A Fully-Integrated Quad-Band GSM/GPRS CMOS Power Amplifier

Ichiro Aoki, *Member, IEEE*, Scott Kee, *Member, IEEE*, Rahul Magoon, *Member, IEEE*, Roberto Aparicio, *Member, IEEE*, Florian Bohn, *Student Member, IEEE*, Jeff Zachan, *Member, IEEE*, Geoff Hatcher, Donald McClymont, and Ali Hajimiri, *Member, IEEE*

***Abstract*—Concentric distributed active transformers (DAT) are used to implement a fully-integrated quad-band power amplifier (PA) in a standard 130 nm CMOS process. The DAT enables the power amplifier to integrate the input and output matching networks on the same silicon die. The PA integrates on-chip closed-loop power control and operates under supply voltages from 2.9 V to 5.5 V in a standard micro-lead-frame package. It shows no oscillations, degradation, or failures for over 2000 hours of operation with a supply of 6 V at 135 °C under a VSWR of 15:1 at all phase angles and has also been tested for more than 2 million device-hours (with ongoing reliability monitoring) without a single failure under nominal operation conditions. It produces up to +35 dBm of RF power with power-added efficiency of 51%.**

***Index Terms*—CMOSFET power amplifiers, CMOS RF, distributed active transformer (DAT), global system for mobile communications (GSM), power control, reliability.**

I. Introduction

THERE is a strong drive toward handheld communication devices with fully integrated CMOS components. Although there are obvious cost, size, and repeatability advantages, implementing fully integrated power amplifiers with no external matching in standard CMOS is particularly challenging due to the large supply variations and harsh load mismatches present in handheld devices. Such integration enables many new opportunities for low-cost wireless solutions for a broad range of applications with no need for individual fine tuning of the blocks.

The power amplifier (PA) has been one of the last bastions of exotic technologies in wireless systems. The perceived need for exotic technologies fundamentally arises from the requirement of delivering RF power in excess of +30 dBm to a 50 Ω load at a high operation frequency. This bottleneck requires a peak-to-peak voltage swing in excess of 20 V at the nominal 50 Ω load; potentially higher voltages can occur under certain VSWR conditions. This usually leads the designer to using a device technology capable of handling such voltage swings and/or providing impedance transformation means to lower the 50 Ω nominal impedance to a lower value that requires a less severe voltage swing.

Manuscript received April 12, 2008; revised June 19, 2008. Current version published December 10, 2008.

I. Aoki and S. Kee are with AyDeeKay LLC, San Clemente, CA 92673 USA.

R. Magoon, R. Aparicio, J. Zachan, G. Hatcher, and D. McClymont are with Axiom Microdevices Inc., Irvine, CA 92618 USA.

F. Bohn and A. Hajimiri are with the California Institute of Technology, Pasadena, CA 91125 USA (e-mail: hajimiri@caltech.edu).

Digital Object Identifier 10.1109/JSSC.2008.2004870

Process technologies providing fast, high-breakdown voltage transistors for power amplifiers include GaAs HBT processes [1]–[5], LDMOS processes [6], [7], SiGe HBT processes [8], [9] or other customized processes [10], [11]. Despite superior breakdown voltages compared to Si CMOS technologies for similar RF performance, these approaches often also require off-chip passive components (e.g., inductors and couplers). To streamline the usage by hand-set manufacturers, these components are typically included with the core amplifier IC on a module, where the die is mounted on an organic or ceramic substrate that typically contains the output matching network as well, as depicted in Fig. 1(a). Despite the cost disadvantage and diminished prospects for integration with the rest of the radio, off-chip passive components and compound semiconductors are used in the absence of any other viable alternative.

CMOS technology, on the other hand, has been recognized as a desirable alternative [12]–[15], offering cost advantages and prospects for integration with the rest of the radio. However, the challenge of fully integrating the required output impedance matching network is greatest in CMOS technologies as required impedance transformation ratios are highest due to the lower device breakdown voltage resulting in comparably higher losses as reactive currents in energy storing elements increase. This situation is exacerbated in CMOS by inductors of lower quality compared with other technologies due to the increasingly highly-doped, conductive substrates. Although watt-level CMOS PAs have been reported in earlier works [12]–[14] dating back to the late 1990s, those PAs as well as more recent research implementations (with the exception of the DAT concept [16]–[20]) require off-chip baluns [21], [22] to meet output power requirement alone or require bonding-wire inductors and large voltage stresses across the transistors [23]. Therefore, fully integrated, transformer-based CMOS PAs typically target standards with lower output power requirements [24]. Other CMOS PA with integrated output matching circuitry show much lower output power levels [25], [26].

This paper will show how one can overcome the challenges associated with the realization of a fully integrated commercially viable CMOS PA at RF and microwave frequencies by moving away from standard approaches without having to use module technologies or exotic substrates. The techniques developed make it possible to have a fully-integrated CMOS PA in a standard lead-frame package, as shown schematically in Fig. 1(b). In Section II, we will review some of the dominant breakdown mechanisms in CMOS. In Section III, we will elaborate on two main challenges posed by the low breakdown voltages in MOS transistors. Section IV demonstrates how concen-

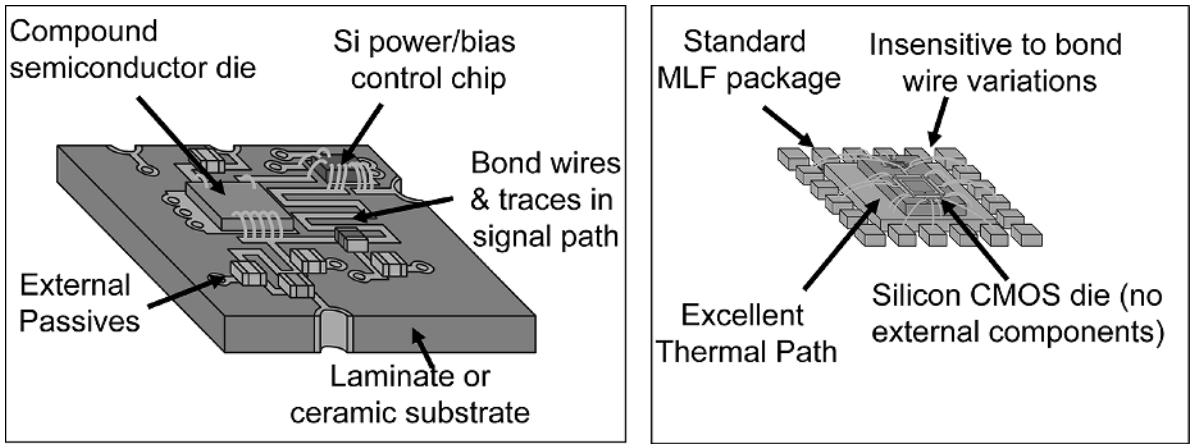


Fig. 1. The module-based compound semiconductor PA versus fully-integrated CMOS PA.

tric DATs can be used to overcome both problems at the same time. We will talk about power control and the input drive circuitry in VI and Section VII, respectively. Finally, we discuss the final product and the measurement results in Section VIII.

II. CMOS Breakdown Mechanisms

Scaling of MOS transistors has been the primary enabling technology behind the self-fulfilling prophecy of Moore's law [27]. Transistor scaling is ideally done by maintaining the same electric field strength and profile inside the transistor, while the physical dimensions of the MOSFET are scaled down [28], [29]. This results in an inevitable reduction in the maximum allowable voltages on the transistor [30]–[32], which is achieved by migrating to smaller power supply voltages to avoid breakdown and performance degradation.

Although digital circuits have coped well with this voltage scaling, analog designers have faced additional hurdles in dealing with scaling. As we will see in Section III, this poses major challenges to use of conventional PA architectures in CMOS, particularly in battery operated handheld units. In order to devise reliable solutions to this problem we must gain a better understanding of the nature of various breakdown and stress-related degradation mechanisms in MOS transistors, as discussed next:

A. Gate Oxide Breakdown

Gate oxide breakdown leads to a catastrophic failure in a MOSFET, resulting in permanent damage to the gate oxide. This failure is caused by tunneling current due to electric field across the gate oxide, which can result in defects within the oxide or at the silicon/oxide interface. The presence of these defects typically results in increased current in the affected region, potentially leading to destructive runaway. Once this occurs, the ruptured oxide can result in an ohmic connection between the gate and the underlying silicon. Considering that the gate oxide is only a few atomic layers thick (e.g., 2 nm in 130 nm CMOS process), it takes a relatively small voltage for such a catastrophic gate oxide breakdown to occur.

Although it is often assumed due to the sudden onset of destruction that this breakdown occurs instantly at a predetermined "breakdown voltage," in fact the failure is a probabilistic event, with increasing probability as voltage, stress time, or oxide area are increased. Even ignoring any areal manufacturing variation in oxide thickness/quality, each unit area of oxide has an independent probability of failure in a given stress level and duration. Even so, due to the particularly strong dependence on voltage (which affects both the number of damage-causing carriers and the energy of each carrier), there is only a narrow range of voltage where the variation due to time and device area has any practical impact, and designers can to first order assume that the design is robust if the oxide voltage is kept sufficiently low.

The allowable gate-oxide voltage becomes smaller as the gate oxide thickness is reduced in each process generation. The introduction of other dielectric materials in the gate can alleviate this problem, but the underlying trade-off between transconductance (which increases with areal gate capacitance) and gate breakdown voltage remains a major challenge to non-digital circuits.

As can be seen in Fig. 2, the voltage at the bottom of the gate oxide varies between the source and drain. As a result, the oxide stress is a function of position. The highest stress areas occur at the source and drain oxide edges and so from a design standpoint, we must ensure that the gate-source and the gate-drain voltages (V_{GS} and V_{GD}) never exceed pre-specified values, *by design*.

B. Hot Carrier Degradation

Hot carrier degradation occurs when the carriers in the channel accelerate in the electric field caused by the drain-source voltage. In short channel devices, this field can be extremely high, and so carriers can achieve a very large energy before losing momentum to a collision with the crystal lattice.

Some of these "hot" charge carriers collide with the lattice before arriving at the drain[1] with sufficient energy to result in impact ionization, as illustrated in Fig. 3. This, in addition to potential avalanche multiplication, can result in surface defects, resulting in a reduced carrier mobility in the channel. This can also result in trapped charge in the gate oxide or oxide/silicon interface, shifting the local threshold voltage. Usually this damage is occurring in the drain region where the electric field is very high, causing the damage to manifest itself as an increase in

[1]All these hot carriers eventually decelerate when they get to the drain, because the mean-free path is much smaller in the highly doped drain.

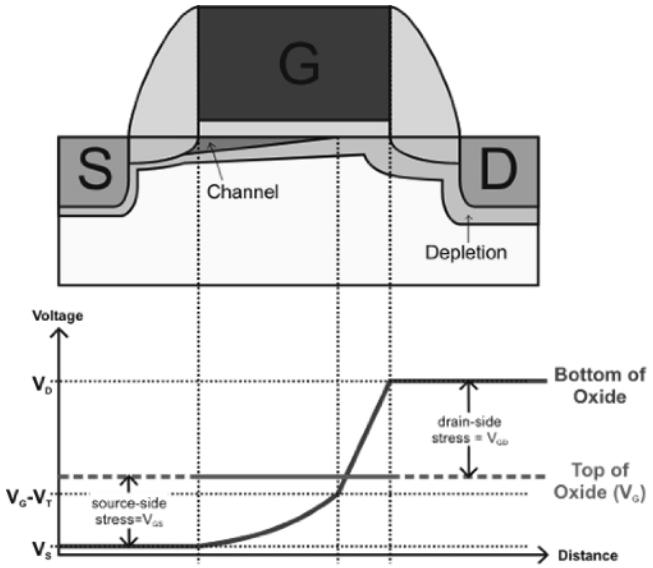


Fig. 2. Gate oxide breakdown in a MOS transistor.

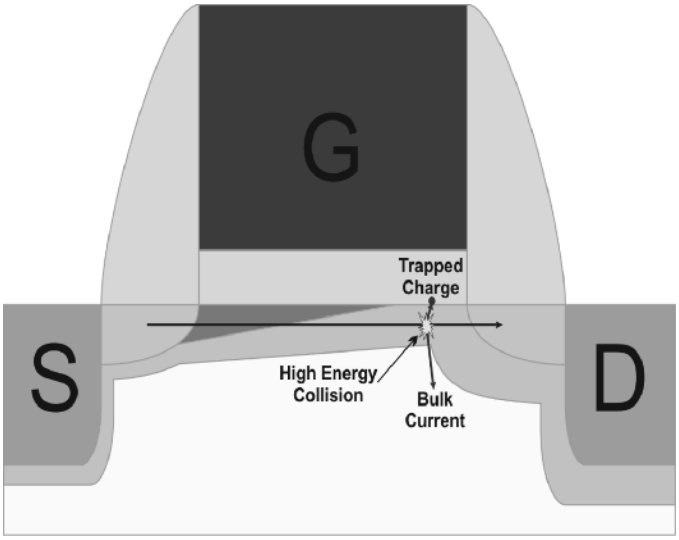


Fig. 3. Hot carrier degradation in a MOS transistor.

on-resistance and knee voltage, reducing the power amplifier performance.

Unlike the gate oxide breakdown, the hot carrier degradation is not intrinsically catastrophic. Instead, it can cause a gradual degradation in the device performance over a period of time. Although, the hot carrier degradation is caused by a different mechanism than gate-oxide breakdown, the damage due to hot carriers can result in increased rate of gate oxide damage.

For the hot carrier degradation to be noticeable, it is necessary to have high drain-source voltage, V_{DS}, and substantial drain current, I_D, at the same time. Thus, from a circuit design perspective, the hot carrier degradation can be prevented by avoiding channel current when drain voltage is high. This happens ordinarily in a switching amplifier, where the high efficiency is achieved by having a small transistor voltage while the transistor conducts current and vice versa, making switching power amplifiers less prone to hot carrier degradation. However, this can be a serious issue in a linear power amplifier.

C. Punch-Through

Under normal circumstances the drain-source current of a MOS transistor flows close to the surface when a large enough gate bias is applied to create a channel by inversion. In the absence of a gate bias, very little current flows because the drain-bulk and the source-bulk diodes are effectively connected back-to-back with opposite polarities. As larger voltages are applied to the drain, the drain-bulk depletion region extends farther to accommodate the electric potential drop. This depletion region will eventually extend all the way to meet the depletion region of the source-bulk junction, thereby diminishing the potential barrier that stops the direct flow of current between the drain and the source, as depicted in Fig. 4(a). This punch-though process results in a current flow that can occur even in the absence of any significant gate bias (Fig. 4(b)).

In practice there is no need for the depletion regions to touch. Even when they approach each other, the potential barrier height drops, increasing current from one side to the other. Punch-though is exacerbated by smaller channel length, L, and larger V_{DS}. The significance of punch-through reduces rapidly with increasing L.

Punch-through is not intrinsically destructive, although the simultaneous occupance of high voltage and high current can easily result in thermal failure if sustained. However, the process can generate its own hot carriers, potentially causing similar reliability issues as previously discussed. From a design perspective, punch-through can be dealt with by making sure that V_{DS} of a single transistor remains at all times below a certain pre-specified value for any given channel length.

D. Drain-Bulk Breakdown

In a standard CMOS process the bulk is connected to a fixed electrical potential and the drain-bulk diode experiences a reverse bias directly proportional to the absolute drain voltage, V_D. This diode has a reverse breakdown voltage primarily determined by the doping of the bulk (the lightly doped side). Therefore, it is important to make sure that V_D does not reach this breakdown voltage in addition to maintaining V_{DS} smaller than its critical level. Fortunately, this breakdown voltage is relatively high in today's CMOS processes (e.g., greater than 10 V in 130 nm CMOS).

As mentioned in the beginning of this section, it is essential to be aware of these breakdown mechanisms and deal with them through design to be able to implement a robust fully-integrated PA in CMOS. Next we will talk about the circuit challenges and solutions.

III. Design Challenges Due to Breakdown

Low breakdown voltages of MOS transistors pose two categories of challenges for PAs used in handset cellular applications. The first one is due to the high output power requirement of many wireless applications. The second challenge is caused by the battery technology where the unit cell voltage is fixed and does not scale. We will address these challenges and how they can be overcome in this and Section IV.

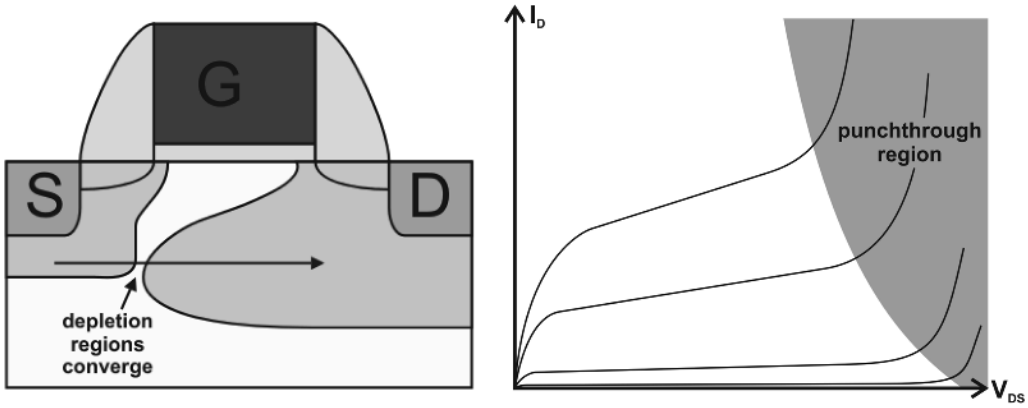


Fig. 4. (a) Punchthrough in a MOS transistor. (b) The effect on the *I-V* curve.

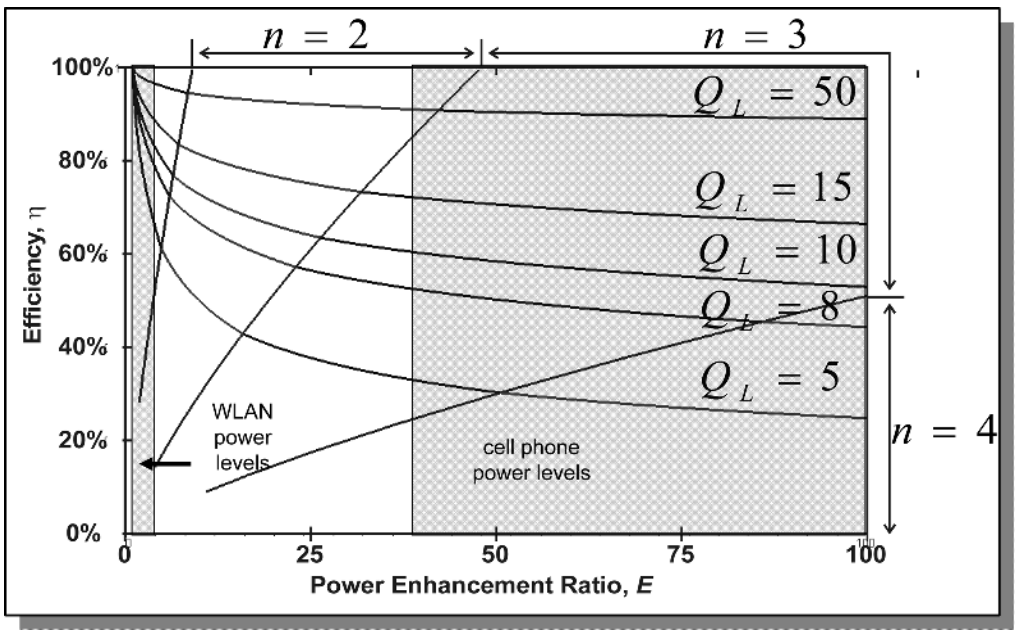


Fig. 5. Passive efficiency versus PER.

A. High Power Generation Using Low Voltage Transistors

For many applications the PA has to generate several watts of RF power. For example, considering the losses in the diplexer following the PA in a 900 MHz GSM transmitter, the PA must produce up to +35 dBm ($\sim$3.1 W at its output). This corresponds to a peak-to-peak voltage swing of approximately 36 V_{pp} across a 50 Ω load.

On the other hand, assuming that we are not willing to let the drain-source and drain-gate of a 130 nm CMOS transistor to experience a peak voltage greater than 1.5 V, the drain RF voltage amplitude cannot exceed 750 mV, assuming a sinusoidal voltage waveform. If this voltage directly appears across a 50 Ω resistor, it delivers only 5.6 mW of power to the load. This major disparity can be captured by the very large power enhancement ratio (PER) defined as the ratio of the required power into the load, $P_{\rm req}$ (e.g., 3.1 W for GSM) to the maximum power directly deliverable to the load by the transistor assuming the same voltage swing, $P_{\rm dir}$ (e.g., 5.6 mW for a single 130 nm drive NMOS), i.e., [18]

$$\mathrm{PER} = \frac{P_{\rm req}}{P_{\rm dir}}. \tag{1}$$

This necessitates the use of an impedance matching network that can present a smaller impedance to the transistor, so it can deliver a larger power by driving a lower impedance load at a lower voltage and higher current swing.

The impedance transformation can be achieved several different ways. LC matching networks are common ways of achieving this. The challenge in this case is that for a given inductor quality factor (assuming lossless capacitor), the *passive* efficiency of the matching network drops with increasing required PER. The optimum number of LC sections depends on the unloaded inductor Q and the PER. The maximum achievable passive efficiency, η_p, with the optimum number of stages for a given PER and unloaded inductor Q is shown in Fig. 5 [18].

As can be seen from Fig. 5, for large values of PER and typical on-chip Q's, the passive efficiency alone drops to prohibitively low values, even in the absence of *any* energy loss in the active device. (In our earlier numerical example with $P_{\rm dir} = 5.6$ mW and $P_{\rm rec} = 3.1$ W, the PER is in excess of 500.) The LC-matching approach may be suitable in applications with

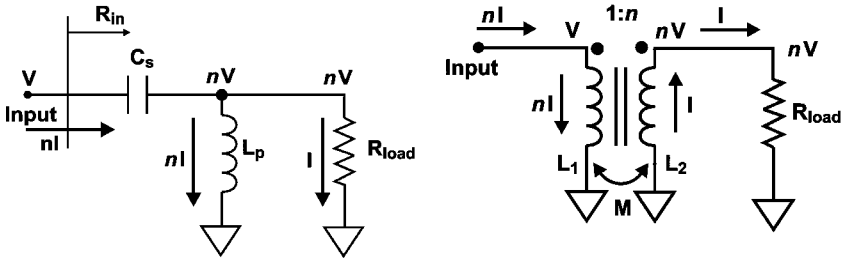


Fig. 6. (a) Resonant LC match. (b) Transformer match.

lower output power specifications (e.g., wireless LAN) due to their lower required PER. However, the large required PER for a fully-integrated watt-level CMOS power amplifier that includes an on-chip output matching circuit poses a serious challenge to integration.

Transformers have an advantage over LC matching section because they store less energy in the inductance for a given transformation ratio [18]. To demonstrate this point let us compare the single-section LC resonant matching circuit shown in Fig. 6(a), where the load resistance $R_{\rm load}$ is match to the output of the transistor, using an LC section. Assuming that the voltage across the load resistor is nV and the current through $R_{\rm load}$ is I and that the input voltage amplitude is V and the input current amplitude is nI, the total energy stored in the inductor is n^2IV and the impedance transformation ratio is n^2. On the contrary if the same impedance transformation ratio is to be achieved using a transformer, as depicted in Fig. 6(b), the total energy stored in the transformer will be nIV which is n-fold smaller than the case of the LC match. The lower stored energy in a transformer results in a smaller power loss for a given Q, thus the advantage of the transformer matching over LC matching.

A direct implementation of an on-chip transformer may be an acceptable option for applications with lower required power levels and PERs, where the primary and secondary inductors are comparable (small impedance transformation ratio). However, in watt-level CMOS PAs, due to the large required impedance transformation ratios (and PERs), there is a large disparity between the inductance and the physical sizes of the primary and the secondary inductors. A large ratio between the primary and secondary inductance results in a lower effective quality factor and degrades the efficiency.

In Section IV, we will see how this challenge can be overcome using a distributed active transformer (DAT) which uses several 1:1 transformers to combine impedance transformation and power combining in the same structure, while always maintaining a low voltage swing across the transistors.

B. The Battery Voltage Challenge

The battery cell voltage depends on its chemistry and does not scale. This poses a challenge for mobile handsets that run off of a battery (today it is most commonly Li-ion batteries with a cell voltage of 3.6 V). Unlike many other circuits, the PA is usually run directly off the battery with no external regulation to preserve the system efficiency. This is further exacerbated by the fact that the phone has to be operational on the charger, which can apply voltages higher than that of the battery when charging. This means that the PA has to operate reliably from a high power supply under various conditions, such as load mismatch, elevated temperature, etc.

In a PA, the drain of a drive transistor can experience an AC signal greater than two-times the DC supply voltage applied to its drain. Unfortunately, submicron standard MOS transistor capable of operating at cellular frequencies cannot withstand such voltages, generally speaking. Using a cascode stage and dividing the voltage between those transistors can relax this requirement by a factor of two, but by itself cannot solve such a large disparity. In Section IV, we will see how this problem can be solved by using stacked concentric DATs.

IV. Distributed Active Transformer

DATs have been shown as the way to solve the first breakdown problem in a fully integrated fashion [16]–[20]. It makes it possible to make fully-integrated watt-level power amplfiers using low-breakdown voltage transistors. The DAT also allows alleviating the effect of the low quality factor passive elements used to integrate the output matching network.

A. Basic DAT

A single-concentric quad-core DAT is shown in Fig. 7. It consists of four differential cores that drive the four corners of the structure differentially. Thus, each primary slab inductor is driven differentially creating a virtual ground in the middle of the slab where the power supply is applied. The resonant capacitors for each inductor can be applied between the drains of two adjacent transistors being driven differentially, instead of being across the slab inductor itself. This can be done because the symmetry of the structure guarantees that it sees the same voltage as when it was place in parallel with the inductor. This makes it possible to create an alternating RF signal in the primary consisting of the slab inductors, capacitors, and transistors.

The RF power is then magnetically coupled to the secondary loop that consists of the secondaries of four 1:1 transformers in series. This allows the secondary voltages to add up in-phase in the voltage domain to achieve the high voltage levels necessary for a watt level output into the load (e.g., 36 $V_{\rm pp}$ for +35 dBm into a 50 Ω load). The large voltage swing on the secondary loop does not pose a reliability issue since no transistor is connected electrically to the output to experience it. More discussion on the detailed operation of DAT can be found in [16]–[20].

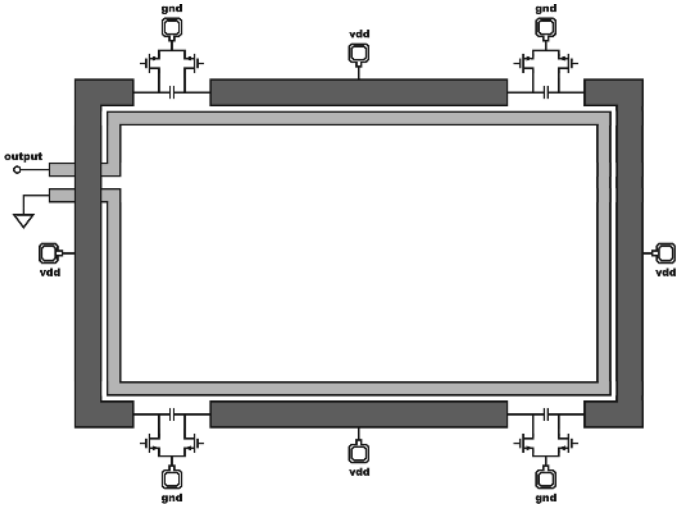


Fig. 7. Single quad-core distributed active transformer.

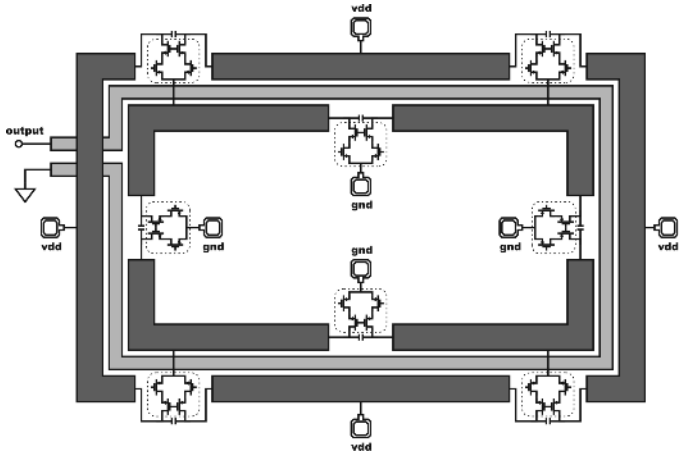


Fig. 8. Double-concentric quad-core DAT with cascode drivers.

In Fig. 7, the DC power supply voltage is provided to the amplifier via the mid-point of the primary slab inductors (a virtual ground). Since this DC voltage appears directly at the drain of the drive transistors, its maximum value is limited by transistor breakdown voltage. A cascode structure can be used to increase the supply voltages that can be tolerated by the transistors. Although the cascode helps by allowing a larger voltage to appear on the drain of the top transistor without degrading the device performance, it is not sufficient for a power amplifier that must withstand a peak voltage in excess of twice the V_{DD} with a high-cell voltage battery (e.g., Li-ion) on the charger under significant load mismatch (large VSWR) without any long-term performance degradation. Next we will see how a concentric DAT can be used to overcome this challenge.

B. Concentric DAT

Although the DAT can be used to produce watt-level output powers at low gigahertz frequency range with good efficiency, its original implementation is more conducive to low supply voltages. This is very useful for a low-battery-voltage solution that may become more commonplace in the future. However, it is not compatible with the higher voltages presented by the Li-ion batteries, which under charging situations can present supply voltages in excess of 3.6 V to the PA.

This problem can be solved by noticing that the mid-points of the slab inductors are at virtual grounds and as such can also serve as the ground connection of another DAT, concentric with the original one. This way, we can form the double-concentric DAT, as shown in double-concentric quad-core DAT example of Fig. 8. This arrangement allows for the same DC current to be shared by the inner and the outer DATs. This way, each DAT experiences roughly half of the supply voltage while each contributes to the total output RF power magnetically coupled to the secondary loop. It is noteworthy that this procedure can be repeated multiple times if necessary.

Another challenge for a fully integrated commercial GSM/GPRS CMOS power amplifier is the potential load mismatch presented to the output of the PA mainly due to the changes in the relative position of conductive objects with respect to the antenna. The PA has to keep meeting its specifications for relatively large voltage standing-wave ratios (VSWRs) (8:1 at the PA output) for all potential phase angles. It also has to remain operational and should not be permanently damaged or degraded under an even greater VSWR (15:1 at the PA output). Additionally, it must not oscillate and its supply current has to be kept below a predetermined maximum under any of these conditions, with the output power tightly controlled over a broad range of power levels. The distributed nature of the PA combined with its concentric design improves its robustness, as each transistor experiences smaller stress even under severe load mismatch conditions. This allows the PA to maintain performance and avoid damage under severe VSWR and other adverse effects.

V. Electromagnetic Simulations

In order to achieve a robust design, it is very important to accurately model all on-chip passive structures, in particular the DAT as it carries large AC and DC currents when the power amplifier is operating at full output power. The modeling challenges are further exacerbated as the DAT in each band occupies a large physical area and creates strong AC magnetic fluxes in its vicinity. As a result, other passive structures such as the harmonic filter inductor, the transformers, and even the bonding wires can pick up strong, potentially undesirable or even destructive interference. Furthermore, the large AC currents drawn create strong common-mode voltages wherever even a small amount of inductance is present. As these voltages could also negatively impact performance or even cause breakdown conditions, it is important to accurately simulate all passive circuitry.

The DAT and the amplifier output network are simulated in commercial, 2.5-D electromagnetic field-solver software (e.g., [33]). The S-parameter output data at the frequencies of interest is used in large-signal, nonlinear harmonic balance simulations to predict amplifier performance as well as to find design values of tuning elements. Both electric and magnetic coupling mechanisms from the DAT to DC supply lines and signal feed structures can also be identified and their impact on the overall circuit

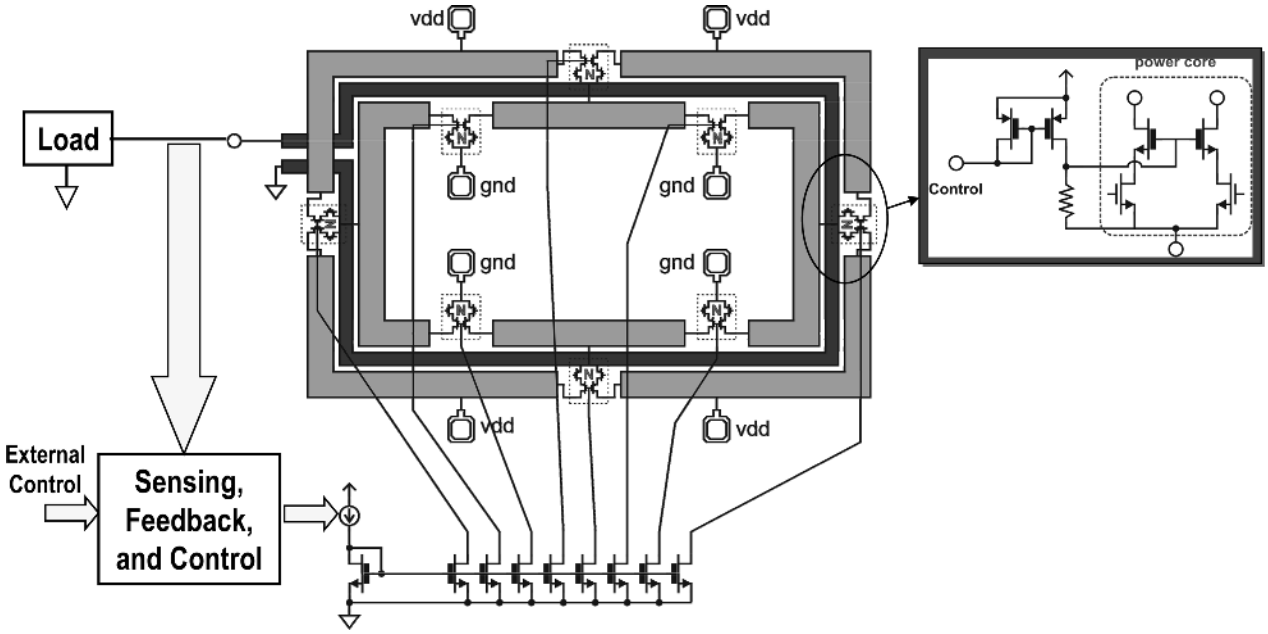


Fig. 9. The distributed power control system.

performance evaluated. In this way, a robust design can be arrived at, in which the coupling mechanisms are reduced to an acceptable minimum. Where necessary, the full-chip electromagnetic simulations were complemented by more accurate, finer grid-size EM simulations of partial structures to evaluate some of the more particular interactions such as between the DAT and the transformers.

The electromagnetic S-parameter simulation results are also used to derive a lumped-element approximation of the passive structures, enabling time-transient simulations including the passive structures to evaluate start-up transients as well as transmit mask requirements.

The impact of tuning capacitor and packaging variations was also evaluated and compared to RF circuit simulations. This also allows an assessment of circuit performance sensitivity to processing and packaging variations, which could negatively impact overall product yield.

VI. Power Control and Bias

Each wireless communication standard imposes stringent requirements on the transmitter under real world operating conditions. The GSM standard, requires accurate control of the power transmitted by the PA as well as the timing of each power burst in addition to placing restrictions on spectral emissions allowed within and outside the intended transmit frequency channel.

These collective specifications must be met under varying combined conditions of operating supply, temperature, frequency band, and output load mismatch experienced by the PA. Also, in order to minimize individual tuning of phone handsets in volume production, it is important that the part-to-part variation on many of these performances be minimal. With the exception of noise and RF isolation, most of these specifications require closed-loop control of the PA.

Typically, closed-loop power control is employed in a PA for such an application. A power detector, which is used to sense

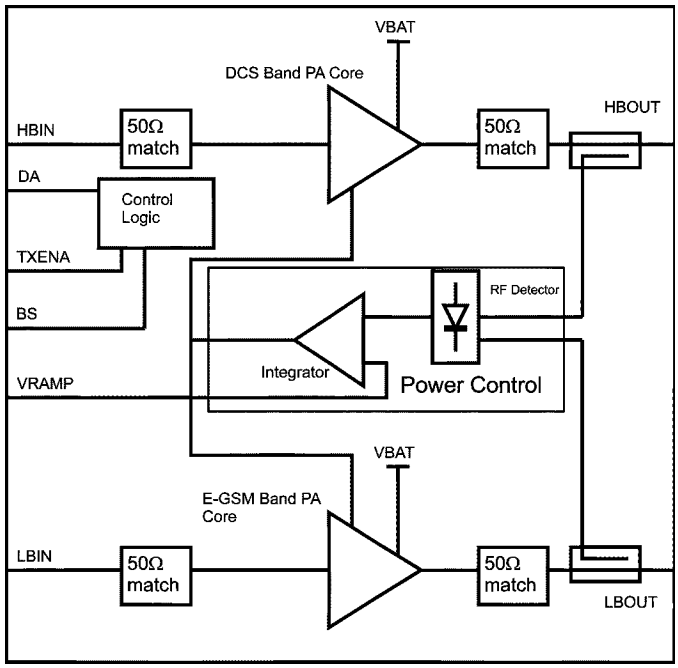


Fig. 10. Functional block diagram of the fully integrated quad-band CMOS.

the output of the PA, generates a signal that is then compared to an externally supplied reference in order to generate the signal that controls the PA. The closed loop power control accuracy is limited fundamentally by the precision of the sensor and by the loop gain that corrects the error. A typical closed-loop power control circuit thus needs to maintain its loop gain and employ a feedback block whose transfer function is as insensitive as possible to the changing operating conditions. The bandwidth of the power control loop is constrained on the one hand by transient requirements, which tend to require higher bandwidths, and on the other hand, by noise emission constraints which tend

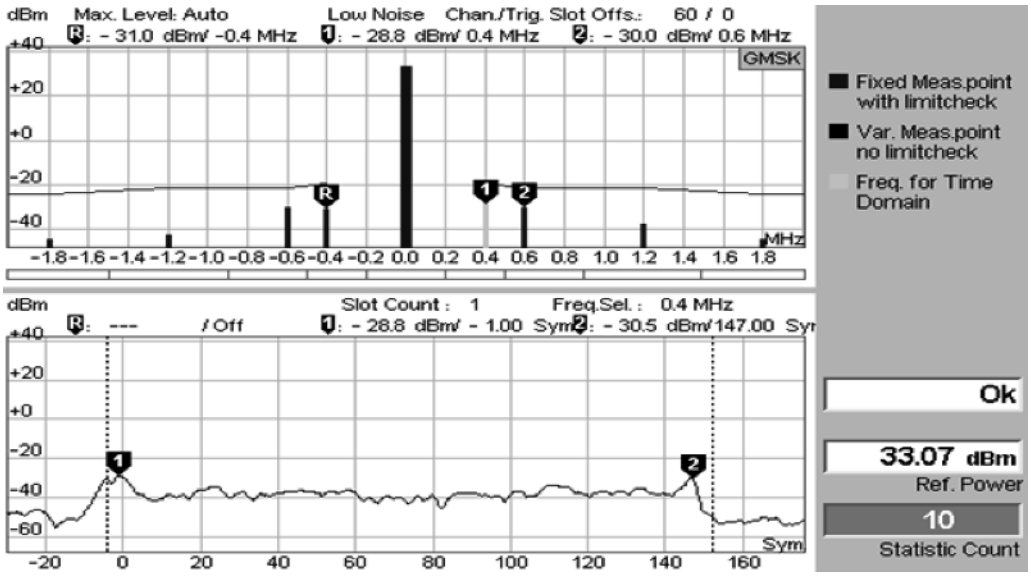


Fig. 11. The output switching transient.

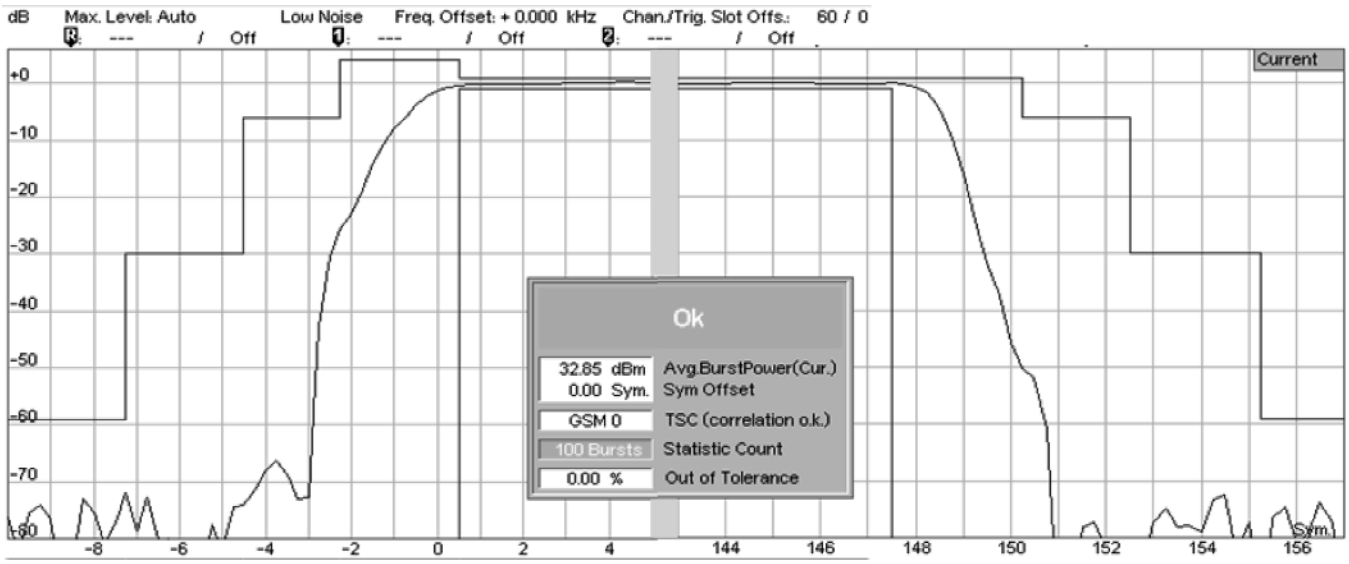


Fig. 12. The power-up and power-down time masks.

to require lower bandwidths. To the extent that the power control loop's open loop gain is kept constant over power level, the bandwidth remains constant.

The implementation of power control specifically for a DAT-based PA presents a number of unique challenges. Firstly, the controlled forward path is distributed, but drives a single output. Secondly, given that the DAT-based PA consists of two distributed concentric primaries which are connected in series in DC through the virtual grounds, half of the forward control elements are different from the other half. Finally, as is discussed in detail in the previous section, the presence of the DAT on the same die as input drive and control circuits dictates that particular attention must be paid to any trace on the die that has to cover any significant routing distance as it may pick up several volts of RF. This is particularly concerning given that the power control has a distributed function to perform as well.

The power control loop topology chosen employs current mode routing throughout the die in order to control the distributed forward elements, as shown in Fig. 9. The forward control paths are matched to each other. A single feedback detector is employed for each band in order to sense the output of the PA and its output is fed back to a single control loop error amplifier, again in current mode.

The power control loop operates under a voltage regulator whose output voltage is maintained constant under all supply conditions. This ensures that the loop dynamics are, to the first order, unaffected by changes in the battery voltage. Since the power control loop has to maintain spectral purity under a range of output load mismatch conditions, the feedback sensor used ensures that the control loop stays closed at all times. This allows the output of the PA to faithfully follow the reference ramping signal in both the time and modulation frequency domains under all conditions and meet the GSM output radio

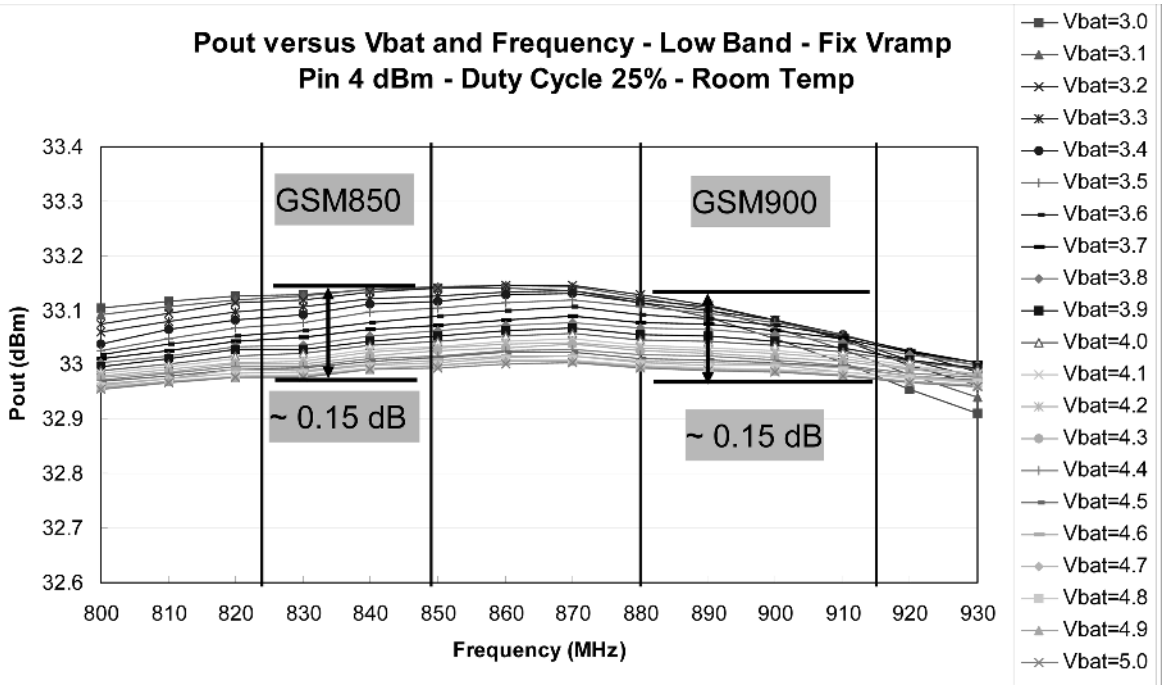


Fig. 13. Power control accuracy for different supply voltages.

frequency spectrum (ORFS) requirements, as demonstrated in Section VI.

VII. Input Drive System

One of the major challenges in the design of a PA is a good input stage to drive the gates of the power transistors. Two of the common problems faced by a poor input circuit design are namely potential for oscillation created by parasitic coupling from the output to the input and input signal degradation due to poor grounding of the output power stage.

The oscillation challenge due to output coupling, is exacerbated if the input connections need to be routed through inside of the DAT in order to save die area. The DAT generates a strong magnetic field inside its loop, hence any metal connection within its perimeter will have a high voltage signal induced on it that can be up to several volts. This unwanted voltage induced in the input connection will add to the input signal becoming a direct RF feedback from PA output, potentially causing instability. A second coupling mechanism, albeit of lesser strength than the first, is capacitive coupling from the DAT through the substrate.

Coupling problems can be more easily addressed if the input signal is made differential and common mode coupling and differential mode coupling from the output are treated independently. An input active balun is used for each of the bands. The balun simultaneously transforms the single-ended input signal to four differential signals for the low-band and three differential signals for the high band.

In order to avoid the differential signal from coupling magnetically to the input through the long connections between these input active baluns and the power transistor gate driver circuits in the DAT, these connections were implemented using a twisted pair configuration, where magnetic coupling in the adjacent lobes is of opposite polarity and gets canceled out. Transformers were used in the drive path to isolate any common mode signal from the inputs to the outputs. The isolated outputs from these transformers further provides common mode isolation among each one of the power transistors pairs in the distributed PA.

In summary, a combination of transformer isolation, shielding and twisted pair transmission line allow maintaining the integrity of the input signal to each one of the power transistor pairs in AX502 avoiding a serious input signal degradation and possible oscillation.

VIII. The End Product

Fig. 10 shows the block diagram of the presented quad-band fully integrated CMOS PA. It comprises two separate PAs on the same die, one for the two lower frequency bands (GSM at 850 MHz and E-GSM at 900 MHz) and another for the high bands (DCS at 1.75 GHz and PCS at 1.9 GHz). Both the low- and the high-band PAs are matched to 50 Ω on-chip at their single-ended input and output terminals. The low-band (LB) PA, which must be able to deliver a peak power of at least $+34.5$ dBm is implemented using a double-concentric quad-core DAT. The high-band (HB) PA, which must have a peak power in excess of $+32.5$ dBm can achieve this peak power with a triple core as opposed to the LB quad-core and is thus realized as a double-concentric triple-core DAT. An on-chip power control monitors the supply current and output power of the PA using on-chip sensors and adjusts its operating point to a tight control of the output power level and supply current.

The fully integrated quad-band CMOS PA is implemented in a standard 130 nm CMOS process and runs off a nominal supply of 3.5 V. It is housed in a 5 mm $\times$ 5 mm $\times$ 0.9 mm MLF package. In the E-GSM band it produces a maximum RF power of $+35$ dBm (3.2 W) with a PAE of 51% that includes all the on-chip losses of the PA, the power control, and other supporting circuitry and package losses. It operates reliably with input power levels ranging from 0.2 dBm to $+8$ dBm . The HB

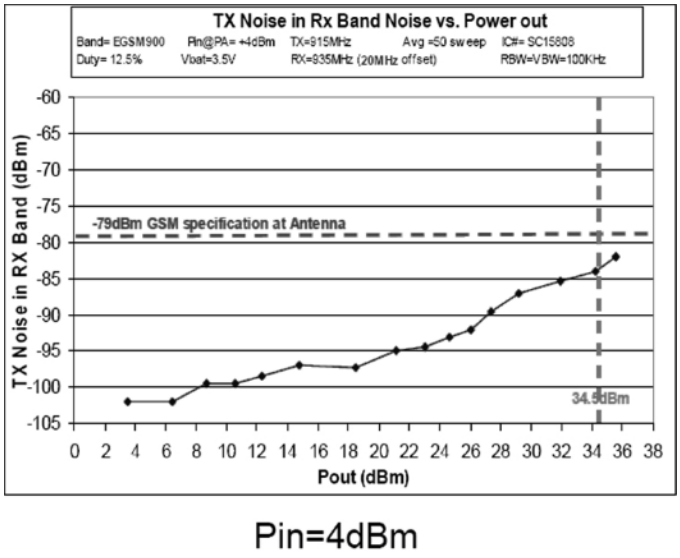


Fig. 14. Transmit noise in the receive band (20 MHz offset) versus Pout.

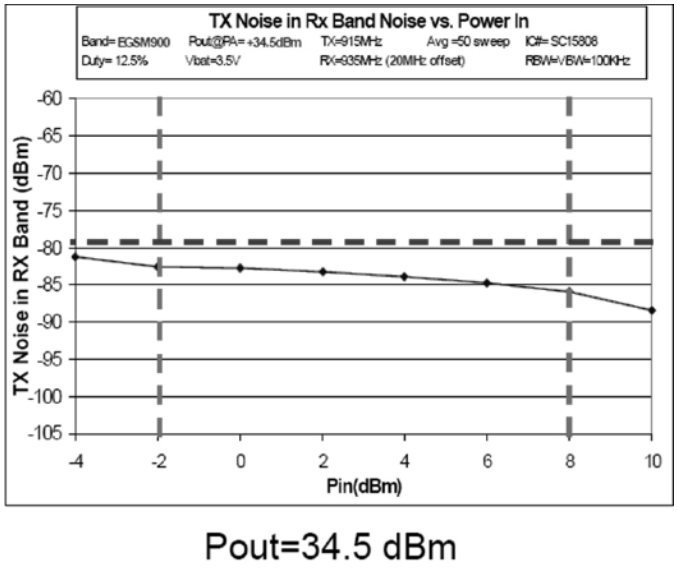


Fig. 15. Transmit noise in the receive band (20 MHz offset) versus Pin.

power amplifier in the PCS band (1850 to 1910 MHz) produces up to 33 dBm (2 W) of power at an overall efficiency of 45%.

The PA is fully compliant with the FCC and GSM requirements, has passed GSM/GPRS full-type-approval (FTA) certification for use in phones, and is in high volume production. The measured output switching transient spectrum is shown in Fig. 11. The power-up and power-down time masks are shown in Fig. 12. The power control accuracy and performance under supply voltage variations is shown in Fig. 13. Its transmit noise in the receive band (at 20 MHz offset) varies between -102 dBm to -84 dBm for output power levels ranging from 0 dBm to $+35$ dBm, which meets the GSM requirements with a margin, as seen in Fig. 14. The transmit noise in the receive band (20 MHz offset) versus Pin is shown in Fig. 15. The $V_{\rm ramp}$ control characteristics are shown in Fig. 16, which provides a very easy way of controlling the power in the system. This characteristic is very repeatable across PVT variations.

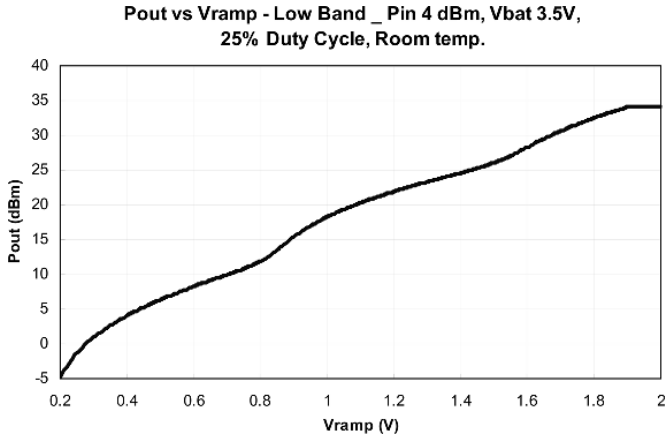


Fig. 16. $V_{\rm ramp}$ control characteristics.

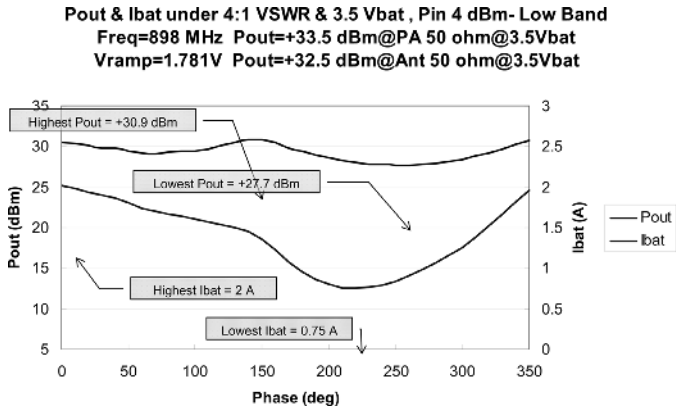


Fig. 17. Output power and the battery current under 4:1 VSWR load mismatch ($\sim$ 6:1 at the antenna).

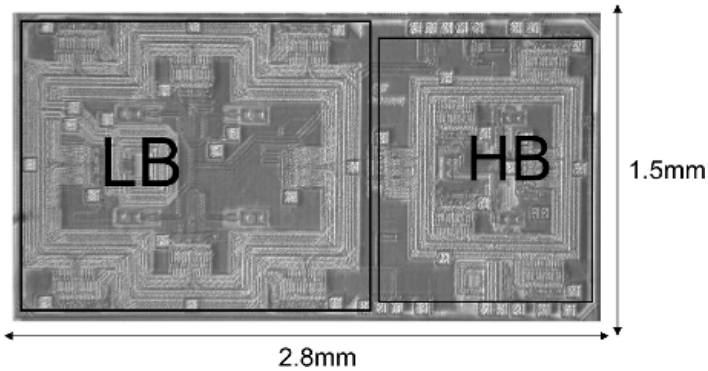


Fig. 18. Die photograph of quad-band fully integrated CMOS PA.

The PA operates for any supply voltage between 2.9 V and 5.5 V, and can withstand a supply of up to 6 V indefinitely under VSWR of greater than 15:1 for all phase angles with no oscillation, breakdown, or degradation. Fig. 17 shows the variations of the output power and battery current with different VSWR angles. This has been demonstrated by more than 2000 hours of failure free operation for hundreds of units at 87.5% duty-cycle under these conditions. The PA has also been tested for more than 2 million device-hours (with ongoing reliability monitoring) without a single failure under nominal operation conditions. Its instantaneous supply breakdown voltage is greater than 9 V. A die micrograph of the chip is shown in Fig. 18. The presented PA proves the viability of CMOS technology for

watt-level fully integrated power generation for wireless applications and serves as an important step toward a single-chip cellphone.

ACKNOWLEDGMENT

The authors would like to thank M. Johnson, V. Boyapati, C. Huynh, J. Kim, F. Carr, F. Jarrar, A. Kral, A. Mellati, J. Mehta, S. Martin, H. Wu, T. Wisler, S. Mezouari, D. Kang, T. Trinh, K. Kong, J. Huynh, D. Qiao, D. Hartman, I. Vitomirov, R. Chen, F. Roux, and M. Damgaard.

REFERENCES

[1] I. J. Bahl *et al.*, "Class-B power MMIC amplifiers with 70 percent power-added efficiency," *IEEE Microw. Theory Tech.*, vol. 37, no. 9, pp. 1315–20, Sep. 1989.
[2] J. Portilla *et al.*, "High power-added efficiency MMIC amplifier for 2.4 GHz wireless communications," *IEEE J. Solid-State Circuits*, vol. 34, no. 1, pp. 120–3, Jan. 1999.
[3] D. Ngo *et al.*, "Low voltage GaAs power amplifiers for personal communications at 1.9 GHz," in *IEEE MTT-S Dig.*, Jun. 1993, vol. 3, pp. 1461–4.
[4] P. DiCarlo *et al.*, "A highly integrated quad-band GSM TX-front-end-module," in *Proc. IEEE Gallium Arsenide Integrated Circuit (GaAs IC) Symp.*, Nov. 2003, pp. 280–283.
[5] K. Asada *et al.*, "A GSM/EDGE dual-mode, 900/1800/1900-MHz triple-band HBT MMIC poweramplifier module," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig.*, Jun. 2002, pp. 245–248.
[6] A. Tombak *et al.*, "A flip-chip silicon IPMOS power amplifier and a DC/DC converter for Gsm850/900/1800/1900 MHz systems," in *IEEE Radio Frequency Integrated Circuits (RFIC) Symp. Dig.*, Jun. 2007, pp. 79.83–79.83.
[7] Y. Tan *et al.*, "A 900-MHz fully integrated SOI power amplifier for single-chip wireless transceiver applications," *IEEE J. Solid-State Circuits*, vol. 35, no. 10, pp. 1481–5, Oct. 2000.
[8] H. Visser *et al.*, "A fully integrated, silicon bipolar RF power amplifier for GSM, operating from a single 3 V supply voltage," in *Proc. 28th European Microwave Conf.*, Oct. 1998, vol. 2, pp. 656–658.
[9] W. Simbürger *et al.*, "A monolithic 2.5 V, 1 W silicon bipolar power amplifier with 55% PAE at 1.9 GHz," in *IEEE MTT-S Dig.*, Jun. 2000, vol. 2, pp. 853–6.
[10] E. Glass *et al.*, "High performance single supply power amplifiers for GSM and DCS applications using true enhancement mode FET technology," in *IEEE MTT-S Dig.*, Jun. 2002, pp. 656–658.
[11] W. Bakalski *et al.*, "A load-insensitive quad-band GSM/EDGE SiGeC-bipolar power amplifier with a highly efficient low-power mode," in *IEEE IEEE Radio and Wireless Symp. Dig.*, Jun. 2008, pp. 203–206.
[12] K. C. Tsai *et al.*, "A 1.9 GHz, 1-W CMOS class-E power amplifier for wireless communications," *IEEE J. Solid-State Circuits*, vol. 34, no. 7, pp. 962–9, Jul. 1999.
[13] C. Yoo *et al.*, "A common-gate switched, 0.9 W class-E power amplifier with 41% PAE in 0.25 mm CMOS," in *Proc. VLSI Symp.*, Jun. 2000, pp. 56–7.
[14] T. C. Kuo *et al.*, "A 1.5 W class-F RF power amplifier in 0.2 mm CMOS technology," in *IEEE ISSCC Dig.*, Feb. 2001, pp. 154–5.
[15] C. Fallesen *et al.*, "A 1 W CMOS power amplifier for GSM-1800 with 55% PAE," in *IEEE Microwave Theory and Technique Symp. Dig.*, May 2001, vol. 2, pp. 911–914.
[16] I. Aoki *et al.*, "A fully 2.4-GHz, 2.2-W, 2-V fully-integrated CMOS circular-geometry active-transformer power amplifier," in *Proc. IEEE Custom Integrated Circuits Conf. (CICC)*, May 2001, pp. 57–60.
[17] I. Aoki *et al.*, "Fully-integrated CMOS power amplifier design using the distributed active transformer architecture," *IEEE J. Solid-State Circuits*, vol. 37, no. 3, pp. 371–83, Mar. 2002.
[18] I. Aoki *et al.*, "Distributed active transformer: A new power combining and impedance transformation techniques," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 1, pp. 316–32, Jan. 2002.
[19] I. Aoki *et al.*, "A fully-integrated 1.8 V, 2.8 W, 1.9 GHz, CMOS power amplifier," in *Proc. IEEE RFIC Symp.*, Jun. 2003, pp. 199–202.
[20] A. Hajimiri, "Fully-integrated RF CMOS power amplifier—A prelude to full radio integration," in *Proc. IEEE RFIC Symp.*, Jun. 2005, pp. 439–442.
[21] N. Wongkomet *et al.*, "A +31.5 dBm CMOS RF doherty power amplifier for wireless communications," *IEEE J. Solid-State Circuits*, vol. 41, pp. 2852–2859, Dec. 2006.
[22] J. Jang *et al.*, "A CMOS RF power amplifier using an off-chip transmission line transformer with 62% PAE," *IEEE Microw. Wireless Compon. Lett.*, vol. 17, pp. 385–387, May 2007.
[23] K. H. An *et al.*, "Power-combining transformer techniques for fully-integrated CMOS power amplifiers," *IEEE J. Solid-State Circuits*, vol. 43, no. 5, pp. 1064–1075, May 2008.
[24] G. Liu *et al.*, "Fully-integrated CMOS power amplifier with efficiency enhancement at power back-off," *IEEE J. Solid-State Circuits*, vol. 43, no. 3, pp. 600–609, Mar. 2008.
[25] R. Gupta *et al.*, "Design and optimization of CMOS RF power amplifiers," *IEEE J. Solid-State Circuits*, vol. 36, no. 2, pp. 166–75, Feb. 2001.
[26] Y. J. E. Chen *et al.*, "RF power amplifier integration in CMOS technology," in *IEEE MTT-S Dig.*, Jun. 2000, vol. 1, pp. 545–548.
[27] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, Apr. 1965.
[28] Y. Taur *et al.*, "CMOS scaling into the nanometer regime," *Proc. IEEE*, vol. 85, no. 4, pp. 486–504, Apr. 1997.
[29] D. Goldhaber-Gordon *et al.*, "Overview of nanoelectronic devices," *Proc. IEEE*, vol. 85, no. 4, pp. 521–540, Apr. 1997.
[30] R. Moazzami *et al.*, "Projecting gate oxide reliability and optimizing reliability screens," *IEEE Trans. Electron Devices*, vol. 37, pp. 1643–1650, Jun. 1997.
[31] K. Sakakibara *et al.*, "Field and temperature acceleration model for time-dependent dielectric breakdown," *IEEE Trans. Electron Devices*, vol. 44, pp. 986–992, Jun. 1997.
[32] M. Kimura, "Field and temperature acceleration model for time-dependent dielectric breakdown," *IEEE Trans. Electron Devices*, vol. 46, pp. 220–229, Jan. 1999.
[33] Sonnet User Manuals, Ver. 11. Sonnet Software Inc., 2008.

Ichiro Aoki (S'98–M'01) was born in Kyoto, Japan, in 1965. He received the B.S.E.E. degree from Universidade Estadual de Campinas, SP, Brazil, in 1987, and the M.S.E.E. and Ph.D. degrees from the California Institute of Technology (Caltech), Pasadena, in 1999 and 2001, respectively.

He is currently with AyDeeKay LLC, a fabless semiconductor startup, which he co-founded in 2007. He co-founded Axiom Microdevices, Inc., a fabless semiconductor startup with over 40 million fully integrated CMOS power amplifiers shipped to GSM/GPRS cell phone market to date, and from 2002 to 2007 worked in its engineering and management positions including acting CEO. He cofounded and managed as co-CEO PST Eletrônica S/A, Brazil, a car electronic components manufacturing company from 1988 to 1998. At the time of his departure, PST had 300 employees and has since grown to over 1200 employees and over US$200 million revenue. His current research interests include high-frequency silicon RF analog integrated circuits for wireless communications and low-power mixed-signal circuits.

Mr. Aoki received the Schlumberger Fellowship from 1998 to 1999 and the Walker von Brimer Foundation Award in 2000 at Caltech. He currently serves on the Board of the Directors of Axiom Microdevices, Inc.

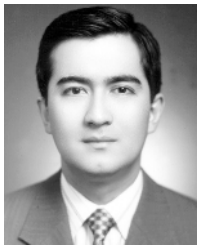
Scott Kee (M'03) received the B.E.E. degree in electrical engineering from the University of Delaware, Newark, in 1998, and the M.S.E.E. and Ph.D. degrees in electrical engineering from the California Institute of Technology (Caltech), Pasadena, in 2002.

He was a founding member of Axiom Microdevices, a fabless semiconductor startup producing CMOS cellular power amplifiers, where he worked from 2002 to 2007 as CTO. He is currently with AyDeeKay, a fabless semiconductor startup, which he co-founded in 2007.

Rahul Magoon (M'97) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Bombay, in 1995, and the M.S. degree in electrical engineering from the University of California, Santa Barbara, in 1997.

From 1997 to 2002, he was with Rockwell Semiconductor Systems, Conexant Systems, and Skyworks Solutions where he was responsible for the GSM RF transceiver design team. He is currently Director of Engineering at Axiom Microdevices, Irvine, CA, and has been with the venture-backed start-up company since its inception in 2002 as one of the initial founding team. At Axiom, his focus has been on the development of monolithic watt level power amplifier and integrated radio products in commercial CMOS processes for the cellular market.

Mr. Magoon holds a number of U.S. patents and has given invited presentations at various academic and industrial forums, in addition to having published several well-received IEEE journal and conference papers.



Roberto Aparicio (S'01–M'04) was born in Puebla, Mexico, in 1975. He received the B.S. degree with honors (*cum laude*) in electronics from Universidad Autonoma de Puebla, Mexico, in 1999, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, in 2001 and 2004, respectively.

He was an Integrated Circuit Design Engineer with Centro Tecnologico para Informatica (CTI), Campinas, Brazil, where he worked on the design of very large CMOS gate arrays in 1998. In 2002, he was a member of the Technical Staff of the Mixed-Signal Communications IC Design group at the IBM T. J. Watson Research Center, Yorktown Heights, NY, where he investigated PLL architectures for optical communications. Since 2004, he has been with Axiom Microdevices Inc., Irvine, CA, where he is Principal PA Architect.

Dr. Aparicio ranked first at the Universidad Autonoma de Puebla class '98 and earned the Phoenix Medal for outstanding academic achievements. He was a Fulbright Scholarship and an IBM Research Fellowship recipient at the California Institute of Technology from 1999 to 2001 and 2002 to 2003, respectively, and received the Walker von Brimer Foundation Outstanding Accomplishment Award and the Analog Devices Inc. Outstanding Student Designer Award, both in 2001.

Florian Bohn (S'07) received the B.S. degree with honors in electrical engineering from the California Institute of Technology, Pasadena, and the M.S. degree in electrical and computer engineering from the University of California, Santa Barbara, in 2001 and 2003, respectively.

From 2003 to 2005, he worked as an RF IC design engineer with Axiom Microdevices Inc., Irvine, CA. At Axiom, he focused on the design and testing of active and passive circuits for GSM/GPRS CMOS power amplifiers. He is currently pursuing the Ph.D. degree in electrical engineering at the California Institute of Technology, where he has worked on frequency synthesizers. His research interests lie in the area of integrated micro- and millimeter wave transceiver circuits and systems.

Mr. Bohn has received a Conexant Scholarship and an Analog Devices Outstanding Student Designer Award.

Jeff Zachan (S'98–M'01) received the B.S. and M.S. degrees in electrical engineering from the University of California, Santa Barbara, in 1998 and 1999, respectively.

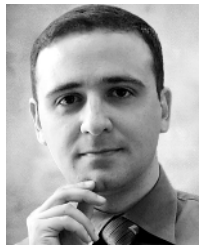
From 2000 to 2004, he was with Skyworks Solutions (formerly Conexant Systems), Newport Beach, CA, where he was a Senior Staff Engineer focused on frequency synthesis and VCO design for integrated GSM transceivers. Since 2004, he has been with Axiom Microdevices where he is a Principal Design Engineer leading the development of transmit architectures and frequency synthesizers for fully integrated CMOS transceivers, including fully integrated power amplifiers, for cellular standards.

Geoff Hatcher received the B.S. and M.S. degrees in electrical engineering from the University of California at Los Angeles in 1999 and 2001, respectively.

From 2000 to 2004, he was with Skyworks Solutions Inc. (formerly Conexant Systems), Newport Beach, CA, where he was a Senior Staff Design Engineer in the RFIC Design Group. He worked on direct conversion transceivers for the GSM/EDGE standard. Since 2004, he has been with Axiom Microdevices, Irvine, CA. He is a Principal Staff Engineer working on transceivers for cellular standards. His interests are in fully integrated CMOS solutions of transceivers and power amplifiers.

Donald McClymont received the Masters degree in engineering from the University of Glasgow, Scotland, U.K.

He is Vice President of Marketing for Axiom Microdevices, responsible for architecting the company's product roadmap. He brings more than 15 years of experience in marketing and engineering for the semiconductor industry to his position. Prior to joining Axiom, he served as Product Line Director for Skyworks Solutions, Inc. and at Conexant Systems, Inc., managing the company's GSM direct conversion transceiver business. During his career, he also has held marketing and engineering positions at Fujitsu Microelectronics GmbH, Thesys GmbH, and Wolfson Microelectronics plc.



Ali Hajimiri (M'99) received the B.S. degree in electronics engineering from the Sharif University of Technology, Tehran, Iran, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1996 and 1998, respectively.

He was a Design Engineer with Philips Semiconductors, where he worked on a BiCMOS chipset for GSM and cellular units from 1993 to 1994. In 1995, he was with Sun Microsystems, where he worked on the UltraSPARC microprocessor's cache RAM design methodology. During the summer of 1997, he was with Lucent Technologies (Bell Labs), Murray Hill, NJ, where he investigated low-phase-noise integrated oscillators. In 1998, he joined the Faculty of the California Institute of Technology, Pasadena, where he is a Professor of electrical engineering and the Director of Microelectronics Laboratory. His research interests are high-speed and RF integrated circuits.

Dr. Hajimiri is the author of *The Design of Low Noise Oscillators* (Springer, 1999) and has authored and coauthored more than 100 refereed journal and conference technical articles. He holds more than 30 U.S. and European patents. He is a member of the Technical Program Committee of the IEEE International Solid-State Circuits Conference (ISSCC). He has also served as an Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS, an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS PART II, a member of the Technical Program Committees of the International Conference on Computer Aided Design (ICCAD), Guest Editor of the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, and the Guest Editorial Board of *Transactions of Institute of Electronics, Information and Communication Engineers of Japan (IEICE)*.

Dr. Hajimiri was selected to the top 100 innovators (TR100) list in 2004 and is a Fellow of Okawa Foundation. He is a Distinguished Lecturer of the IEEE Solid-State and Microwave Societies. He is the recipient of Caltech's Graduate Students Council Teaching and Mentoring award as well as Associated Students of Caltech Undergraduate Excellence in Teaching Award. He was the Gold medal winner of the National Physics Competition and the Bronze Medal winner of the 21st International Physics Olympiad, Groningen, Netherlands. He was a co-recipient of the IEEE JOURNAL OF SOLID-STATE CIRCUITS Best Paper Award of 2004, the ISSCC Jack Kilby Outstanding Paper Award, two times co-recipient of CICC's best paper awards, and a three times winner of the IBM faculty partnership award as well as National Science Foundation CAREER award. He is a cofounder of Axiom Microdevices Inc.