A 1 GHz CMOS RF Front-End IC for a Direct-Conversion Wireless Receiver

Ahmadreza Rofougaran, James Y.-C. Chang, Maryam Rofougaran, and Asad A. Abidi, *Fellow, IEEE*

***Abstract*—An integrated low-noise amplifier and downconversion mixer operating at 1 GHz has been fabricated for the first time in 1 μm CMOS. The overall conversion gain is almost 20 dB, the double-sideband noise figure is 3.2 dB, the IIP3 is +8 dBm, and the circuit takes 9 mA from a 3 V supply. Circuit design methods which exploit the features of CMOS well suited to these functions are in large part responsible for this performance. The front-end is also characterized in several other ways relevant to direct-conversion receivers.**

I. Introduction

MOTIVATED by the growing needs for low-power and low-cost wireless transceivers, mainstream IC technologies are competing to integrate more RF functions onto a single chip. Bipolar circuits dominate integration at 1 GHz today, followed by GaAs IC's. As recent results demonstrate, CMOS too is a viable contender in this frequency range [1]–[6]. If CMOS is shown to perform in certain important respects as well as circuits in other established technologies, and it successfully merges analog and digital blocks, its use at RF may become as compelling as it is in baseband circuits.

To date, most research on CMOS RF circuits shows the feasibility of elementary RF building blocks, such as standalone tuned amplifiers [1], mixer IC's [2], [5], and oscillators [7]. The next development step calls for the integration of these building blocks into subcells, comparable in function to currently available small-scale RF IC's in bipolar or GaAs technology. The most common example of such an IC is an RF low-noise amplifier (LNA) combined with a downconversion mixer, often labeled the *front-end* for an RF receiver. Integrated front-ends are widely used because they combine all the RF signal processing on one chip, often requiring only a small overhead of off-chip components, and they produce an amplified signal translated down to a conveniently low intermediate frequency (IF) at the output. Thereafter, it is relatively simple to implement IF and baseband circuits for the rest of the receiver. The work reported here is the first implementation of a 1 GHz front-end in CMOS.

The front-end of a wireless receiver must meet several exacting specifications. First is *sensitivity*. The input noise of the front-end must be sufficiently low to enable it to detect weak input signals. The front-end gain must be high enough to overcome the noise contributions of later circuits, which may otherwise degrade the receiver sensitivity. Second, a front-end with a wide *input dynamic range* can tolerate large undesired signals nearby in frequency to a weak desired signal, which may otherwise, through intermodulation distortion, create energy at frequencies overlapping the desired channel. Third, the *RF input impedance* of the front-end must be a good match to the antenna characteristic impedance over the frequency band of interest.

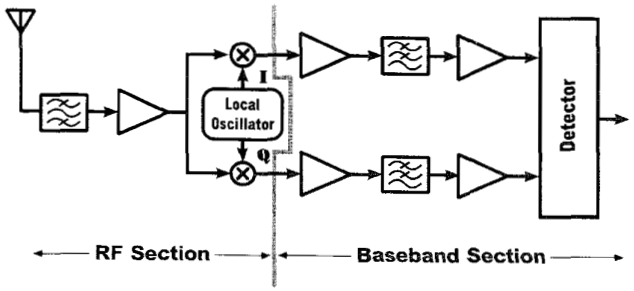

Fig. 1. A direct-conversion receiver suitable for FSK modulation.

The LNA and mixer together determine the performance of the front-end. For instance, although a large LNA gain is desirable as mentioned above, too large a gain may overload the mixer and compromise dynamic range. On the other hand, the gain must be large enough to overcome the fundamentally higher mixer noise. It is also desirable to connect the LNA in some simple way to the mixer input, without a power-hungry RF buffer. The front-end design is influenced by its intended use, as discussed below. Therefore, good performance is only obtained through careful *co-design* of the front-end building blocks.

This front-end is intended for a direct-conversion, or zero-IF, frequency-shift keying (FSK) receiver [8], which simplifies how the blocks are connected together (Fig. 1). In a superheterodyne receiver, a passive filter—usually a second preselect filter of the type connected to the antenna—follows the LNA [9] to suppress the amplifier noise in the image of the RF input band, where there is no signal. Without this filter, the downconverted signal must contend with downconverted noise from both the signal-bearing and the idle sidebands. In a direct-conversion receiver, on the other hand, the local oscillator (LO) is centered in the desired channel, so useful signal energy, and noise, occupy *both* upper and lower sidebands. As there is now no idle sideband to be filtered, the LNA is directly connected to the downconversion mixer. Therefore, when a

Manuscript received January 23, 1996; revised March 5, 1996. This work was supported by the U.S. Advanced Research Projects Agency, Rockwell International, Texas Instruments, Harris Semiconductor, Advanced Micro Devices, Hewlett-Packard Company, and the State of California MICRO Program.

The authors are with the Integrated Circuits & Systems Laboratory, Electrical Engineering Department, University of California, Los Angeles, CA 90095-1594 USA.

Publisher Item Identifier S 0018-9200(96)04472-1.

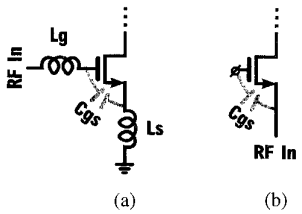

Fig. 2. Candidate FET LNA input stages. (a) Common-source stage, with lossless matching network. (b) Common-gate stage.

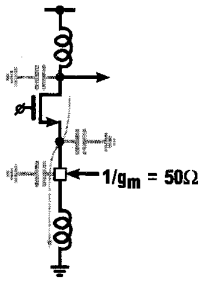

Fig. 3. Low-noise amplifier with tuned load and an off-chip tuning inductor at the input port.

modulation such as FSK permits use of direct-conversion [10], the receiver may be integrated with a need for very few off-chip components. There are several advantages to this, mainly small physical volume, less wasted power in buffering high-frequency signals off-chip, and lower assembly costs. However, the receiver requires a vector baseband signal path, consisting of two branches downconverted in quadrature to prevent the signal-bearing image sidebands from aliasing on one another.

II. Circuit Design

A. Low-Noise Amplifier

The LNA must simultaneously attain high RF gain, low input noise, and a good input match to 50 Ω. These requirements are often interdependent in a simple circuit, and may require iterative design for all to be fulfilled. The following discussion on a CMOS LNA design covers input impedance matching first, then input noise, and finally voltage gain.

The gate of a FET fabricated in 1-μm technology is capacitive to frequencies beyond 1 GHz.[1] However, a lossless *matching network*, consisting only of inductors and capacitors can transform the FET input into a pure resistance over some frequency band of interest. The most common matching network for FET's consists of a series feedback inductor, L_s, in the FET source, and another inductor, L_g, in series with the gate to tune out the capacitance $C_{\rm gs}$, resulting in an input resistance $g_m L_s/C_{\rm gs}$ where g_m is the FET transconductance [Fig. 2(a)] [11]. This method is preferred over resistor feedback because the matching network introduces no noise of its own. However, loss in practical inductors L_g and L_s will tend to degrade noise figure. With sufficiently good inductors, though, a noise figure well below 3 dB may be obtained with this technique, ultimately limited by such transistor imperfections as nonzero gate resistance.

When a noise figure of 3 dB is acceptable, as it is in our receiver, it is simpler to regulate the input impedance with a common-gate input stage [12] [Fig. 2(b)]. For the sake of discussion, first suppose that the load resistance at the drain is much less than the FET $r_{\rm ds}$, and that $g_m r_{\rm ds} \gtrsim 10$. Then the input resistance at the FET source is $1/g_m$. At 1 GHz, the FET $C_{\rm gs}$ and parasitic input capacitance, C_P, due to the bonding pads and external strays significantly shunt this resistance. Therefore, to achieve a good impedance match, the size and bias of the FET are selected for $1/g_m = 50\ \Omega$, and an inductor tunes out the shunt capacitance by parallel resonance in a frequency band around 1 GHz. As the capacitance at the LNA input is to be tuned, it makes good sense to do so with a grounded off-chip low-loss inductor, which also carries the LNA bias current.

Fortuitously, a FET with a small-signal channel resistance of 50 Ω produces a lower thermal noise current than a linear resistor of the same value [13]. The noise current spectral density in the FET is $4kT\gamma g_m A^2$/Hz, where $\gamma \simeq 0.67$ owing to the distributed inversion layer. Thus, ruling out any other noise sources in a matched LNA, the noise figure due to the FET alone is $10 \log 1.67 = 2.2$ dB. In a short-channel FET biased at unfavorable conditions, hot-electron effects may augment this [14] to raise the noise figure. Flicker noise in the FET is unimportant at RF.

A tuned load peaks the frequency response of the LNA in the band of interest (Fig. 3), in effect transforming the inherent lowpass characteristic of the amplifier to a bandpass. The load also helps to reject out-of-band signals and noise. However, the LNA passband is seldom sufficiently flat and narrow for RF *preselection*, that is, to suppress image channels and out-of-band interferers. Rather, a sharply tuned discrete filter, such as SAW or dielectric resonator, is inserted before the LNA for this purpose. Discrete filters usually operate at a characteristic impedance of 50 Ω or 75 Ω. As explained above, no preselect filter need follow this LNA, nor the RF buffer required to drive the filter.

The tuned load, therefore, comprises an inductor resonating with the FET drain capacitance, C_D, and the sum, C_P, of the input capacitance of the subsequent downconversion mixer and any other parasitics. Another advantage of the common-gate stage is that the somewhat large $C_{\rm GD}$ of the FET returns its current to a fixed bias, rather than to the input node as it would in a common-source amplifier. This current undergoes rapid phase-shifts with frequency in an RF tuned amplifier, and makes it difficult to design an input matching circuit.

The inductance, L, to tune this total capacitive load to the resonant frequency ω_0, in our case 2π Grad/s, is

$$L = 1/\omega_0^2(C_D + C_P). \tag{1}$$

In most modern FET's, the drain junction capacitance $C_D \simeq C_{\rm GS}$. Furthermore, the unity-current gain frequency, $\omega_T =$

[1] This assumes the FET is laid out sensibly. The gate resistance of a FET with a certain channel width is lowered by an interdigitated gate, whereas without such a layout, this resistance may dominate the FET input impedance at high frequencies.

g_m/C_{GS}. The inductance may then be expressed as

$$L = \frac{1}{\omega_0^2(g_m/\omega_T + C_P)}. \quad (2)$$

If inductor loss, as modeled by a series resistance R_S, limits the impedance of the tuned load at resonance, then using (2), the voltage gain of the common-gate stage is

$$\text{Gain} = g_m \frac{(\omega_0 L)^2}{R_S} = \frac{\omega_T}{1 + \omega_T C_P/g_m} \frac{L}{R_S}. \quad (3)$$

This form makes it clear which parameters are within the circuit designer's reach to determine RF gain. ω_T mainly depends on FET channel length, but is also controlled by gate bias. However, even with infinite ω_T, parasitic-related quantities will limit the maximum achievable gain to

$$\text{Max Gain} = \min\left(\frac{g_m}{C_P}\frac{L}{R_S}, g_m r_{DS}\right). \quad (4)$$

In the common-gate amplifier, the desired input impedance sets g_m. Thus, a large parasitic capacitance at the drain means a smaller achievable gain, unless the loss in the load inductor is somehow lowered to boost the gain. The relative quality of the inductor, L/R_S, depends on how it is physically realized, and there are limits to how large this may be in practice. For instance, at 1 GHz the L/R_S is 4 nH/Ω for a discrete 10 nH chip inductor meant for RF applications [15]. This argues for an on-chip inductor load, because it is unlikely that a discrete off-chip inductor can overcome, simply because of a higher quality, the RF gain loss due to the large parasitic capacitance of the bond pads, bondwires, package leads, and board traces.

The LNA fulfills its specifications at the price of power dissipation. At low values of gate drive voltage, long-channel FET laws fairly well describe the dependence of bias current, I_D, on the transconductance

$$I_D = \frac{1}{2} g_m (V_{GS} - V_t). \quad (5)$$

Similarly, at low gate drives, the ω_T of a FET theoretically follows the dependence

$$\omega_T = \frac{g_m}{C_{GS}} = \frac{\mu}{1 + \theta(V_{GS} - V_t)} \frac{(V_{GS} - V_t)}{L^2} \quad (6)$$

where θ captures how the inversion-layer mobility, μ, degrades with gate electric field [16]. While (5) is readily verified by measurement, there is little data in the literature on CMOS ω_T to validate (6). Therefore, the s-parameters of a single, large 1-μm NFET were characterized on a Cascade™ probe station, from which ω_T was deduced. The measured data (Fig. 4) shows that $\omega_T = 2\pi f_T$ conforms closely to (6) in the $V_{GS} - V_t$ range of interest. This curve serves as an important design aid.

Using (2), the load inductance may be calculated which tunes the LNA to a certain frequency in the absence of any significant parasitics. For instance, if the FET is biased at an f_T of 5 GHz, then the LNA requires a 40 nH inductor load to achieve a peak at 1 GHz in its frequency response. The inductor may be implemented on-chip as a square spiral in

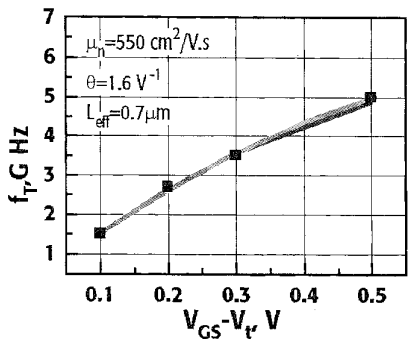

Fig. 4. Measured f_T versus $V_{GS} - V_t$ for MOSIS 1-μm NMOSFET (points) fitted to simple expression (line).

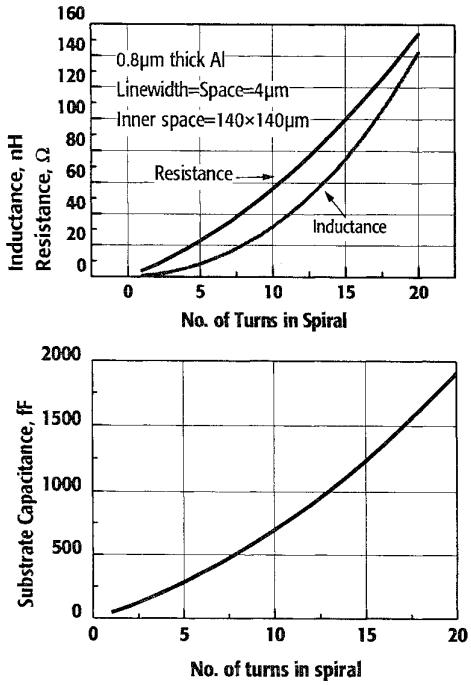

Fig. 5. Design curves for square spiral on-chip inductor. (a) Inductance and resistance of spiral versus number of turns and (b) capacitance of inductor (assuming it is an equipotential) to substrate through 1 μm thick field oxide.

Metal-2, with the return conductor in Metal-1. Greenhouse's formula accurately specifies how the inductance grows with the number of turns [17], while the series resistance accumulates with the number of squares of metal comprising the spiral [Fig. 5(a)]. The spiral grows outwards from a 140 μm square hole in the middle. Any turns within the hole would enclose relatively little magnetic flux, but would contribute a nonnegligible unwanted resistance. These curves show that in the inductance range of 40 to 50 nH, the relative quality, L/R_S, of the spiral is about 0.7 nH/Ω. It is seen from (3), (5), and Fig. 4 that the LNA may achieve a gain as large as 20 dB at 1 GHz in the absence of parasitic capacitance, while drawing 1.5 mA of current per FET.

The main impediment to a practical implementation of the tuned amplifier arises from the parasitic capacitance of a 50 nH

spiral inductor to the semiconductor substrate. This is so large through the typical 1 μm-thick field oxide [Fig. 5(b)] that the spiral self-resonates at 700 MHz, and at 1 GHz appears as a capacitive, rather than inductive, load on the LNA. This is why it is generally believed that medium- to large-value inductors may only be integrated on semi-insulating substrates, while on a standard silicon substrate inductors no larger in value than 5 to 10 nH are usable at 1 GHz [18].

In earlier work, we have described a method to eliminate the parasitic capacitance under the spiral inductor by selectively removing the underlying silicon substrate [1]. This leaves the spiral encased in a layer of oxide suspended above an air-gap. Inductors as large as 100 nH may be fabricated in CMOS, whose self-resonance frequency, now limited by the small fringing capacitance through the air gap to the distant ground plane, lies beyond 2 GHz. This maskless, post-fabrication etching of the substrate does not require any foundry modifications in the 1-μm CMOS process available through MOSIS. Via holes which will expose the surface of the silicon after fabrication and passivation surround the spiral. A selective etchant removes the exposed silicon at a much higher rate than it does oxide and metal. After sufficient exposure to the etchant, a deep cavity forms under the inductor, while the remaining active area on the chip is left intact. A passivating coat protects the exposed silicon on the sides and back of the chip from inadvertent etching.

The substrate was removed in the earlier work by a liquid-phase, anisotropic etchant [1], resulting in a spiral inductor surrounded by large trapezoidal openings, and attached by four oxide bridges to the rest of the CMOS substrate. This has since been replaced by a gas-phase, isotropic etchant. Through small circular openings surrounding the spiral, the etchant now excavates hemispherical pits whose radius increases with exposure time. Etching is stopped when the multiple pits coalesce into one large pit with a depth of roughly half of one side of the spiral (Fig. 6). This suspended inductor enables wholly-integrated RF components in CMOS. In addition to the LNA, it is also used in the local oscillator, power amplifier, and even as a low quality bandpass filter. As a survey in a recent publication shows [19], no simpler method has yet been found to realize large-value integrated inductors.

It was assumed in the earlier analysis that the impedance, Z, of the LNA tuned load is much less than the FET $r_{\rm ds}$. When this is not so, the expressions for the gain and input impedance must be modified to

$$\text{Gain} = \frac{g_m Z}{\left(1+\frac{Z}{r_{\rm ds}}\right)}; \qquad r_{\rm in} = \frac{1}{g_m}\left(1+\frac{Z}{r_{\rm ds}}\right). \tag{7}$$

The complete LNA is a balanced circuit (Fig. 7). A power-conserving, passive balun converts a single-ended antenna signal into a balanced input drive to the LNA. A printed-circuit balun may even be integrated into the transceiver case. The bias, VG, at the common-gate FET's regulates the LNA input impedance. An on-chip scaled-down replica circuit stabilizes this impedance against variations in process and temperature as follows. An op-amp drives the VG bias voltage of the replica to servo the dc value of $1/g_m$ to an off-chip reference

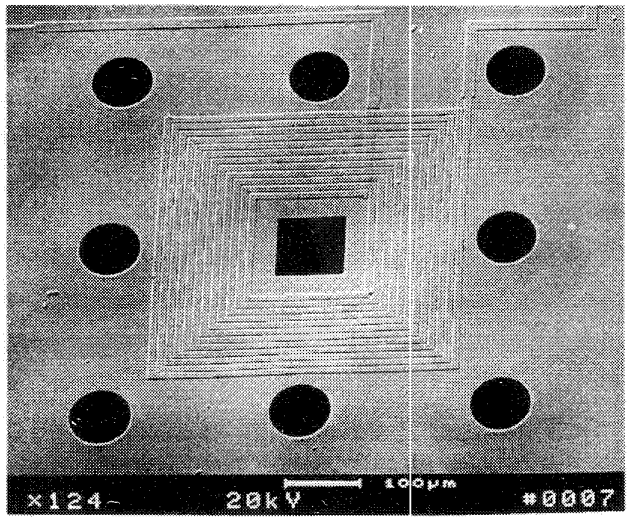

Fig. 6. Suspended 50 nH spiral inductor over a hemispherical cavity etched underneath through surrounding via holes.

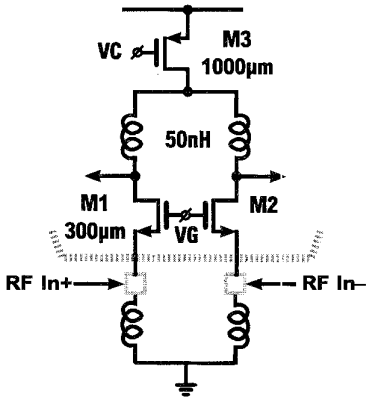

Fig. 7. Low-noise amplifier circuit diagram.

resistance, and the same voltage is then applied to the main LNA, thereby regulating its input impedance to within the FET matching in the replica and the main circuits. As the feedback loop bandwidth is well below 1 MHz, the op-amp does not contribute any RF noise. The inductor loads on each half of the circuit share a common top-node, which connects to the 3 V power supply through a triode-region PFET. By adjusting the gate voltage, VC, the dc voltage drop across this PFET resistor may be changed, and this sets the dc level at the LNA output.

The LNA FET's are biased at $V_{\rm GS}-V_t = 0.35$ V, and drain 2.2 mA each. Taking into account the capacitive load of the mixer, the LNA requires 50 nH load inductors to obtain a peak at 1 GHz in its frequency response. From (4) and Fig. 4, this means that in the absence of parasitics ($C_P = 0$), the peak RF gain is at most 26 dB.

B. Downconversion Mixer

The amplified signal at the LNA output is converted down in frequency for further amplification, channel-select filtering,

and detection. The frequency mixer is an integral part of the RF front-end.

There are two fundamentally different ways to implement a mixer in CMOS. The first, somewhat unconventional method, uses a MOSFET as a wideband analog switch. A passive track-and-hold circuit, consisting of a 1-μm FET switch and a grounded capacitor, is designed for a track-mode bandwidth of greater than 1 GHz. This follows the waveform of a modulated 1 GHz carrier and samples it at a much lower rate, which must be at least twice the *modulation* bandwidth. An op-amp feedback circuit clocked at this low sample rate buffers the held output, and removes signal-dependent charge injected by the switch. When the sample rate is an integer submultiple of the carrier frequency, the interpolated samples directly downconvert the RF signal to dc. A prototype evaluated at a 900 MHz RF [2] shows very good linearity, but fundamentally suffers from a large noise figure, because while tracking the narrowband signal, it also tracks and aliases wideband noise. This, and the difficulty of buffering such a switched mixer to the inductive load of an integrated LNA, make it inappropriate in a sensitive receiver.

The second, more conventional, mixer resembles the Gilbert analog multiplier. It consists of a linear RF voltage-to-current (V-I) converter, or RF transconductor, whose output current is commutated by the local oscillator (LO). As commutation conserves the total current, it downconverts a fraction of the RF current to the IF, and the remaining RF current upconverts around one or more harmonics of the LO. The voltage conversion-gain of the active mixer is independently set by choice of the transconductance and load resistance. The internal current conversion-loss penalizes mixer noise, as is analyzed in a later section.

A good mixer is highly linear, and its input-referred noise does not overwhelm the amplified noise of the preceding LNA. The mixer handles larger signals than the LNA, and therefore its nonlinearity must be lower by at least a factor of the LNA gain if it is not to become the bottleneck to receiver dynamic range. This is why the following discussion concentrates on mixer nonlinearity, as the LNA, with the choice of bias voltages, is not the bottleneck to front-end linearity.

Third-order intermodulation distortion in a double-balanced mixer may cause two large adjacent-channel signals to create energy at spurious frequencies coincident with a weak desired channel. The linearity of a front-end is specified by the input-referred third-order intercept point (IIP3) [20]. Often this is set by the static and dynamic nonlinearity in the RF V-I converter of the mixer, or by the static nonlinearity in the mixer load.

Linear MOS transconductance circuits have been studied extensively in the context of continuous-time active filters [21] operating at frequencies up to tens of MHz. These circuits exploit the property that the dominant second-order nonlinearity in a MOSFET circuit cancels in balanced differential inputs and outputs. For instance, if two identical common-source FET's conforming to the classic long-channel I-V characteristics are biased at some V_{GS} [Fig. 8(a)] and excited differentially by a large signal V_{in} whose amplitude is less

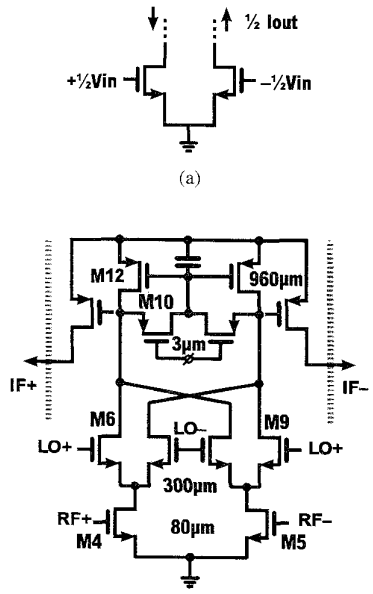

Fig. 8. (a) Linear MOS transconductor. (b) Downconversion mixer circuit diagram.

than $2(V_{GS} - V_t)$, then the differential output current

$$I_{out} = \mu C'_{ox}(W/L)(V_{GS} - V_t)V_{in} \tag{8}$$

depends *linearly* on V_{in}, and the bias $V_{GS} - V_t$ sets the transconductance [22].

Residual third-order nonlinearity produces a small distortion in the transconductor. Its large signal handling is limited by clipping when an input swing of $V_{GS} - V_t$ turns off one of the FET's in the circuit. These sources of static nonlinearity are expected to govern the mixer up to and beyond 1 GHz, as no significant nonquasi-static effects are likely to set in given the short carrier transit time in the 1-μm channel. Dynamic nonlinear currents which grow with frequency will flow in any voltage-dependent FET capacitance and might even become the significant form of distortion at 1 GHz. The MOSFET, however, is benign in this respect, as its main capacitance, C_{GS}, is relatively independent of bias for $V_{GS} > V_t$ and behaves like a linear capacitor in the saturation region of operation.

The MOS downconversion mixer is a balanced circuit [Fig. 8(b)] comprising a linear common-source FET transconductor (as opposed to a differential pair in the bipolar Gilbert multiplier), four commutating FET switches, and a high-swing load consisting of a center-tapped FET resistor across pull-up current sources. Common-mode feedback from the center tap biases the current sources at a well-defined voltage. The LNA output is directly connected into the differential mixer input.

The mixer attains its peak conversion gain when a sinewave of at least +5 dBm (1 V ptp) is applied to the commutating switches $M6$–$M9$. This also lowers the total front-end noise figure. While it is obvious that incomplete commutation leads to conversion loss, what may not be evident is how it also degrades noise figure. The transconductor FET's and the loads clearly contribute noise in the mixer. In addition, the

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.