Fully Integrated RF CMOS Power Amplifiers - A Prelude to Full Radio Integration

INVITED PAPER

Ali Hajimiri

Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

Abstract

A fully integrated radio on a chip offers innumerable advantages and several challenges. In this article, we will discuss some of these opportunities and challenges for a fully-integrated CMOS PA and see how the distributed active transformer (DAT) technology can over come them.

Introduction

Full integration of a complete radio system on a single CMOS die in a standard package with no external components has been considered the Holy Grail of the RF CMOS quest. This level of integration is shown symbolically in Fig. 1. This level of integration will enable numerous new opportunities for versatile low-cost wireless solutions for a plethora of applications with no need for individual fine tuning of the building blocks.

The power amplifier (PA) has been one of the last bastions of exotic technologies in wireless systems, particularly in cellular handsets. The combination of the high power levels and high operation frequencies often necessitate accurately controlled passive impedance transformation networks to match the load to a lower impedance level. Currently, PAs are implemented using complex module technologies, where a chip containing power transistors is mounted on an organic or ceramic substrate that carries or contains the output matching passives, as depicted in Fig. 2a. These PA modules offer matched input and output (usually to 50Ω) to streamline their usage by the handset original equipment manufacturers (OEMs). Today, the power transistor is often implemented using compound semiconductor processes, mostly due to their higher breakdown voltages. The power control components are usually included in the module either as external components on the module or as a part of the same chip containing the power transistors. Despite the cost disadvantages and lack of prospects for integration with the rest of the radio, off-chip passives and compound semiconductors have been chosen for lack of any other viable alternative. Compound semiconductor (e.g., GaAs HBT) and other specialty (e.g., LDMOS) power transistors are used primarily due to their higher breakdown voltages. As we will see later in this article, classical power amplifier design methodologies mandate transistors with high breakdown voltages to obtain the necessary power levels in cellular handsets. For example, GSM/GPRS/EDGE handsets must provide up to three watts of peak power.

The compound semiconductor PAs generally suffer from a performance hit with a lower power supply and have therefore resisted lowering of the battery voltages in cell phones. This has resulted in a constantly growing, and generally unsustainable, disparity between the power supplies used for the PA and that of the rest of the components of a cell phone that benefit from CMOS scaling (e.g., digital baseband and radio transceivers), as shown qualitatively in Fig. 3. A fully-integrated CMOS PA can serve as the key to break this deadlock without a substantial loss in performance.

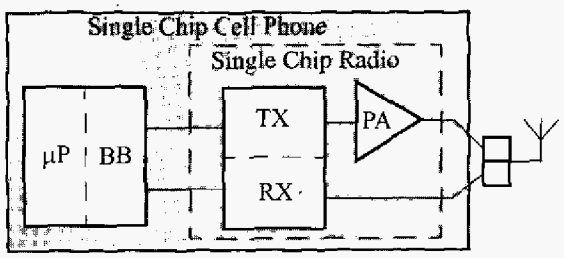

Figure 1: Handset integration prospects.

While there are challenges on the active device side, perhaps the greatest challenge is the successful implementation of an on-chip fully-integrated output impedance matching. Today, this problem is often solved using external components on the printed circuit board (PCB) or the module. In the absence of integrated matching circuits with acceptable power loss, external passive components, such as printed metal trace inductors and chip capacitors, are implemented using module technology. Unfortunately, they must be low loss and maintain very tight tolerance control due to the sensitivity of the classical power amplifier topologies to output matching network component values. This necessitates the use of either expensive substrate technologies, such as ceramic substrate and/or mounting additional high-quality passive components on the base PCB. Both will result in additional cost and size limitations but most importantly pose serious hurdles to full integration of the PA and transceiver to form a complete CMOS radio on a chip. Even in the absence of cost disadvantages of compound power transistors, such module-based approaches suffer from higher cost and the absence of a direct path to integration with the rest of the transceiver. Although these two major disadvantages, namely, the need for passives in the module and high voltage specialty transistors were perceived to be fundamental, they can be completely circumvented via techniques such as the distributed active transformer (DAT) approach, which makes it possible to have a fully integrated CMOS PA in a standard lead-frame package, as shown schematically in Fig. 2b. Such technology can naturally lead to an economical CMOS radio on a chip where the entire radio including the PA can be integrated onto a single CMOS chip in a standard package.

There are two major challenges for watt-level fully-integrated RF power amplifiers in CMOS with no external components: the high ohmic and substrate energy loss of the on-chip passive components (mainly inductors) and the low breakdown voltage of the active devices. It should be noted that it is the full integration of the output matching where the big prize and the major challenge lies. As a matter of fact, watt level CMOS PAs with external passive components such as printed circuit metal traces, external baluns, and/or bonding wires have been reported in earlier works [1][2][3]

0-7803-8983-2/05/$20.00 © 2005 IEEE

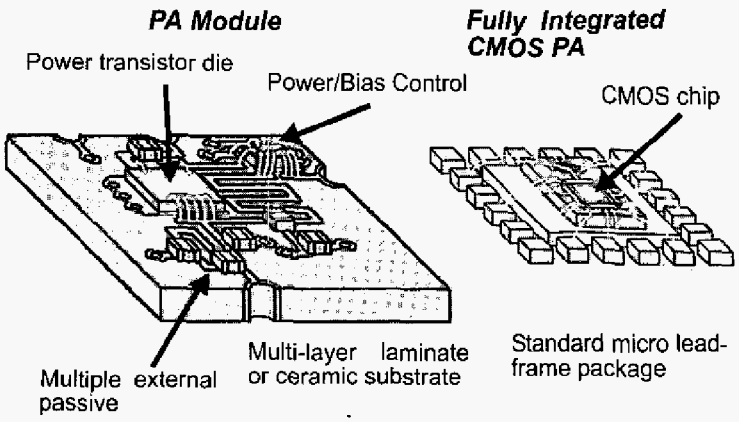

Figure 2: Evolution of PA technology: a) PA modules b) fully integrated CMOS PA

dating back to 1990s. One the other hand, to this date, with the exception of the DAT concept in [6]-[8], the highest power levels achieved using a CMOS PA with integrated output matching circuitry are on the order of 100mW [4][5].

In addition to CMOS implementations, Si-Bipolar transistors with multiple external components such as bonding wires and external baluns have been used to produce output power levels in excess of 1W [9]. Alternative technologies with higher breakdown voltage devices or higher substrate resistivity have been used to increase the efficiency and output power of integrated amplifiers. In particular, LDMOS transistors with a breakdown voltage of 20V [10] and GaAs MMICs (monolithic microwave integrated circuit) with semi-insulating substrate [11][12][13] have been used to integrate power amplifiers.

This article will show how one can overcome the challenges associated with the realization of a fully integrated CMOS PA at RF and Microwave frequencies by moving away from standard approaches without having to use module technologies or exotic substrates. Thus, creating for the first time a direct path to the complete radio-on-a-chip solution.

Design Challenges for a Fully-Integrated CMOS PA

In a traditional power amplifier design a resonant LC impedance transformation network is used to limit the voltage swing seen by the active device driving the output. The matching network plays an important role by lowering the impedance level seen and hence lowering transistor voltage swing and increasing its current. The impedance transformation ratio, r, is the ratio of the load resistance (usually 50Ω) to its transformed impedance at input port (transistor side) of the matching circuit. The voltage swing of the active device in conjunction with the required output power determines the input impedance.

To illustrate the significance of impedance transformation, consider the following example: Delivering 3W (~35dBm) of power to a 50Ω load corresponds to a peak-to-peak voltage swing in excess of 35V on the load. If we were to limit the peak-to-peak voltage swing across the drain of the transistor to 3.6V, with no impedance transformation network in place we can only deliver 32mW (~15dBm) to a 50Ω load. Based on this observation, we can define the power enhancement

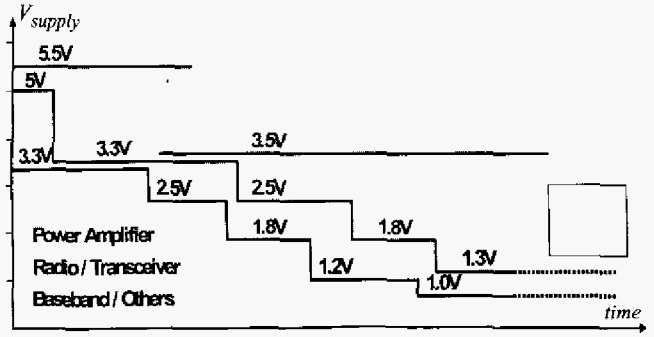

Figure 3. Supply voltage trends in cellular phone building blocks.

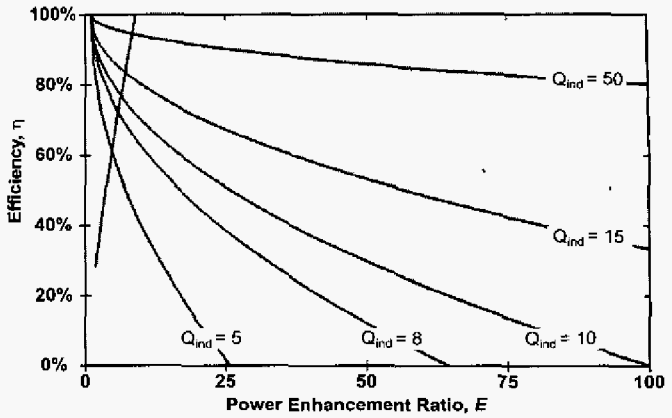

Figure 4. Passive efficiency vs. PER for different inductor Q in a single-section resonant impedance transformation network.

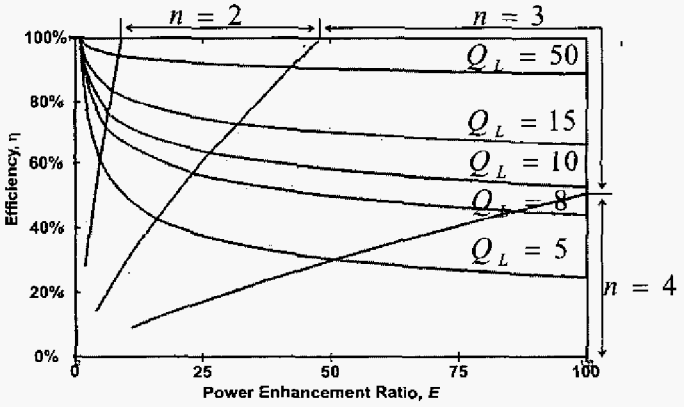

Figure 5. Passive efficiency vs. PER for different inductor Q in a resonant impedance transformation network with the optimum number of stages.

ratio (PER) as the ratio of the RF power delivered to the load with a transformation network in place (e.g., 3W) to the power delivered to the load for the same sinusoidal input voltage source when it drives the load directly (e.g., 32mW). Unlike impedance transformation ratio, the PER accounts for the loss in the passive impedance transformation network and is thus particularly important for lossy on-chip passive components in silicon technology.

It is possible to calculate the passive network efficiency, η_p, for a necessary PER and a given available inductor Q_{ind}. Fig. 4 shows plots of η vs. PER for several different Q_{ind} for a single section LC matching network. In the example given above, the required

PER is more than 90. So in this case, even with a Q_{ind} of 10, the maximum theoretically achievable passive efficiency of a single LC section match is less than 4%. This does not include the additional energy losses in the active device, external connections, or driving circuitry which will degrade the efficiency even further. We can also see in the Fig. 4 that for a given inductor quality factor, Q_{ind}, there is a maximum achievable PER beyond which the required transformation becomes unattainable.

Multi-section transformation networks can be used as a means of lowering loss for high PER compared to a single section networks. The analysis can be extended to multi-section transformation networks [7]. For example, one can see from Fig. 5 that with a PER of 90 and an inductor quality factor, Q_{ind}, of 10, the best matching network will have 3 LC-sections and will have a theoretical maximum passive efficiency of around 55%. Again, this figure does not include any loss in the active device, the dc feeds, or the external connections. Additionally, a multi-section transformation network requires more complex layouts with inductors and capacitors having a very broad range of reactances compared to a single section. This results in a lower overall quality factors, Q, for the network.

The dependence of η_p on PER and Q_{ind} has important implications regarding the necessary reactance, transformation efficiency, and the PER. In particular, the inductor reactance necessary for this type of matching network with a single-section decreases rapidly as the desired PER increases, making it more susceptible to parasitics and loss. More importantly, the passive efficiency, η_p, also decreases quickly with higher PER, as can be seen in Figures 4 and 5. In a multi-section approach, the loss is improved significantly compared to the single-section network, but still increases with higher PER. Also for given number of stages, a higher PER usually corresponds to a smaller bandwidth in this kind of design. PA designers have long understood this trade-off by intuition and experience. This basic trade-off between passive efficiency and PER is has been used as one of the main justifications for transistors with higher breakdown voltages that can be used to reduce the required PER and hence improve the passive efficiency of the matching network for a given range of quality factors.

Using conventional PA topologies, the low Q passives currently available on chip fundamentally limit achievable output power of a fully-integrated design. No amount of complexity in a single- or multi-stage LC transformation network can overcome this, making it necessary to pursue alternative approaches hence the introduction of distributed active transformers (DAT).

Distributed Active Transformer (DAT)

"Divide and conquer" has been the underlying principle used to solve many engineering problems. Integrated circuit design is a perfect example of this process, where the chip is initially defined at the application level, then described using system level terms, leading to an architecture using a number of sub-blocks. The integrated circuit design process is then

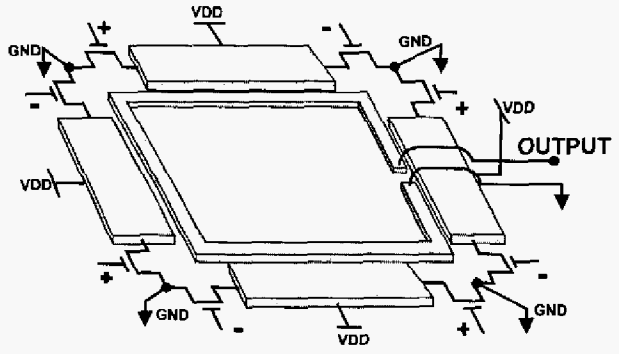

Figure 6. The basic structure of a square DAT.

divided further by defining the specifications for circuit building blocks and their interfaces that together form the system. The circuit designer works with the specifications at a lower level of abstraction dealing with transistors and passive components whose models have been extracted from the measurements, device simulations, or analytical calculations based on the underlying physical principles of semiconductor physics and electrodynamics.

While this approach has been quite successful in mainstream areas of integrated circuit design, it can lead to sub-optimal designs. Further advancements beyond these local optima can be achieved by looking at the problem across multiple levels of abstraction to find solutions not easily seen when one confines the design space to one level (e.g., transistor level circuit design).

Distributed active transformer (DAT) is one such multi-level approach allowing a more integral co-design of the building blocks at the circuit, device, and electromagnetism levels. Unlike most conventional circuits, it relies on multiple parallel signal paths operating in harmony to achieve the design objective applying the divide-and-conquer mentality at a different level. The DAT defies basic trade-offs discussed earlier between the passive efficiency and PER using lossy passive components in classical PA design by taking advantage of multiple parallel signal paths. This multiple signal path feature results in strong electromagnetic couplings between circuit components across multiple levels of abstraction. The strong coupling makes it necessary to perform the analysis and the design of the DAT across different levels. Due to the larger solution space, the design is much closer to a global optimum and hence outperforms conventional approaches. It is obvious that such multi-level approaches are more challenging due to the larger number of disciplines that need to be mastered.

The DAT's primary circuit consists of multiple distributed push-pull circuits in a polygon geometry, as seen in the square geometry of Fig. 6. Each side of the square is a single push-pull amplifier consisting of a load transmission line, a differential driver, and input matching circuitry. The push-pull structure creates a virtual ac ground at the supply node that makes it unnecessary to use choke inductors and large on-chip bypass capacitors at supply. This particular positioning of the push-pull amplifiers makes it possible to use a

straight wide metal lines as the drain inductor. These inductors provide natural low resistance paths for the dc current to flow from the supply to the drain of the transistors, as shown in Fig. 6. This results in a higher quality factors for on-chip inductors due to the smaller required length of each primary side inductor. The virtual grounds formed in each corner make the DAT power amplifier insensitive to the exact length of the package lead inductance and wirebonds.

A power distribution network in the center of the DAT (not shown in Fig. 6) provides the differential input to the transistor pairs in the corner. By driving the two adjacent transistors of two different push-pull amplifiers in opposite phases, we can create a virtual ac ground in each corner of the square. This is an essential feature for realizing a lumped inductor characteristic using a long and wide piece of metal whose two ends are at two different physical locations.

A metal loop inside the square is used to harness this RF magnetic flux and acts as the transformer secondary. This is where multiple signal paths converge to generate the large required voltage swing across the 50Ω load by adding the voltages on the secondary. The strong electromagnetic coupling between multiple signal paths in a DAT necessitates an analysis and design approach spanning architecture, circuits, device physics, and electromagnetics.

The $n/2$ push-pull amplifiers, (four in this example), conduct identical synchronized ac currents at the fundamental, inducing corresponding ac magnetic fields in this secondary loop. The internal metal loop generates a voltage between its terminals equivalent to the sum of the differential voltages of the $n/2$ push-pull amplifiers. The DAT architecture results in a simultaneous $1{:}n$ impedance transformation and n transistor series power combining, circumventing the basic tradeoff between the efficiency and impedance transformation.

The distributed nature of the DAT structure reduces the sensitivity of the power amplifier's efficiency to the substrate power losses while providing a large overall output power using low-breakdown-voltage MOS transistors. The transformer based matching used in DAT is inherently more broadband and has lower loss than the conventional LC approach.

Its distributed nature also makes it more robust to parameter variations and facilitates the heat sinking of the power amplifier, as there are multiple small sources of heat as opposed to the conventional approach where a large power transistor is responsible for all of the output power generation. The thermal issues are further improved due to the superior thermal conductivity of silicon substrate compare to compound semiconductors (roughly 4 times better than GaAs) and the fact that the die is directly mounted on a copper heat slug.

The DAT technology is used to realize a fully-integrated quad-band (GSM/DCS/PCS) CMOS power amplifier by Axiom Microdevices Inc. [14]. It comes in a 5mm x 5mm micro-lead-frame (MLF) package with high-accuracy on-chip closed loop power control and achieves integrated 50Ω input/output matching circuits without the use of module or ceramic substrate technology. It requires no external components and supports GPRS class 12, allowing a power supply voltage range of 2.9-5.5V, direct from cell-phone battery pack. It offers the handset designer an easy, zero-component interface to transceiver and transmit-receive switch module. The device is fully protected against load mismatch, over-voltage, over-current, and over-temperature events. The functional block diagram of this part is shown in Fig. 7.

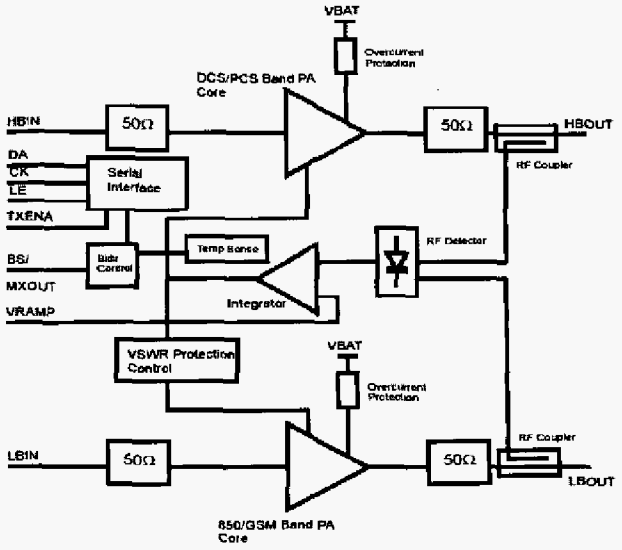

Figure 7. Block diagram of the CMOS fully-integrated power amplifier.

Acknowledgments:

The author would like to thank A. Aoki, S. Kee, and D. Rutledge as well as many employees of Axiom Microdevices Inc. for their inputs and contributions.

References:

1. K. C. Tsai, *et al.*, "A 1.9 GHz, 1-W CMOS Class-E Power Amplifier for Wireless Communications," *IEEE JSSC*, vol. 34, no. 7, pp. 962-9, July 1999.
2. C. Yoo, *et al.*, "A Common-Gate Switched, 0.9W Class-E Power Amplifier with 41% PAE in 0.25μm CMOS," *VLSI Symp.*, pp. 56-7, June 2000.
3. T. C. Kuo, *et al.*, "A 1.5W Class-F RF Power Amplifier in 0.2mm CMOS Technology," *IEEE ISSCC*, pp. 154-5, Feb. 2001.
4. R. Gupta, *et al.*, "Design and Optimization of CMOS RF Power Amplifiers," *IEEE JSSC*, vol. 36, no. 2, pp. 166-75, Feb. 2001.
5. Y. J. E. Chen, *et al.*, "RF power Amplifier Integration in CMOS Technology," *IEEE MTT-S*, vol. 1, pp. 545-548, June 2000.
6. I. Aoki, *et al.*, "A 2.4-GHz, 2.2-W, 2-V fully-Integrated CMOS Circular-Geometry Active-Transformer Power Amplifier," *IEEE CICC*, May 2001.
7. I. Aoki, *et al.*, "DAT- A New Power Combining and Impedance Transformation Technique," *IEEE Proc. MTT*, vol. 50, no. 1, Jan. 2002.
8. I. Aoki, *et al.*, "A Fully-Integrated 1.8-V, 2.8-W, 1.9-GHz CMOS Power Amplifier," *IEEE RFIC*, pp. 199-203, June 2003.
9. W. Simbürger, *et al.*, "A Monolithic 2.5V, 1W Silicon Bipolar Power Amplifier with 55% PAE at 1.9GHz," *IEEE MTT-S*, vol. 2, pp. 853-6, June 2000.
10. Y. Tan, *et al.*, "A 900-MHz Fully Integrated SOI Power Amplifier for Single-Chip Wireless Transceiver Applications," *IEEE JSSC*, vol. 35, no. 10, pp. 1481-5, Oct. 2000.
11. J. Portilla, *et al.*, "High Power-Added Efficiency MMIC Amplifier for 2.4 GHz Wireless Communications," IEEE JSSC, vol. 34, no. 1, pp. 120-3, Jan. 1999.
12. I. J. Bahl, *et al.*, "Class-B Power MMIC Amplifiers with 70 Percent Power-Added Efficiency," *IEEE Proc. MTT*, vol. 37, no. 9, pp. 1315-20, Sept. 1989.
13. D. Ngo, *et al.*, "Low Voltage GaAs Power Amplifiers for Personal Communications at 1.9GHz," *IEEE MTT-S*, vol. 3, pp. 1461-4, June 1993.
14. Axiom Microdevices Inc. web page: www.axiom-micro.com