16

Power Amplifier Design for High Spectrum-Efficiency Wireless Communications

Steve Hung-Lung Tu, Ph.D.
Fu Jen Catholic University
Taiwan

1. Introduction

The growing market of wireless communications has generated increasing interest in technologies that will enable higher data rates and capacity than initially deployed systems. The IEEE 802.11a standard for wireless LAN (WLAN), which is based on orthogonal frequency division multiplexing (OFDM) modulation, provides nearly five times the data rate and as much as ten times the overall system capacity as currently available 802.11b wireless LAN systems (Eberle et al., 2001; Zargari et al., 2002; Thomson et al., 2002). The modulation format of the IEEE 802.11a is OFDM (Orthogonal Frequency Division Multiplexing) which is not a constant-envelope modulation scheme; more sensitive to frequency offset and phase noise, and has a relatively large peak-to-average power ratio. These reasons induce the linearity requirements, which are crucial for power amplifier design. Among various linearization techniques, transistor-level predistortion is the simplest approach to implement and can be realized in a small area, which makes it be the most compatible with RFIC implementation.

Conventionally, WLAN has been implemented with multi-chip approach or single chip with processes other than CMOS. For example, the radio frequency (RF) and intermediate frequency (IF) sections are fabricated with GaAs or BiCMOS processes, and the baseband DSP section with a CMOS process. Note that the complexity and cost will be dramatic in a wireless LAN system with hybrid processes, which makes the CMOS process be the most promising approach to achieve high integration level, low power consumption, and low cost for the integration of baseband and RF front-end circuits of a WLAN system.

There are two different modulation schemes employed in wireless communication standards: linear modulation and nonlinear modulation, in which the former one is employed in North American Digital Cellular (NADC) standard whereas the latter one is also called constant-envelope modulation which employed in the European Standard for Mobile Communications (GSM). The main requirements for power amplifiers (PA's) employed in wireless communications are generally high power efficiency and low supply voltage operating at high frequencies. Class-E PA's have demonstrated the potential of high power efficiency whereas due to the operation characteristics, it can only be adopted in constant-envelope modulation applications. The linear modulation scheme, on the other hand can achieve high spectrum efficiency, which is especially suitable for the application of

wireless communications. A power amplifier that can achieve high power efficiency while providing high spectrum efficiency is therefore highly desired. To discuss this issue, in this chapter a class-AB type amplifier in a standard CMOS process is investigated together with the presentation of a transistor-level predistortion compensation techniques.
The main theme of this chapter is aimed at providing the fundamental background knowledge concerned with linear PA design for high spectrum-efficiency wireless communications. Nevertheless, we also present the design considerations of the state-of-the-art linear PA's together with the design techniques operating at the gigahertz bands in CMOS technologies. To conclude the chapter, we investigate a design and implementation of a Class-AB PA operating at GHz for IEEE 802.11 wireless LAN to demonstrate the feasibility.

2. Design Concept of CMOS Power Amplifiers

Conjugate matching is fully understood as making the value of load resistance equals to the real part of the generator's impedance. Since the maximum power will be delivered to the load, however, this delivering power will be limited by the maximum rating of the transistor. This phenomenon can be shown in Fig. 1. For utilizing the maximum current and voltage swing of the transistor, a lower than the real part of generator's impedance is chosen for maximum power transformation.

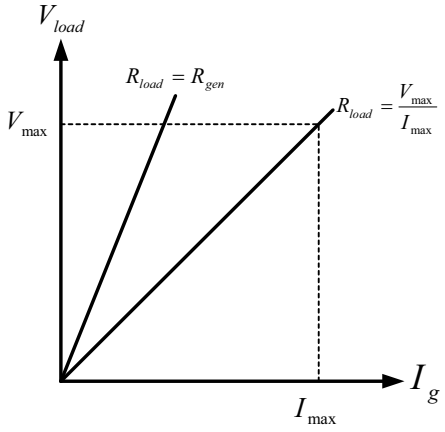

Fig. 1. Conjugate Matching and Power Matching

The load-line match represents an actually compromise that extracts the maximum power from the power devices, and simultaneously maintains the output swing within the limitation of the power devices and the available DC supply. In a typical situation, the conjugate matching yields a 1-dB compression power about 2dB lower than that can be attained by the correct load line matching (power matching), which means the power device can deliver 2dB lower power than the manufactures specify. So the power matching condition has to be taken seriously, despite the fact that the gain of the PA circuits is lower than conjugate matching at lower signal levels.

Another design concept of CMOS power amplifiers is the knee voltage effect of deep sub-micron CMOS transistors. The knee voltage (pinch-off voltage) divides the saturation and linear operation region of the transistor. Typically, for a power transistor may be 10% or 15% of the supply voltage, and the optimum load impedance is

$$R_{opt} = \frac{V_{\max} - V_{Knee}}{I_{\max}} \tag{1}$$

Notice that the knee voltage can be as high as 50% of the supply voltage for deep sub-micron CMOS technologies as shown in Fig. 2. Therefore, preventing the CMOS transistor from operating in the linear region doesn't result in the optimum output power. Also, both the saturation and linear operating regions must be considered in determining the optimum output load impedance since a lot portion of RF cycle could be in the linear operating region.

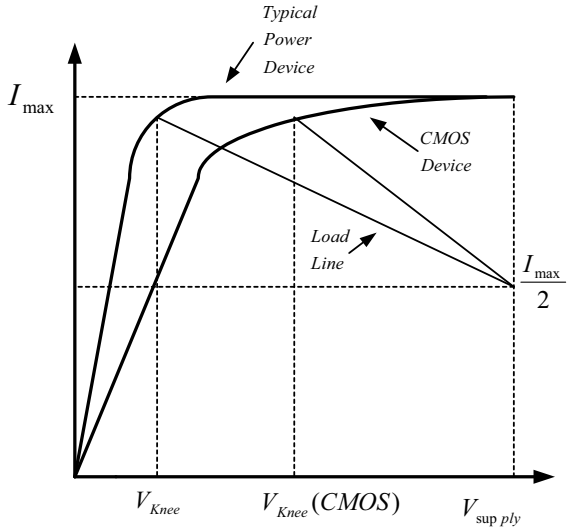

Fig. 2. Knee Voltage of Typical Power Device and CMOS device.

Another issue is the choice of device size of each amplifying stage. A simple class-A amplifier can briefly explain this issue as shown in Fig.3, in which RFC means radio frequency choke with large impedance compared with load impedance R_L.

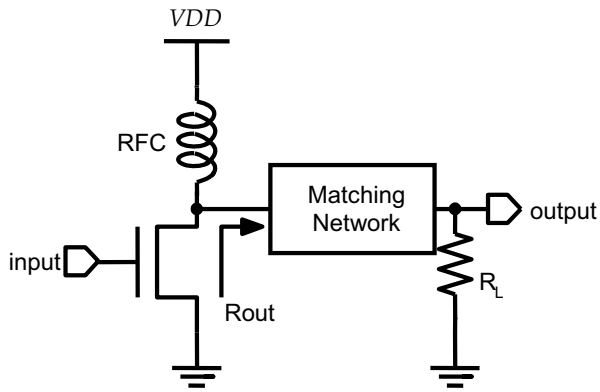

Fig. 3. Simple circuit of class-A amplifier.

Load impedance R_L is generally equal to 50 ohms and the matching network is tuned to obtain the device maximum output power. When the device output power is reaching to the maximum, the output impedance *Rout* is defined as the optimum load impedance *Ropt*. In a class-A amplifier, the device plays a role of a voltage-dependant current source as shown in Fig.4, in which I_{max} is the maximum available current of the device, I_{min} is the minimum current of the device. V_{max} is the maximum tolerance voltage of the device between drain and source of the device. V_{min} is the knee voltage of the device. V_{dc} and I_{dc} are the DC bias of the device. Therefore, the device voltage swing and current swing are V_{max}-V_{min} and I_{max}–I_{min}, respectively.

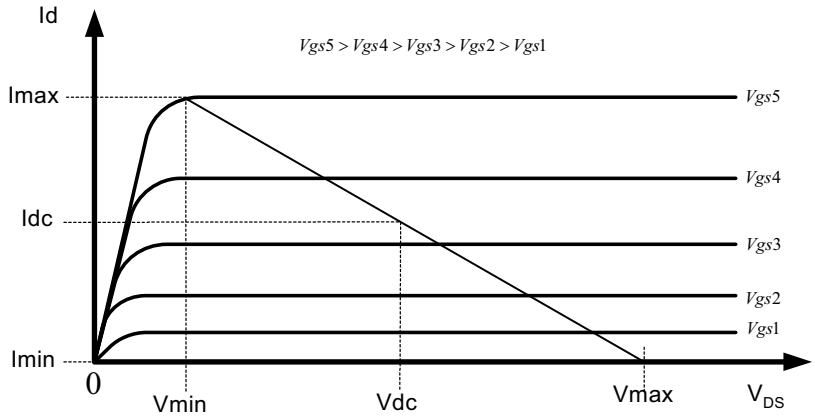

Fig. 4. I-V Curve of a NMOS device.

The optimum load impedance *Ropt* for maximum AC swing can thus be described as

$$R_{opt} = \frac{V_{\max} - V_{\min}}{I_{\max} - I_{\min}} \tag{2}$$

and the output power of the device can be expressed as

$$P_{RF} = V_{rms} I_{rms} = \frac{V_{\max} - V_{\min}}{2\sqrt{2}} \cdot \frac{I_{\max} - I_{\min}}{2\sqrt{2}} = \frac{(V_{\max} - V_{\min})(I_{\max} - I_{\min})}{8} \tag{3}$$

Since DC power consumption is

$$P_{DC} = \frac{V_{\max} + V_{\min}}{2} \cdot \frac{I_{\max} + I_{\min}}{2} = \frac{(V_{\max} + V_{\min})(I_{\max} + I_{\min})}{4} \tag{4}$$

the power efficiency η is thus given by

$$\eta = \frac{P_{RF}}{P_{DC}} = 0.5 \frac{(V_{\max} - V_{\min})(I_{\max} - I_{\min})}{(V_{\max} + V_{\min})(I_{\max} + I_{\min})} \tag{5}$$

Theoretically, V_{min} and I_{min} are zeros, and ideal drain efficiency of a class-A amplifier is 50%. However, in fact V_{min} and I_{min} are not equal to zeros, which implies that the drain efficiency should be less than 50%.

3. Power Amplifier Linearization Techniques

Feedback linearization techniques are the most general approaches employed in RF power amplifier design such as in the North American Digital Cellular (NADC) standard, a CMOS power feedback linearization is employed to linearize an efficient power amplifier transmitting a DQPSK modulated signal (Shi & Sundstrom, 1999), in which a reduction of more than 10dB in the adjacent channel interference was achieved according to the experimental results.

Fig. 5 shows a PMOS cancellation transistor-level linearization technique (Wang et al., 2001). The measurement results demonstrate that the amplifier with nonlinear input capacitance compensation has at least 6-dB IM3 (Third-order intermodulation intercept point) improvement in a wide range of output powers compared with the non-compensated amplifier whereas the disadvantages are low power gain and increasing input capacitance.

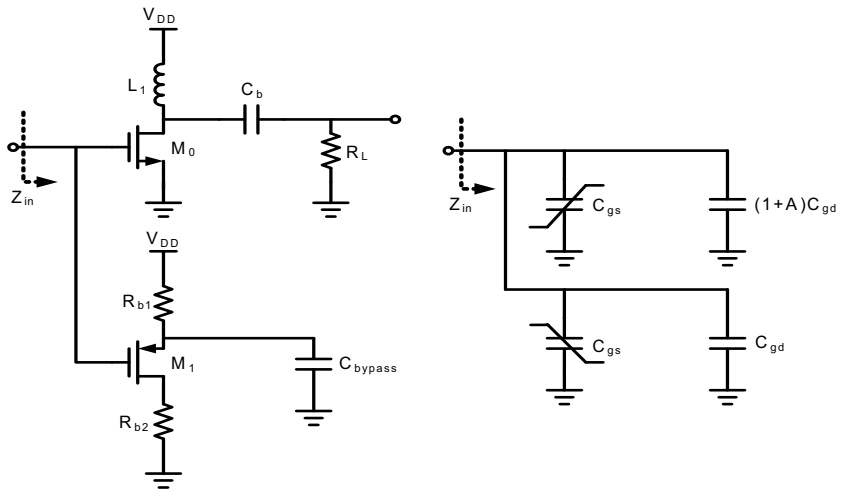

Fig. 5. Transistor-level linearization techniques – PMOS cancellation.

A miniaturized linearizer using a parallel diode with a bias feed resistance in an S-band power amplifier was also proposed (Yamauchi st al., 1997). The diode linearizer can improve adjacent channel leakage power of 5dB and power-added efficiency of 8.5%. Note that the improvement is based on 32 kb/ps, $\pi/4$ shift QPSK modulated signal at 28.6 KHz offset with a bandwidth of 16 KHz.

A miniaturized "active" predistorter using cascode FET structures was also applied to linearize a 2-GHz CDMA handset power amplifier. The ACPR (Adjacent Channel Power Ratio) improvement of 5dB was achieved (Jeon et al., 2002). Unlike the previously reported predistorters, this "active" predistorter can provide 7 to 17-dB gain which alleviates the requirement of additional buffer amplifiers to compensate the loss of the predistorter.

Another transistor-level linearization technique using varactor cancellation is shown in Fig.6 (Yu et al., 2000), which the approach improves 10-dB spectral regrowth with a low loss at 2GHz. However, the GaAs FET amplifier has AM-PM distortion under large-signal operating conditions due to the non-linear gate-to-source capacitance C_{gs} and the disadvantages are high cost, low integration with other transmitter circuits, and occupy a large PCB footprint.

A complex-valued predistortioner chip in CMOS for baseband or IF linearization of RF power amplifiers has been implemented (Westesson & Sundstrom, 1999). By choosing the coefficients for the predistortion polynomial properly, the lower-order distortion components can be cancelled out. Results of measurement performed as two-tone tests at an IF of 200MHz with 1MHz tone separation, using the chip for linearization gives a reduction of IM3 and IM5 with more than 30 and 10dB, respectively

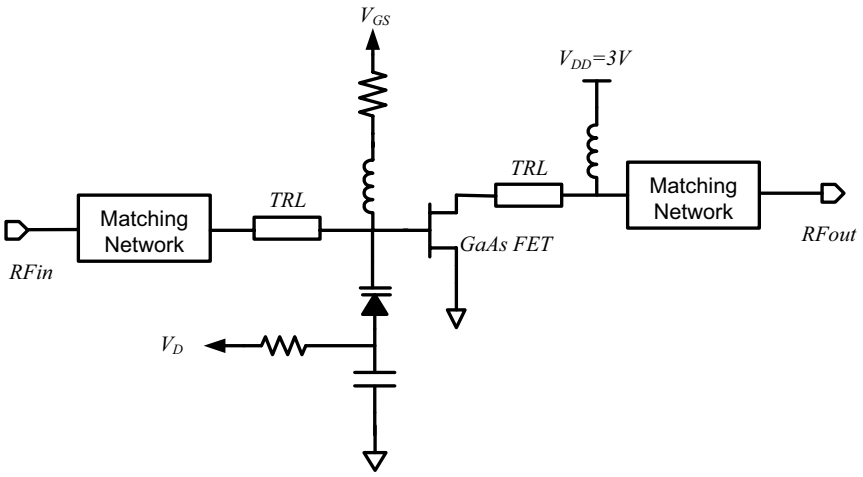

Fig. 6. Transistor-level linearization techniques – varactor cancellation.

Digital predistortion is a technique that counteracts both adjacent channel interference and BER degradation of power amplifiers. By employing digital feedback and a complex gain predistortion present, the experimental results demonstrate that a reduction in out of band spectra in excess of 20dB can be achieved (Wright & Durtler, 1992).

4. Predistortion Techniques for Linearization

Predistortion techniques are popular approaches for linearity improvement in power amplifier design. The concept is placing a black box on the PA input, which consumes little power and provides an acceptable linearity improvement instead of employing more complex circuitry to enhance system linearity. Basically, all predistortion approaches are open loop and can only achieve the level of linearization of closed-loop systems for limited periods of time and dynamic range. Recent research focuses on predistortion techniques offered by DSP. The basic concept is shown in Fig.7, where a predistorter preceding the nonlinear RF power amplifier implements a complementary nonlinearity, such that the combination of the two nonlinearities results in a linearized output signal. In practice, the lower orders nonlinear terms, such as third and fifth, is the most troublesome in communication applications. Even in practical PA models that consist of a couple of lower order nonlinear polynomial terms cannot be accurately estimated.

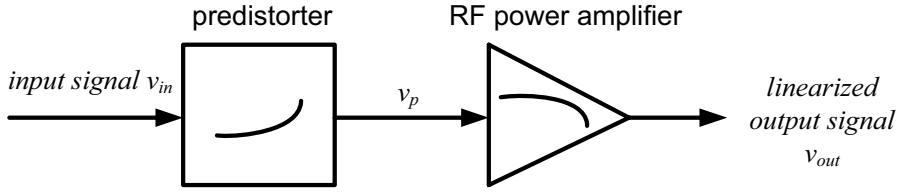

Fig. 7. Concept diagram of predistortion linearization.

4.1 Analog predistorters

Analog predistorters can be classified into two categories: 'simple' predistorters and 'compound' predistorters. The simple predistorters comprise one or more diodes, and the compound predistorters synthesize the required nonlinear characteristic using several sections to compensate different degree of distortion.

Simple analog predistorters mainly use a nonlinear resistive element such as a diode or an FET device as an RF voltage-control resistor that can be configured to provide higher attenuation at low drive levels and lower attenuation at high drive levels. A simple predistorter linearized RF power amplifier has been developed for 1.95-GHz wide-band CDMA (Hau et al., 1999), in which the amplifier is based on a heterojunction FET and its linearity and efficiency are improved by the employment of a MMIC simple analog predistorter which is shown in Fig.8. Gain expansion is observed when V_c is lower than –1V. Insertion loss (IL) is less than 5dB for a gain expansion of 2dB. Phase compensation was obtained from the MMIC predistorter as a result of the use of two inductors.

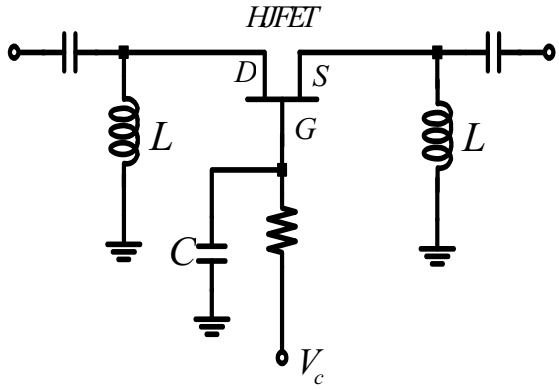

Fig. 8. Schematic of the MMIC predistorter.

The block diagram of a compound cuber predistortion system is shown in Fig.9, in which the input signal is split into two paths, and recombined in 180° phase shift at the output preceding PA (Morris & McGeehan, 2000). The key point of cuber predistorter is that the distortion terms can be scaled and phase shifted independently from the original undistorted input signal. Since the out of phase path can be set only for the third-order term, only the distortion term can be cancelled. For the reasons, this system is sometimes called a "cuber". However, there is a significant insertion loss in the combiner and splitter. Note that the lower coupling factors into and out of the cuber will result in a few losses in the main path. The independent two paths for high levels of IMD correction need a good gain and phase match.

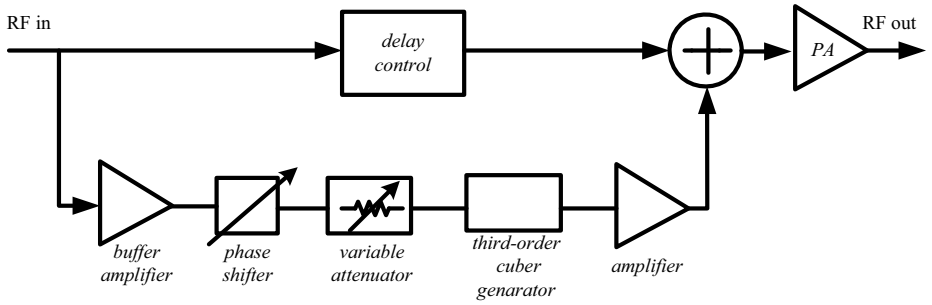

Fig. 9. Block diagram of a compound cuber predistorter.

4.2 DSP predistortion techniques

This approach is attractive since most modern radio frequency transceivers employ some form of DSP in their baseband processing as illustrated in Fig. 10.

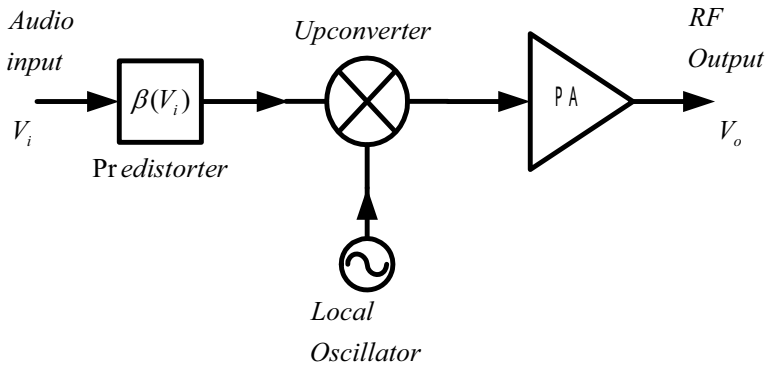

Fig. 10. Baseband predistortion system.

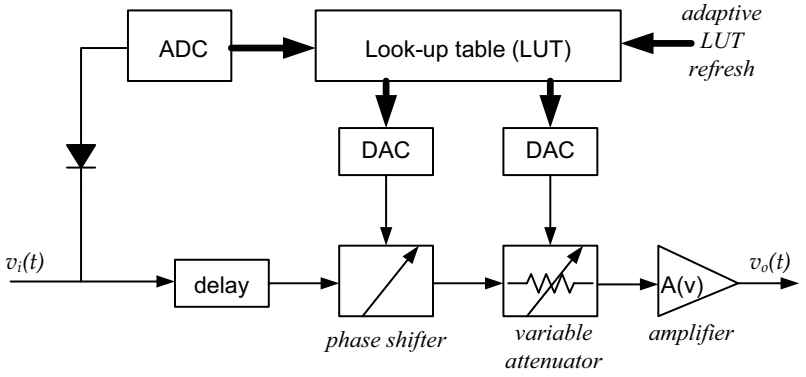

Fig. 11. DSP look-up table predistortion scheme.

A DSP look-up table predistortion system illustrates in Fig. 11. It should be noted that the system employs an input signal delay element to compensate the processing delays in the detection and DSP signal processing. The main limitation of the scheme is the speed of the detection and DSP itself.
The correction signals contain multiple harmonics of the baseband signal in order to perform the necessary predistortion function, which imposes a stringent requirement on the data converters. The precision of the look-up table is an important issue, which it can be implemented either physically or by a suitable algorithm. Moreover, the envelope input sensing is also a difficult task when the input signal throughputs continue rising. Note that a trade-off between the precision of detection process and the number of RF cycles employed to determine the final detector output is existed for the classical envelop detectors.

5. Linearity Improvement Circuit Techniques

Modern communication standards employ bandwidth-efficient modulation schemes such as non-constant envelope modulation techniques to prevent spectral re-growth problem, AM-AM, and AM-PM distortions, which means that some extra circuits for linearization purpose in power amplifier design are required.
Nevertheless, employing a linear PA's is a straightforward approach whereas it is also an inefficient method to meet the requirement of linearity. By taking advantage of the characteristics of high efficiency and applying some linearization techniques, nonlinear PA's may be a promising alternative. In this section, we investigate two transistor-level linear techniques to improve linearity of CMOS PA's namely, one is the nonlinear capacitance compensation scheme and the other is a parallel inductor compensation scheme. These two approaches will be described in the following subsections.

5.1 Nonlinear capacitance compensation technique

A deep sub-micron MOSFET RF large signal model that incorporates a new breakdown current model and drain-to-substrate nonlinear coupling is shown in Fig. 12 (Heo et al., 2000). This model includes a new breakdown current I_{dsB} with breakdown voltage turnover behavior and a new nonlinear coupling network of a series connection of C_{dd} and R_{dd} between the drain and a lossy substrate. The robustness of the new nonlinear deep sub-micron MOSFET model has been verified through load-pull measurements including IMD and harmonics at different termination impedance and bias conditions.

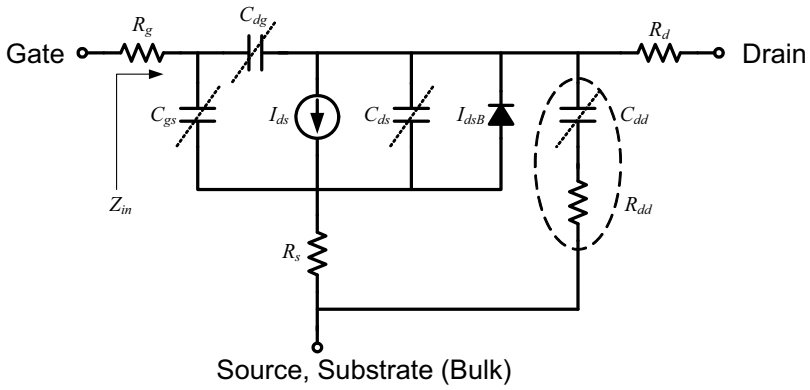

Fig. 12. Equivalent nonlinear model of a deep sub-micron NMOS device (slashed components are bias dependent).

A nonlinear capacitance cancellation technique to cancel the bias dependent input capacitance of the amplifier has been proposed and a prototype single-stage amplifier with a measured drain efficiency of 40% and a power gain of 7dB at 1.9GHz was reported in (Wang et al., 2001). The measured results indicate that the amplifier with nonlinear capacitance compensation has at least 6-dB IM_3 improvement in a wide range of output powers compared with the original amplifier without compensation.
The idea of the nonlinear capacitor compensation technique is that during the drain current clipping when the input signal is large enough to turn device 'on' and 'off', which the dramatical change in C_{GS} will generate distortion since Z_{in} is not keeping constant in signal amplification. The input impedance of the amplifier is approximately (ignore the R_S)

$$Z_{in} = \frac{1}{j\omega C_{in}} = \frac{1}{j\omega\left(C_{GS} + (1+A)C_{GD}\right)} \tag{6}$$

in which A is the voltage gain of the amplifier, the relationship of V_{GS} and V_{in} is

$$V_{GS} = \frac{Z_{in}}{Z_g + Z_{in}} V_{in} \tag{7}$$

and V_{GS} is a linear and delayed version of V_{in} on linear amplification

$$V_{GS}(t) = CV_{in}(t - t_0) \tag{8}$$

Note that by introducing a parallel inverse nonlinear characteristic component at the input of the amplifier can reduce the distortion, which a PMOS capacitance can be a good choice to compensate the nonlinearity of NMOS input capacitance. In other words, the input impedance Z_{in} is near a constant for a wide range of V_{GS} due to the inverse characteristic of the PMOS capacitance from the NMOS counterpart. The Behavior of NMOS C_{GS} and C_{GD} in

different operation region is shown in Fig.13 (Razavi, 2000), where W is the width of the NMOS device, L is the effective length of the NMOS device. C_{OX} is the oxide capacitance per unit width, and the overlap capacitance per unit width is denoted by C_{OV}. If the device is off, $C_{GD} = C_{GS} = WC_{OV}$ and the gate-bulk capacitance comprises the series combination of the gate oxide capacitance and the depletion region capacitance.

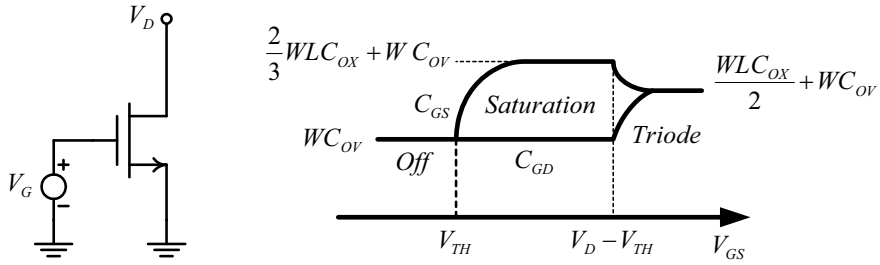

Fig. 13. Variation of gate-source and gate-drain capacitance versus V_{GS}.

If the device is operating at triode region, such that S and D have approximately equal voltages, then the gate-channel (WLC_{OX}) is divided equally and $C_{GD}=C_{GS}= (WLC_{OX})/2+WC_{OV}$. On the other hand, the gate-drain capacitance of a MOSFET is roughly equal to WC_{OV} for the saturation mode operation. The potential difference between the gate and channel varying from V_{TH} at the source to V_D-V_{TH} at the pinch-off point results in a non-uniform vertical electric field in the gate oxide along the channel. It can be proved that the gate-source capacitance equals to $(2/3)WLC_{OX}$ (Muller & Kamins, 1986). Thus, $C_{GS}=(2/3)WLC_{OX}+WC_{OV}$.
The dependence of a p-substrate MOS capacitance on voltage is shown in Fig.14 (Singh, 1994), in which V_{fb} represents flat-band voltage and V_T represents threshold voltage. In accumulation region (negative V_G), the holes accumulate at the oxide-semiconductor interface. Because holes are majority carriers, the response time is fast enough. As the gate voltage becomes positive, the interface is depleted of holes and attracts minority carriers. The depletion capacitance becomes important in this region. When the device gets more and more depleted, the value of C_{MOS} decreases to C_{MOS}(min).
At inversion condition, the depletion width reaches its maximum width. If the bias increases further, the free electrons in the p-substrate start to collect in the inversion region, whereas the depletion width remains unchanged with bias. The required excess free electrons are introduced into the channel by electron-hole generation. Since the generation process takes a certain amount of time, the inversion sheet charge can follow the bias voltage only if the voltage change speed is slow. If the variations are fast, the electron-hole generation cannot catch up the variations. The capacitance due to the free electrons has no contribution and the MOS capacitance is dominated by the original depletion capacitance. Therefore, under high-frequency conditions, the capacitance does not show a turnaround and remains at the C_{MOS}(min) as shown in Fig. 14.

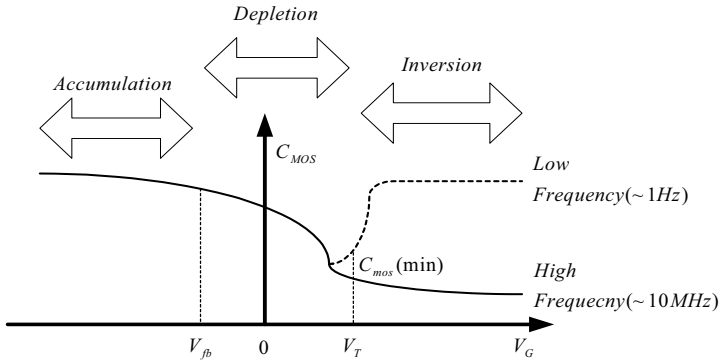

Fig. 14. Dependence of a P-substrate MOS capacitor versus voltage.

5.2 PMOS capacitance compensation technique

As shown in Fig.15, we can use this inverse capacitance characteristic to compensate the nonlinearity of NMOS input capacitance.

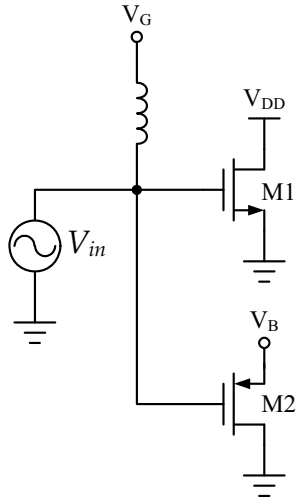

Fig. 15. Schematic of the PMOS capacitance compensation PA.

The Hspice simulation results of the NMOS and PMOS input capacitance (C_{gs} and C_{gd}) are shown in Fig.16 (a) and (b), respectively.

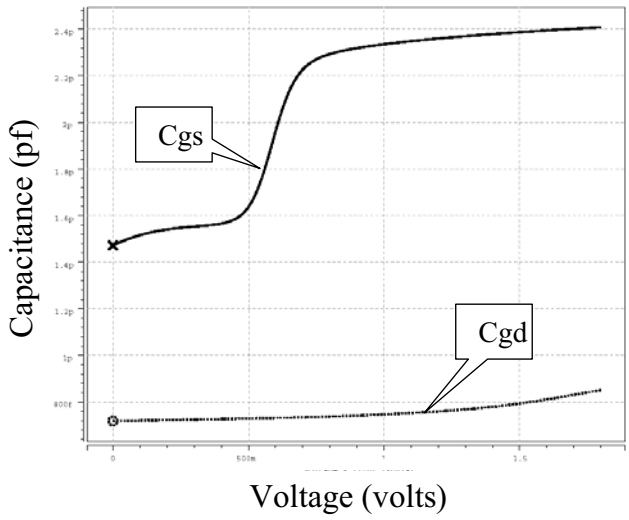

Fig. 16(a). Capacitances of C_{gs} and C_{gd} versus V_{gs} for NMOS device (W=1920μm, L=0.18μm).

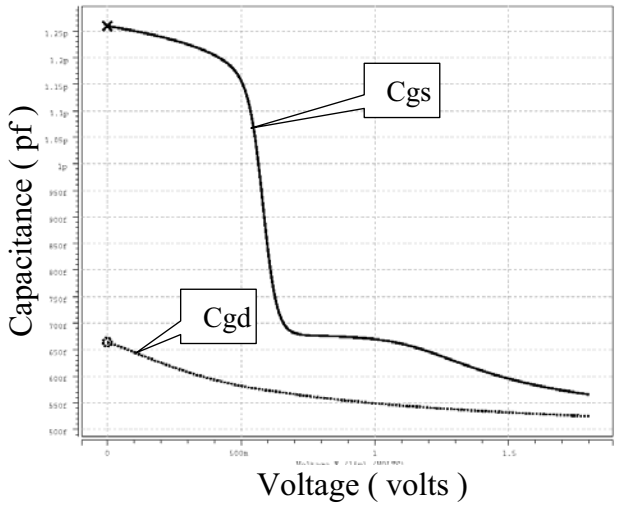

Fig. 16(b). Capacitances of C_{gs} and C_{gd} versus V_{gs} for PMOS device (W=1280μm, L=0.18μm).

The total input capacitance of the NMOS and PMOS devices is shown in Fig.17. Obviously, we can use this inverse capacitance characteristic to compensate the nonlinearity of NMOS input capacitance.

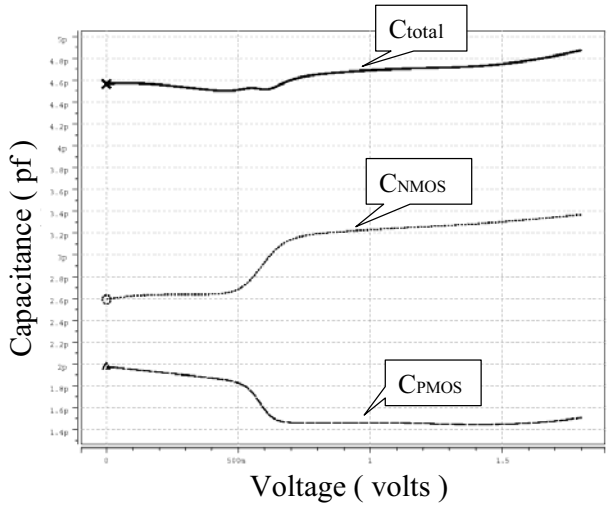

Fig. 17. Total gate input capacitance with PMOS capacitance compensation.

5.3 NMOS diode linearizer technique

The newly proposed approach is the diode linearizer which can be integrated in the PA design. The integrated diode linearizer in HBT PA can effectively improve the gain compression and phase distortion performances from the gate dc bias level (V_{GS}). Notice that the dc bias level decreases as the input power increases. A PA uses an integrated diode-connected NMOS transistor as the function of diode linearizer is shown in Fig.18. A similar technique by using nonlinear capacitance cancellation in CMOS PA designs has been reported in (Yen & Chuang, 2003).

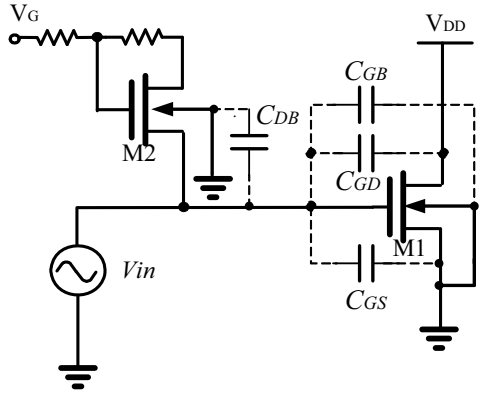

Fig. 18. Schematic of NMOS diode linearizer PA with parasitic capacitors.

For a first-order approximation, the oxide–related gate capacitances C_{GS}, C_{GD}, and C_{GB} of M1 are given by (Massobrio & Antognetti, 1993)

$$C_{GS} = \frac{2}{3} C_{OX} + C_{GS0} W \tag{9}$$

$$C_{GD} = C_{GD0} W \tag{10}$$

$$C_{GB} = C_{GB0} L_{eff} \tag{11}$$

for M1 operating at the saturation mode, in which L_{eff} is the effective channel length, W is the width of the channel, C_{OX} is the gate oxide capacitance, C_{GS0}, C_{GD0}, C_{GB0} are the voltage-independent overlap capacitances per meter among the gate and the other terminals outside the channel region and

$$C_{GS} = C_{OX} \left\{ 1 - \left[\frac{V_{GS} - V_{DS} - V_{TH}}{2(V_{GS} - V_{TH}) - V_{DS}} \right]^2 \right\} + C_{GS0} W \tag{12}$$

$$C_{GD} = C_{OX} \left\{ 1 - \left[\frac{V_{GS} - V_{TH}}{2(V_{GS} - V_{TH}) - V_{DS}} \right]^2 \right\} + C_{GD0} W \tag{13}$$

$$C_{GB} = C_{GB0} L_{eff} \tag{14}$$

for M1 operating at the triode mode, where V_{TH} is threshold voltage.
On the other hand, due to the depletion charge surrounding the respective drain diffusion region embedded in the substrate, the junction capacitance C_{DB} of M2 is given by

$$C_{DB} = \frac{C_j A_D}{\left(1 + V_{DB} / \varphi_j\right)^{m_j}} + \frac{C_{jsw} P_D}{\left(1 + V_{DB} / \varphi_j\right)^{m_{jsw}}} \tag{15}$$

in which C_j and C_{jsw} are the capacitances at zero-bias voltage for square meter of area and for meter of perimeter, respectively, m_j and m_{jsw} are the substrate-junction and perimeter capacitance grading coefficients, φ_j is the junction potential, and drain-to-gate overlap capacitance C_{DG} of M2 can be described as

$$C_{DG} = C_{DG0} W \tag{16}$$

Notice that the input-voltage-dependent capacitances C_{GS}, C_{GD} of M1 indicated in (12) and (13) increase with an increase of V_{GS} whereas the junction capacitance C_{DB} of M2 described in (15) decreases with an increase of V_{DB} (=V_{GS} of M1). Therefore, with a proper choice of the dimensions of M1 and M2, a near constant total input capacitance can be achieved.
Fig. 19 shows the simulation results of the NMOS gate capacitance (C_{GS}, C_{GD}, and C_{GB}) and the NMOS diode total capacitance at drain (C_{DG} and C_{DB}). The total input capacitance of these two devices has flat curve characteristic at each V_{GS}. Clearly, it also implies the distortion due to the nonlinearity of the input capacitance can be reduced.

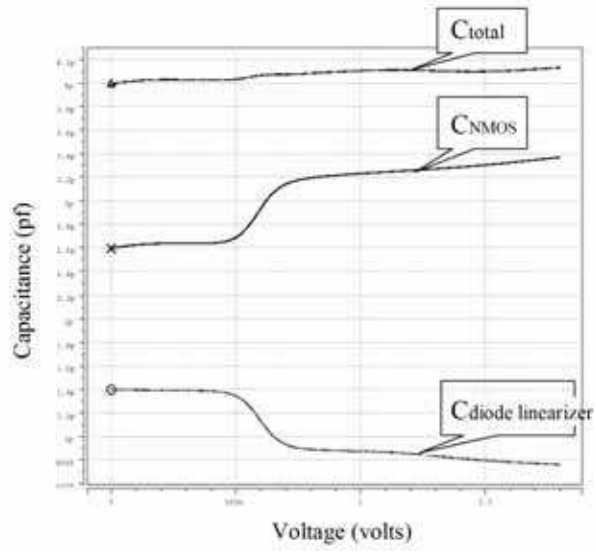

Fig. 19. Total gate input capacitance with diode linearizer for TSMC 1.8V RF MOS device.

Also, the P1dB simulation results indicate this diode-linearizer bias technique can improve 2-dB linear gain than the conventional resistance bias approach as shown in Fig.20. Note that the device dimensions of the CMOS PA can reach millimeter scale, which implies that the parasitic capacitances C_{p1} and C_{p2} can degrade the gain and power-added efficiency of the PA due to the large parasitic capacitances (Jeffrey et al., 2001).

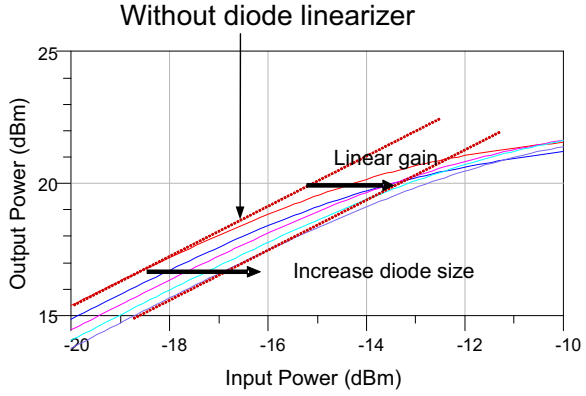

Fig. 20. P1dB simulation results of the diode-linearizer bias approach for different diode size.

5.4 Parallel inductor compensation diode technique

CMOS cascode amplifier architecture with the parallel inductor is shown in Fig. 21. Notice that the large device sizes of CMOS PA can lead to large parasitic capacitances, C_{p1} and C_{p2},