with a small-signal gain of nearly 24 dB and an overall power-added efficiency of 29%, demonstrating the potential utility of the design approach for 3GPP WCDMA applications.

ACKNOWLEDGMENT

The authors wish to thank Prof. P. M. Asbeck, Dr. L. Sheng, Mr. D. Kimball, and Mr. J. Deng of UCSD for invaluable discussions.

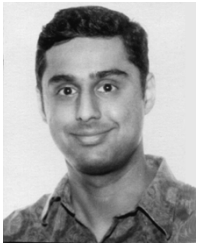
REFERENCES

[1] D. Su and W. McFarland, "A 2.5 V, 1-W monolithic CMOS RF power amplifier," in *Proc. IEEE Custom Integrated Circuits Conf.*, May 1997, pp. 189–190.
[2] K.-C. Tsai and P. Gray, "A 1.9-GHz, 1-W CMOS Class-E power amplifier for wireless communications," *IEEE J. Solid-State Circuits*, vol. 34, pp. 962–970, July 1999.
[3] I. Aoki, S. Kee, D. Rutledge, and A. Hajimiri, "Fully integrated CMOS power amplifier design using the distributed active-transformer architecture," *IEEE J. Solid-State Circuits*, vol. 37, pp. 371–383, Mar. 2002.
[4] K. Mertens and M. Steyaert, "A 700-MHz 1-W fully differential CMOS class-E power amplifier," *IEEE J. Solid-State Circuits*, vol. 37, pp. 137–141, Feb. 2002.
[5] S. C. Cripps, *RF Power Amplifiers for Wireless Communications*. Boston, MA: Artech House, 1999.
[6] C. Wang, L. E. Larson, and P. M. Asbeck, "A nonlinear capacitance cancellation technique and its application to a CMOS class AB power amplifier," in *Proc. IEEE Radio Frequency Integrated Circuits Symp.*, 2001, pp. 39–42.
[7] S. H.-M. Jen, C. C. Enz, D. R. Pehlke, M. Schröter, and B. J. Sheu, "Accurate modeling and parameter extraction for MOS transistors valid up to 10 GHz," *IEEE Trans. Electron Devices*, vol. 46, pp. 2217–2227, Nov. 1999.
[8] C. C. Enz, "MOS transistor modeling for RF integrated circuit design," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2000, pp. 189–196.
[9] C. C. Enz and Y. Cheng, "MOS transistor modeling for RF IC design," *IEEE Trans. Electron Devices*, vol. 35, pp. 201–231, Feb. 2000.
[10] Y. Cheng, C.-H. Chen, C. Enz, M. Matloubian, and M. J. Deen, "MOS modeling for RF circuit design," in *Proc. 3rd IEEE Int. Caracas Conf. Devices, Circuits and Systems*, 2000, pp. D23/1–D23/8.
[11] Y. Cheng, C.-H. Chen, M. Matloubian, and M. J. Deen, "High-frequency small signal AC and noise modeling of MOSFETs for RF IC design," *IEEE Trans. Electron Devices*, vol. 49, pp. 400–408, Mar. 2002.
[12] Y. Tsividis, *Operation and Modeling of the MOS Transistor*, 2nd ed. Boston, MA: McGraw-Hill, 1999.
[13] *SiGe5AM Model Reference Guide*, IBM Corp., Sept. 2002.
[14] S. A. Maas, *Nonlinear Microwave Circuits*. Piscataway, NJ: IEEE Press, 1997.
[15] W. Liu, *MOSFET Models for SPICE Simulation Including BSIM3v3 and BSIM4*. New York: Wiley, 2001.
[16] C. Wang, "CMOS power amplifiers for wireless communications," Ph.D. thesis, Univ. California, San Diego, Dec. 2003.
[17] *Technical Specification 3GPP TS 25.101 V6.1.0*, June 2003.
[18] D. Su and W. McFarland, "An IC for linearizing RF power amplifiers using envelope elimination and restoration," *IEEE J. Solid-State Circuits*, vol. 33, pp. 2252–2258, Dec. 1998.
[19] A. Giry, J.-M. Fourier, and M. Pons, "A 1.9 GHz low voltage CMOS power amplifier for medium power RF applications," in *Proc. IEEE RFIC Symp.*, 2000, pp. 121–124.
[20] C.-C. Yen and H.-R. Chuang, "A 0.25-μm 20-dBm 2.4-GHz CMOS power amplifier with an integrated diode linearizer," *IEEE Microwave Wireless Compon. Lett.*, vol. 13, pp. 45–47, Feb. 2003.
[21] V. Vintola, M. Matilainen, S. Kalajo, and E. Jarvinen, "Variable-gain power amplifier for mobile WCDMA applications," *IEEE Trans. Microwave Theory Tech.*, vol. 49, pp. 2464–2471, Dec. 2001.
[22] H. Jager, A. Grebennikov, E. Heaney, and R. Weigel, "Broadband high-efficiency monolithic InGaP/GaAs HBT power amplifiers for 3G handset applications," in *IEEE MTT-S Dig.*, 2002, pp. 1035–1038.

[23] N. Srirattana, A. Raghavan, D. Heo, P. Allen, and J. Laskar, "A high-efficiency multistage Doherty power amplifier for WCDMA," in *Proc. IEEE MTT Symp.*, 2003, pp. 397–400.

Chengzhou Wang (S'01–M'03) received the B.S.E.E. degree from Beijing University, China, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at San Diego in 1999 and 2003, respectively.

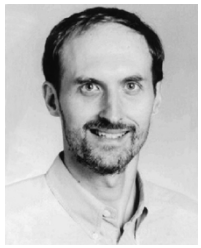
He is currently a Senior Design Engineer with Marvell Semiconductor, Sunnyvale, CA, focusing on RF design blocks for wireless applications. His research interests include linearization techniques for CMOS class-AB power amplifiers in wireless communications.



Mani Vaidyanathan (S'94–M'99) received the B.A.Sc. degree in computer engineering (including co-op work terms with firms such as Nortel Networks, IBM, and Genesis Microchip) and the M.A.Sc. degree in electrical engineering from the University of Waterloo (UW), Waterloo, ON, Canada, in 1990 and 1992, respectively. From 1993 to 1994, he was an Adjunct Lecturer and Research Assistant at UW, teaching courses and performing research in the area of semiconductor devices. He began his Ph.D. work at the University of British Columbia (UBC), Vancouver, BC, Canada, in 1994, and completed his dissertation in 1998.

In 1999, he joined the Department of Electrical and Computer Engineering at the University of California at San Diego (UCSD), La Jolla, CA, as a Postdoctoral Fellow, and was promoted to Assistant Research Scientist in Fall 2001. During 2002–2003, he was on leave from UCSD and held the position of Visiting Assistant Professor at Purdue University, West Lafayette, IN, and in Fall 2004, he joined the faculty of the University of Alberta, Edmonton, AB, Canada, where he is currently an Assistant Professor. His research interests are in the theory and modeling of semiconductor devices, where he has worked on topics ranging from studying carrier transport in small-dimension devices to modeling high-frequency distortion for wireless applications.

Dr. Vaidyanathan received postgraduate scholarships from the Natural Sciences and Engineering Research Council (NSERC) of Canada for the M.A.Sc. and Ph.D. degrees, as well as an NSERC postdoctoral fellowship. He was a Killam Scholar and Governor-General's Gold Medal nominee at UBC, and received the Sandford Fleming Award for teaching excellence at UW.



Lawrence E. Larson (M'86–SM'90–F'00) received the BS degree in electrical engineering in 1979 and the M. Eng. degree in 1980, both from Cornell University, Ithaca, NY. He received the Ph.D. degree in electrical engineering from the University of California at Los Angeles in 1986.

From 1980 to 1996, he was at Hughes Research Laboratories, Malibu, CA, where he directed the development of high-frequency microelectronics in GaAs, InP, and Si/SiGe and MEMS technologies. He joined the faculty at the University of California at San Diego in 1996, where he is the inaugural holder of the Communications Industry Chair. He is currently Director of the UCSD Center for Wireless Communications. During the 2000–2001 academic year, he was on leave at IBM Research, San Diego, CA, where he directed the development of RFICs for 3G applications. He has published over 200 papers, coauthored three books, and holds 27 U.S. patents.

Dr. Larson was the recipient of the 1995 Hughes Electronices sector Patent Award for his work on RF MEMs, a co-recipient of the 1996 Lawrence A. Hyland Patent Award of Hughes Electronics for his work on low-noise millimeterwave HEMTs, and the 1999 IBM Microelectronics Excellence Award for his work in Si/SiGe HBT technology.