Millimeter-Wave CMOS Design

Chinh H. Doan, *Student Member, IEEE*, Sohrab Emami, *Student Member, IEEE*, Ali M. Niknejad, *Member, IEEE*, and Robert W. Brodersen, *Fellow, IEEE*

***Abstract*—This paper describes the design and modeling of CMOS transistors, integrated passives, and circuit blocks at millimeter-wave (mm-wave) frequencies. The effects of parasitics on the high-frequency performance of 130-nm CMOS transistors are investigated, and a peak $f_{\max}$ of 135 GHz has been achieved with optimal device layout. The inductive quality factor (Q_L) is proposed as a more representative metric for transmission lines, and for a standard CMOS back-end process, coplanar waveguide (CPW) lines are determined to possess a higher Q_L than microstrip lines. Techniques for accurate modeling of active and passive components at mm-wave frequencies are presented. The proposed methodology was used to design two wideband mm-wave CMOS amplifiers operating at 40 GHz and 60 GHz. The 40-GHz amplifier achieves a peak $|S_{21}| = 19$ dB, output $P_{1\,dB} = -0.9$ dBm, IIP3 $= -7.4$ dBm, and consumes 24 mA from a 1.5-V supply. The 60-GHz amplifier achieves a peak $|S_{21}| = 12$ dB, output $P_{1\,dB} = +2.0$ dBm, NF $= 8.8$ dB, and consumes 36 mA from a 1.5-V supply. The amplifiers were fabricated in a standard 130-nm 6-metal layer bulk-CMOS process, demonstrating that complex mm-wave circuits are possible in today's mainstream CMOS technologies.**

***Index Terms*—CMOS millimeter-wave integrated circuits, coplanar waveguides, $f_{\max}$, integrated circuit modeling, high-speed integrated circuits, millimeter-wave amplifiers, Q-factor, transmission lines, wideband amplifiers.**

I. INTRODUCTION

IN THE last few years, 7 GHz of contiguous bandwidth have been opened for unlicensed use at millimeter-wave (mm-wave) frequencies around 60 GHz in the U.S. (57–64 GHz) and Japan (59–66 GHz). This allows for a variety of applications including gigabit/s point-to-point links, wireless local area networks (WLANs) with extraordinary capacity, short-range high data-rate wireless personal area networks (WPANs), and vehicular radar. In order for these applications to meet marketplace requirements, the cost, size, and power consumption of any solution has to be significantly below what is being achieved today using compound semiconductor technology. Thus, an alternative approach is required, such as using a mainstream digital 130-nm bulk-CMOS process.

Historically, monolithic microwave integrated circuits (MMICs) have been designed using III-V semiconductor technologies, such as GaAs and InP, which have superior performance compared to CMOS due to their higher electron mobility, higher breakdown voltage, and the availability of high quality-factor (Q) passives. Still, a CMOS implementation promises higher levels of integration and reduced cost. Several recent developments have combined to enable CMOS circuit blocks to operate at ever-increasing frequencies. First, mm-wave CMOS circuits directly benefit from the higher speed of the scaled technology. Additionally, improved circuit topologies and new design approaches to fully exploit the intrinsically faster devices have been introduced. Up to now, only CMOS oscillators [1]–[3] have been demonstrated beyond 30 GHz, while CMOS amplifiers [4]–[6] and mixers [6], [7] have only achieved operation up to 26 GHz and 21.8 GHz, respectively. A key reason for this large discrepancy is the lack of accurate CMOS active and passive device models at mm-wave frequencies.

In this paper, a design and modeling methodology, based on MMIC approaches but tailored for the specific details of CMOS, will be presented that enables, for the first time, CMOS amplifiers operating above 30 GHz. Section II gives a brief description of the CMOS technology used. The theoretical and practical high-frequency limitations of active devices are explored in Section III, and a transistor modeling methodology that results in simple, highly accurate models up to 65 GHz is discussed in Section IV. Transmission lines are important passive components at mm-wave frequencies, and the design and modeling of integrated CMOS transmission lines are presented in Sections V and VI. Finally, to demonstrate the effectiveness of our approach, 40-GHz and 60-GHz wideband amplifiers have been designed and fabricated in a 130-nm digital CMOS process. The 40-GHz amplifier attains 19-dB gain, output $P_{1\,dB} = -0.9$ dBm, IIP3 $= -7.4$ dBm, and consumes 36 mW. The 60-GHz amplifier achieves 12-dB gain, output $P_{1\,dB} = +2.0$ dBm, NF $= 8.8$ dB, and dissipates 54 mW.

II. CMOS TECHNOLOGY

This section provides a brief comparison between a modern digital 130-nm CMOS process and a dedicated microwave technology such as GaAs. Some of the key differences, which motivate our design choices, are highlighted.

A. Front-End Features

Two important disadvantages of a silicon metal–oxide–semiconductor field-effect transistor (MOSFET) compared to a GaAs field-effect transistor (FET) are: 1) the low-resistivity substrate and 2) the high sheet resistance of the polysilicon gates. The substrate resistivity of most modern standard silicon processes is ~10 Ω-cm, which is many orders of magnitude lower than that of GaAs (~10^7–10^9 Ω-cm) [8]. Signals that

Manuscript received April 20, 2004; revised June 14, 2004. This work was supported in part by the Army Communications-Electronics Command (CECOM) Contract DAAB07-02-1-L428 and the industrial members of the Berkeley Wireless Research Center.

The authors are with the Berkeley Wireless Research Center, University of California, Berkeley, CA 94704 USA (e-mail: cdoan@eecs.berkeley.edu; sohrab@eecs.berkeley.edu; niknejad@eecs.berkeley.edu; rb@eecs.berkeley.edu).

Digital Object Identifier 10.1109/JSSC.2004.837251

0018-9200/$20.00 © 2005 IEEE

couple to the low-resistivity silicon substrate incur significant losses, especially at mm-wave frequencies. Furthermore, whereas a GaAs FET can effectively be treated as a three-terminal device, the existence of the bulk terminal and the body-effect complicate matters for MOS designs.

The gate material used for CMOS devices is polysilicon, which has a much higher sheet resistance ($\sim$10 $\Omega/\square$) than the metal used for the gates of GaAs FETs. A higher gate resistance can reduce the transistor power gain and increase noise. Fortunately, simple layout techniques can be used to minimize the detrimental effects of the polysilicon gate.

B. Back-End Features

The core back-end stack consists of six levels of copper metallization. Top-layer metal is 0.9 μm thick and the distance from the substrate is 5 μm. Chemical mechanical polishing (CMP) is used to planarize all metals and dielectrics, providing better repeatability of the conductor and oxide thicknesses compared to GaAs. Due to the use of CMP, though, uniform density is required on all metals. Thus, floating dummy fill metal is needed to increase the local density, while large areas of metal (e.g., for ground planes) are forced to have slots. To reduce costs, metal–insulator–metal (MIM) capacitors were not used.

III. Transistor Design

In this section, the effect of parasitics on the high-frequency performance limits of CMOS transistors is presented. Design guidelines for optimal transistor layout are provided along with experimental verification.

A. Maximum Frequency of Oscillation

The most relevant figure-of-merit for the high-frequency capabilities of a process is the maximum frequency of oscillation ($f_{\max}$). The value of $f_{\max}$ is determined not only by sizing and bias conditions, but is also highly dependent on resistive losses due to transistor and layout parasitics. Using Mason's unilateral gain, $f_{\max}$ of a CMOS transistor can be determined by finding the frequency where $U = 1$ [9].

Although it is common practice to use low-frequency measurements of U and report $f_{\max}$ as the *extrapolated* frequency where $U = 1$ (assuming a 20 dB/decade slope), at frequencies approaching the true $f_{\max}, U$ often drops at a rate much faster than 20 dB/decade (Fig. 1). The $f_{\max}$ values reported in this paper are extrapolated from the circuit models described in Section IV and do *not* assume a 20-dB/decade slope.

B. Layout for Optimal $f_{\max}$

It can be shown theoretically that $f_{\max}$ is independent of the number of fingers of a multi-finger transistor. Therefore, it is sufficient to only consider the optimal layout for a single finger. The physical layout of a single finger is shown in Fig. 2, along with a physical model depicting the dominant high-frequency loss mechanisms. As mentioned, $f_{\max}$ is limited by resistive losses, the most significant being the gate resistance (R_G), series source/drain resistances (R_S, R_D), nonquasi-static channel resistance ($r_{\rm nqs}$), and resistive substrate network (R_{sb}, R_{db}, and R_{bb}) [10].

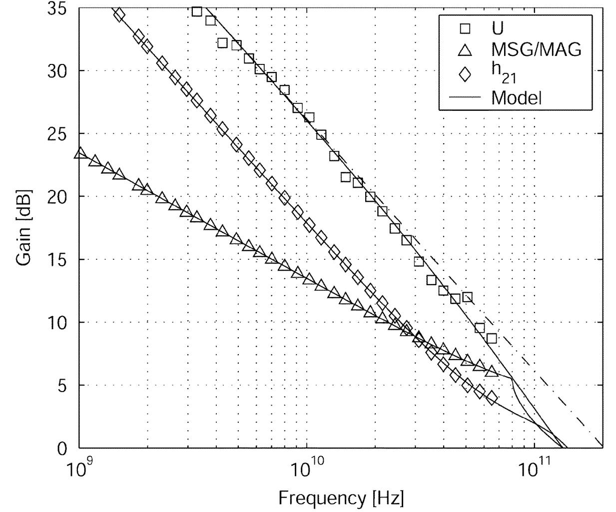


Fig. 1. Measured (markers) and modeled (solid lines) unilateral gain U, maximum stable gain MSG, maximum available gain MAG, and current gain h_{21}, for a typical NMOS device ($W/L = 100 \times 1$ μm/0.13 μm, $I_{\rm DS}/W = 300$ μA/μm, $V_{\rm DS} = 1.2$ V). The maximum frequency of oscillation, based upon the device circuit model, is $f_{\max} = 135$ GHz. This is much lower than the value of 200 GHz attained if a 20-dB/decade slope is assumed.

By using narrow finger widths, the effect of the gate resistance can be made negligible compared to the other parasitics resistors. The polysilicon gate sheet resistance only affects how narrow the fingers must be made. Therefore, with optimal layout, $f_{\max}$ is not limited by the gate resistance, but is primarily determined by the series source/drain resistances and substrate losses.

C. Measured $f_{\max}$

The optimal transistor finger width for our 130-nm digital CMOS process has been determined empirically. The measured $f_{\max}$ for NMOS transistors with minimum channel length as a function of finger width and bias current density is displayed in Fig. 3. Nine devices with $W_F = 1$–8 μm and $N_F = 40$–100 in common-source configuration with the bulk and source grounded and the gate contacted on one side were fabricated. For six bias points ($I_{\rm DS}/W = 20$–300 μA/μm) per device, a transistor model was extracted from the measured data in order to find $f_{\max}$ (see Section IV). The constant $f_{\max}$ contours shown in Fig. 3 were linearly interpolated between the measured data points.

For a constant current density, the device f_t remains fixed (e.g., $I_{\rm DS}/W = 100$ μA/μm, f_t is 70 GHz). It is clear from Fig. 3 that, depending on the finger width, $f_{\max}$ can be much larger or smaller than f_t. Thus, the optimal layout for mm-wave applications requires CMOS transistors to be designed using many extremely narrow fingers in parallel (less than 1 μm each). This is in stark contrast to GaAs FETs with metal gates, where relatively few fingers of wide devices ($W_F \approx 30$–75 μm) are typically used [8]. Furthermore, the device must be biased well into strong inversion (around 100–300 μA/μm) for mm-wave operation. By proper layout and biasing, though, the $f_{\max}$ of an NMOS transistor in a standard 130-nm CMOS technology can easily surpass 100 GHz, opening the possibility for mm-wave circuits.

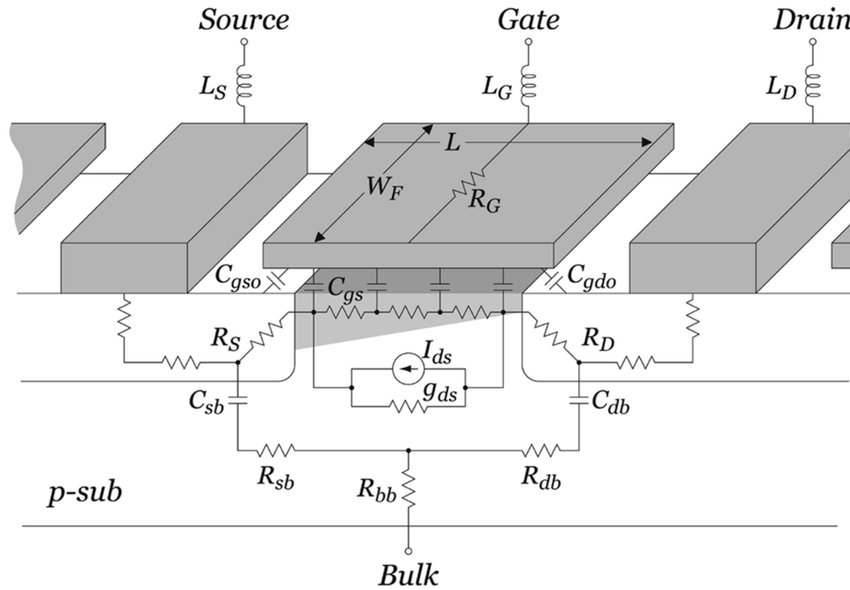


Fig. 2. Simplified physical model for one finger of an NMOS device.

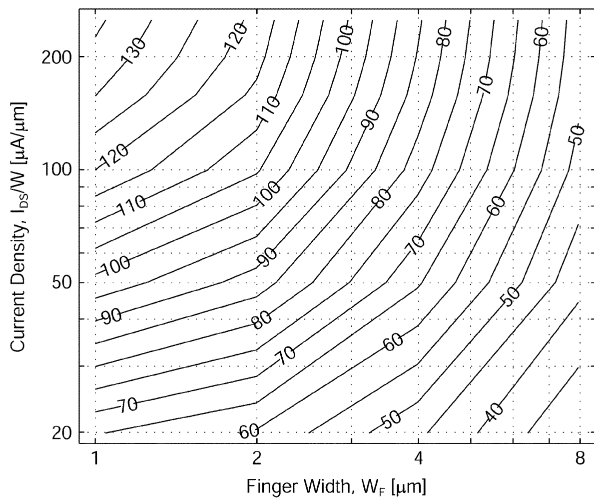


Fig. 3. Measured $f_{\max}$ [GHz] for minimum channel length ($L = 0.13$ μm) NMOS transistors. The constant $f_{\max}$ contour lines are linearly interpolated between the measurement data. The peak measured $f_{\max}$ for a 100×1 μm/0.13 μm device biased at $I_{\rm DS}/W = 250$ μA/μm is 135 GHz.

IV. Transistor Modeling

The traditional microwave approach to transistor modeling uses measured S-parameter data for circuit design. Although S-parameter models are accurate and sufficient for many designs, a circuit model provides the ability to extrapolate to frequencies beyond the measurement capabilities of the test equipment. Additionally, an accurate nonlinear large-signal transistor model [11] is required for the design of mixers, oscillators, and power amplifiers.

A. Extended CMOS Transistor Models

At mm-wave frequencies, series resistive and inductive parasitics become more significant. Consequently, it is critical to properly model these parasitics, in addition to the capacitive effects that are traditionally captured by digital CMOS models

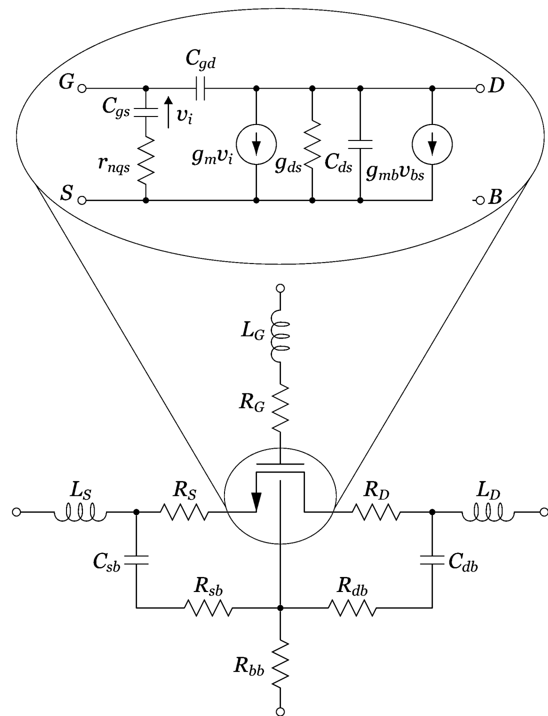


Fig. 4. Small-signal transistor model for an NMOS device showing the important parasitic elements.

[10]. Considering the small margins for modeling errors, the following modeling methodology for active devices was adopted to yield a model with the highest possible accuracy:

- Since the precise layout details—connections to the gate, drain, source, and bulk, location of the substrate contacts, number of fingers, etc.—have a major impact on the parasitic elements, models were extracted only for fixed layouts.
- The transistors in the circuit have the identical layout as the devices used for the model extraction.

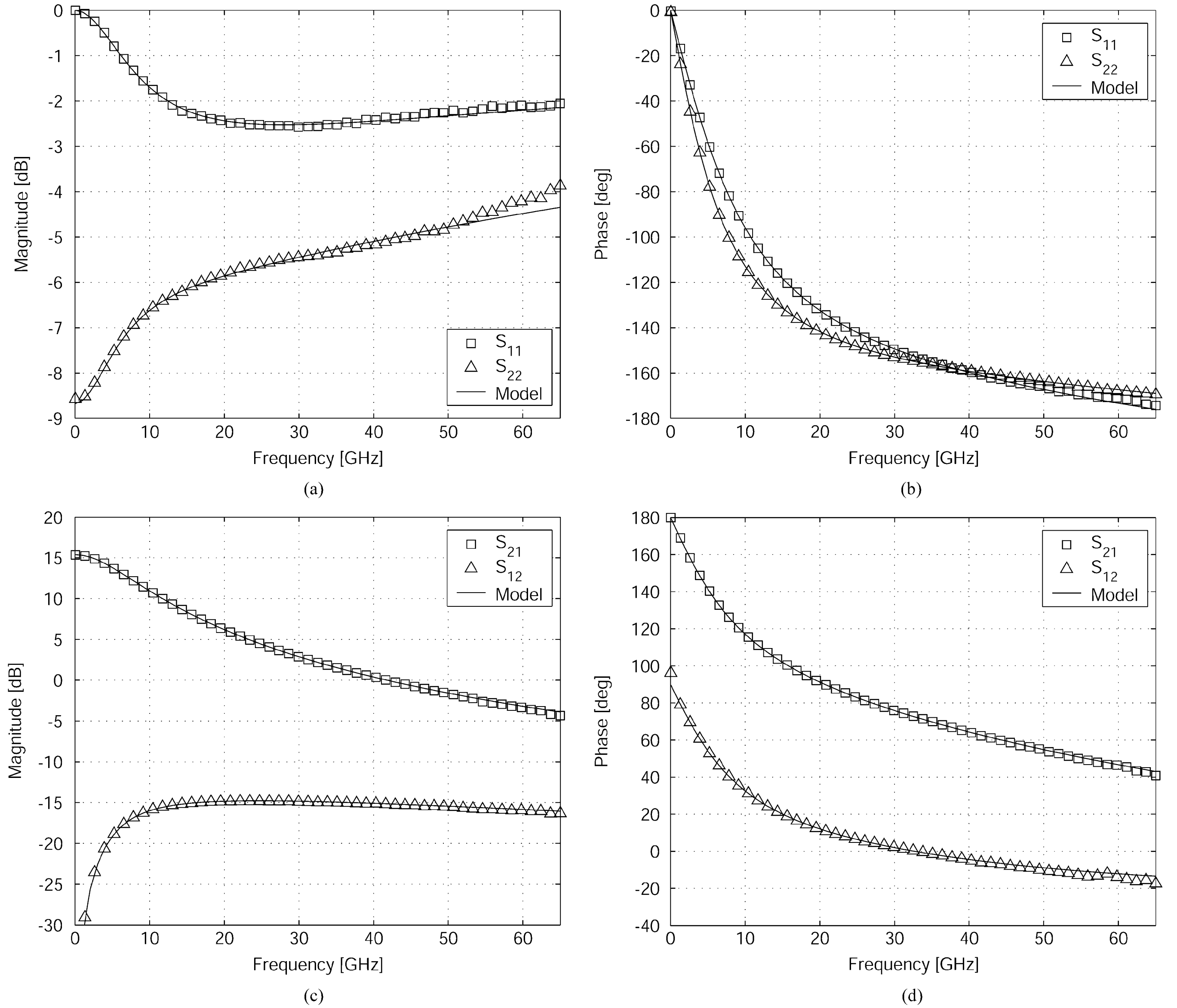


Fig. 5. Measured (markers) and simulated (solid lines) S-parameters for a typical NMOS device ($W/L = 100 \times 1$ μm/0.13 μm, $I_{\rm DS}/W = 300$ μA/μm, $V_{\rm DS} = 1.2$ V).

- For the highest accuracy, a bias-dependent small-signal model is extracted. For increased flexibility, a large-signal transistor model based on BSIM3 has also been demonstrated to provide good results up to 65 GHz [11].

The physical model depicting the significant high-frequency parasitics was shown in Fig. 2, and Fig. 4 shows the corresponding extended circuit model. The core device is modeled using either a lumped small-signal model (Fig. 4) or using a standard BSIM3 model card. In addition to the parasitic resistors, series inductors must be added to all terminals—L_G, L_D, L_S—to properly model the delay effects associated with interconnect wiring. Notice that all of the capacitors (e.g., C_{gd}, C_{gs}, C_{ds}) account for both the traditionally "intrinsic" channel and overlap capacitances as well as the traditionally "extrinsic" wiring capacitances.

For each model, the extrinsic component values and device parameters were extracted from measured data using a hybrid optimization algorithm in Agilent IC-CAP [19]. S-parameters for the simulated small-signal model and measured data up to 65 GHz are shown in Fig. 5 for a 100×1 μm/0.13 μm NMOS transistor biased at $V_{\rm GS} = 0.65$ V and $V_{\rm DS} = 1.2$ V. The excellent broadband accuracy of the simulation compared to the measured data verifies that the topology of our model is correct and complete. Furthermore, it also demonstrates that distributed effects and frequency-dependent losses caused by the skin effect can be adequately accounted for using only lumped extrinsic components with frequency-independent values. Large-signal verification has also been performed on the BSIM3 model [11], showing good distortion prediction at mm-wave frequencies. The default BSIM3 noise model was used since no device noise measurements were available.

The transistor gains—Mason's unilateral gain, maximum stable gain (MSG), maximum available gain (MAG), and current gain—for this device are plotted in Fig. 1. The accurate

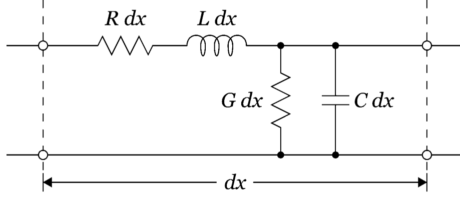


Fig. 6. Distributed RLGC lossy transmission line model.

modeling of the unilateral gain is particularly important. Unlike the MSG and current gain, Mason's unilateral gain is a very strong function of all resistive losses. Therefore, accurately fitting the unilateral gain validates that the important loss mechanisms have been properly modeled. As mentioned earlier, these resistive losses are critical because they ultimately limit the high-frequency capabilities of the transistor.

V. Transmission Lines

Transmission lines (T-lines) are important structures for mm-wave design. At these frequencies, the reactive elements needed for matching networks and resonators become increasingly small, requiring inductance values on the order of 50–250 pH. Given the quasi-transverse electromagnetic (quasi-TEM) mode of propagation, T-lines are inherently scalable in length and are capable of realizing precise values of small reactances. Additionally, interconnect wiring can be modeled directly when implemented using T-lines. Another benefit of using T-lines is that the well-defined ground return path significantly reduces magnetic and electric field coupling to adjacent structures.

Any quasi-TEM T-line can be completely characterized by its equivalent frequency-dependent RLGC distributed circuit model (Fig. 6). The line can also be characterized by the following four *real* parameters:

$$Z \equiv \sqrt{L/C} \tag{1}$$

$$\lambda \equiv \frac{2\pi}{\omega_0\sqrt{LC}} \tag{2}$$

$$Q_L \equiv \omega_0 L/R \tag{3}$$

$$Q_C \equiv \omega_0 C/G. \tag{4}$$

Unlike T-lines implemented on GaAs, where G is essentially zero, T-lines implemented on low-resistivity silicon often have low capacitive quality factors (Q_C) due to the substrate coupling. For T-lines that store mostly magnetic energy, the *inductive quality factor* (Q_L) is the most critical parameter when determining the loss of the line, as opposed to the resonator quality factor or the attenuation constant.

A. Inductive Quality Factor

Transmission lines are often used to resonate with the intrinsic capacitance of the transistors (e.g., when used in matching networks). In this case, the line stores mostly magnetic energy, and it is therefore most appropriate to consider the power lost for a given amount of *net* reactive energy stored in the line, as opposed to the *total* stored energy [13]. Thus, for matching networks, the most meaningful metric is

$$Q_{\text{net}} \equiv 2\omega_0 \frac{\text{net energy stored}}{\text{avg. power loss}} = \frac{2\omega_0(W_m - W_e)}{P_R + P_G} \tag{5}$$

where ω_0 is the resonance frequency, W_m and W_e are the average magnetic and electric energy stored, and P_R and P_G are the average power dissipated in the resistance and conductance, respectively. If Q_L and Q_C are rewritten as

$$Q_L = 2\omega_0 W_m / P_R \tag{6}$$

$$Q_C = 2\omega_0 W_e / P_G \tag{7}$$

it is straightforward to show that

$$\frac{1}{Q_{\text{net}}} = \frac{1}{\eta_L Q_L} + \frac{1}{\eta_C Q_C} \tag{8}$$

where

$$\eta_L = 1 - \frac{W_e}{W_m} \tag{9}$$

$$\eta_C = \frac{W_m}{W_e} - 1. \tag{10}$$

If the line is inductive (i.e., $W_m \gg W_e$), then $\eta_C \gg \eta_L$, and $Q_{\text{net}} \approx \eta_L Q_L$. The loss in the line is therefore almost completely determined by Q_L. For example, consider a shorted transmission line with $l < 0.1\lambda$. In this case, it can be shown that $\eta_C > 7.2\eta_L$, which greatly reduces the impact of the shunt losses on the inductive line. This is particularly important for integrated T-lines on silicon, where the low-resistivity substrate causes Q_C to be nonnegligible. A similar qualitative discussion and conclusion for inductive lines has been presented in [14].

B. Microstrip Versus Coplanar Waveguides

Microstrip lines on silicon are typically implemented using the top-layer metal as the signal line, and the bottom-layer metal for the ground plane. Fig. 7(a) illustrates the effectiveness of the metal shield, with essentially no electric field penetration into the substrate. The shunt loss, G, is therefore due only to the loss tangent of the oxide, yielding a capacitive quality factor, Q_C, of around 30 at mm-wave frequencies [Fig. 8(b)]. The biggest drawback to microstrip lines on standard CMOS is the close proximity of the ground plane to the signal line ($\sim$4 μm), yielding very small distributed inductance, L. This significantly degrades the inductive quality factor, Q_L [Fig. 8(a)].

Another option for on-chip transmission lines is the use of coplanar waveguides (CPWs) [16], [17], which are implemented with one signal line surrounded by two adjacent grounds [Fig. 7(b)]. The signal width, W, can be used to minimize conductor loss, while the signal-to-ground spacing, S, controls the Z_0 and the tradeoff between Q_L, and Q_C. As an example, a CPW with $W = 10$ μm and $S = 7$ μm has a Z_0 of 59 Ω and a Q_L measured to be about double that of the microstrip [Fig. 8(a)]. Therefore, CPW T-lines were used in our designs due to their considerably higher Q_L compared to microstrip lines.

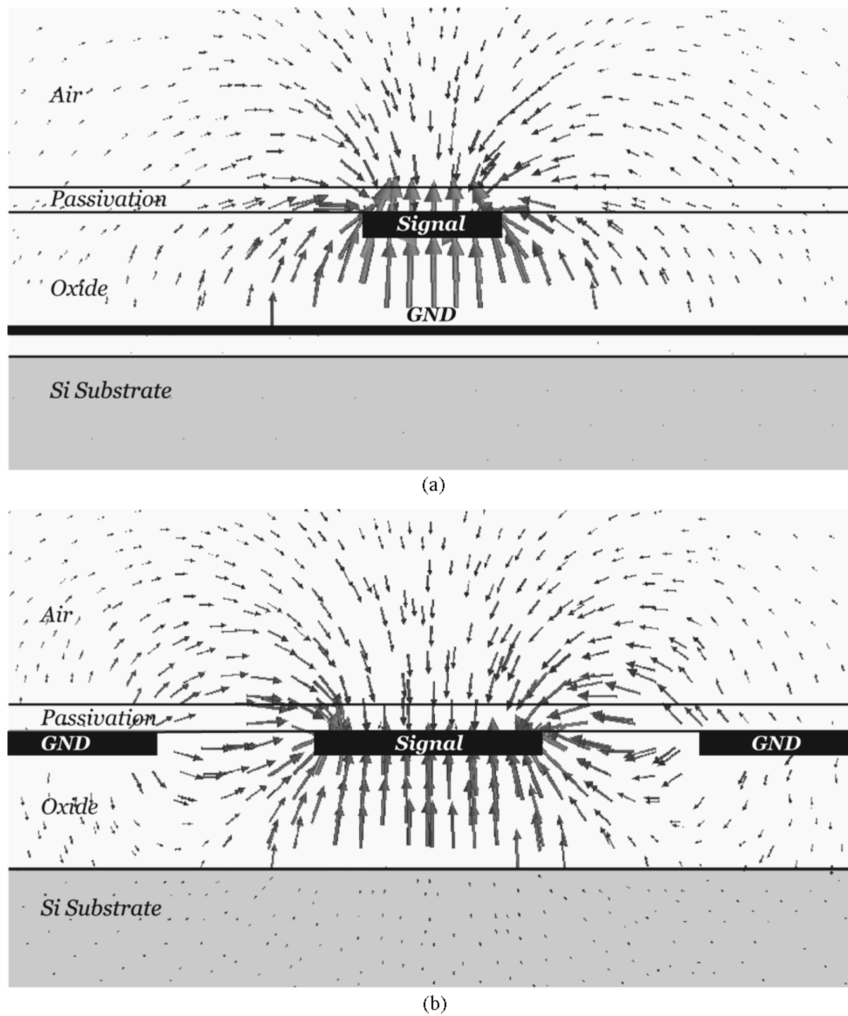


Fig. 7. Electric field distributions from 3-D EM simulations of (a) microstrip and (b) coplanar waveguide transmission lines.

By varying the signal-to-ground spacing, it is possible to design CPW lines to have either large Q_L and high-impedance ($S = 7\ \mu$m) or large Q_C and low-impedance ($S = 2\ \mu$m) (Fig. 9). On the other hand microstrip lines have, to first-order, constant Q_L and Q_C regardless of geometry. Another important issue when designing with CPWs is the unwanted odd CPW mode, which arises because CPW lines inherently have three conductors. To suppress this parasitic propagation mode, the two grounds should be forced to the same potential [16]. In MMICs, this requires the availability of air bridge technology, which is costly and not supported by all foundries. Underpasses using a lower metal level in a modern CMOS process can be used to suppress this mode.

VI. Transmission Line Modeling

To achieve the highest level of accuracy for the transmission line models, a design-oriented modeling methodology, similar to [18], has been chosen for this work. The modeling approach is based on measured transmission line data and the models are optimized to fit most accurately at mm-wave frequencies. Scalable (in length) electrical models, which capture the high-level behavior of the lines, have been used and are supported in most simulators such as SpectreRF, ADS, and Eldo. The model parameters are easy to obtain from measured data or physical EM simulations since only a relatively small number of parameters are required to model the broadband performance of each transmission line: characteristic impedance, effective dielectric constant, attenuation constant, and loss tangent. A first-order frequency-dependent loss model is used. The model assumes that the conductor loss is only caused by the skin effect losses, and the shunt loss is due to a constant loss tangent. From Figs. 8 and 9, it can be seen that the losses are well-modeled.

Using simple electrical models has many advantages. The simulation time is very fast, and the models can be easily integrated into circuit simulators and optimizers. The T-line models assume no coupling to adjacent structures. This assumption is justified as the well-defined ground return path, which helps confine the magnetic and electric fields, and the close proximity of the adjacent grounds to the signal line help to minimize any second-order effects.

A CMOS test chip was fabricated which included 1-mm long CPW and microstrip transmission lines of different cross-sectional dimensions. A different transmission line model was extracted for each geometry. Notice that although the overall loss

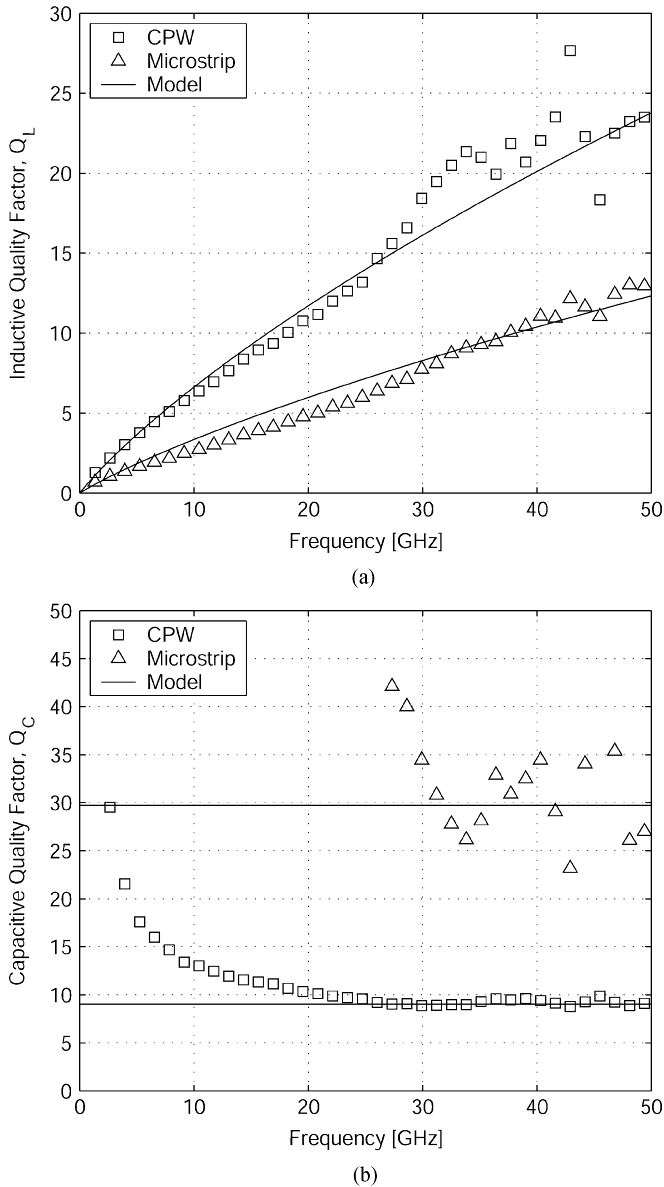


Fig. 8. Measured (markers) Q_L and Q_C for a coplanar waveguide and a microstrip line. The solid lines are for an empirical first-order model where the R variation is due only to skin effect and G is caused by a constant dielectric loss tangent.

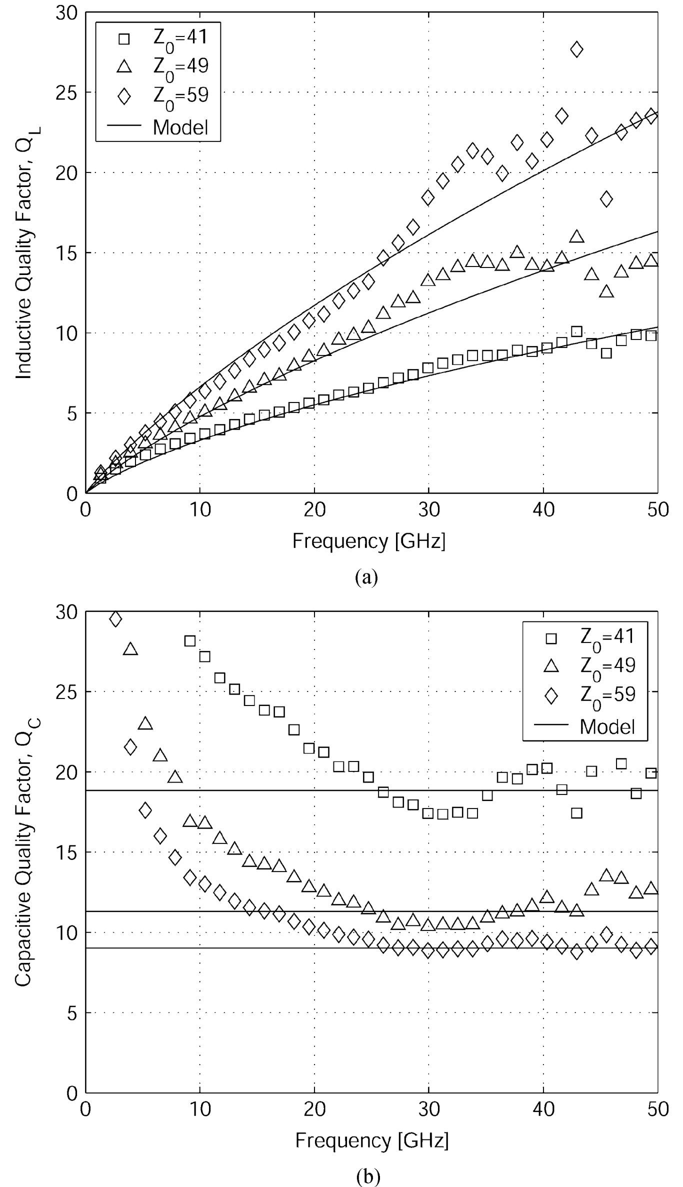


Fig. 9. Measured (markers) Q_L and Q_C for coplanar waveguides with varying geometries. The solid lines are for an empirical first-order model where the R variation is due only to skin effect and G is caused by a constant dielectric loss tangent.

of the T-line can be accurately extracted from measurements, decomposing this loss into individual loss components causes the measured high-Q data (e.g., Q_C for microstrip) to exhibit more measurement uncertainty than the low-Q data.

VII. Coplanar Waveguide Filter

To validate the electrical passive models, a 30-GHz bandpass filter, composed of series and shunt stubs of the modeled CPW lines, was designed (Fig. 10). The topology of the filter is equivalent to a bandpass ladder filter. The shunt resonator is replaced by an open-circuited low-impedance line and a short-circuited high-impedance line. The series transmission line in the center replaces the series resonator. Note that all lines are much shorter than a wavelength ($\lambda = 5$ mm on SiO_2) to minimize loss. Pad models were also extracted from a test chip, and the pads were included as part of the filter. An optimizer was used to fine tune the line lengths.

A die photo of the filter is shown in Fig. 11, measuring 0.93 mm $\times$ 0.64 mm including pads. Although the transmission lines meander, no special modeling of the bends or junctions was performed. The measured and simulated results for the 30-GHz filter are plotted in Fig. 12, demonstrating excellent broadband agreement by using just the simple, scalable electrical models. The measured insertion loss is 2 dB, and the input and output return losses are better than 25 dB. The accurate prediction by the electrical models demonstrates that junctions and bends are not critical, and verifies that there is no significant coupling between the individual lines.

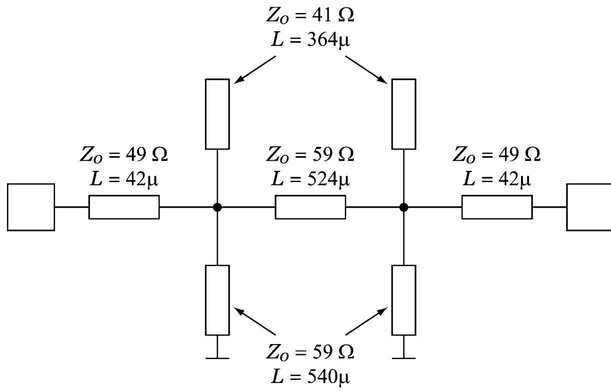


Fig. 10. Schematic of the 30-GHz CPW filter.

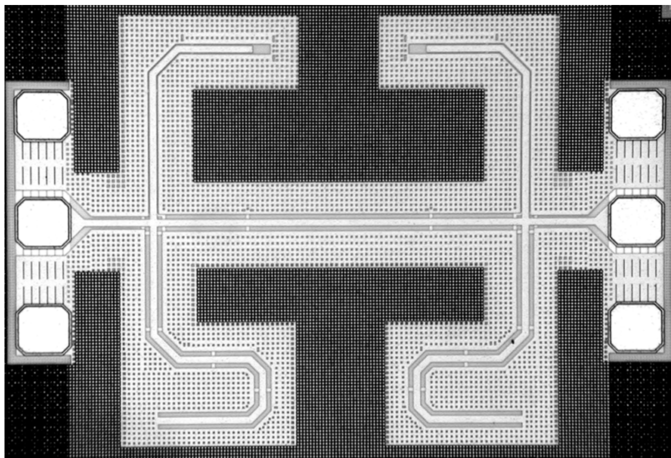


Fig. 11. Chip microphotograph of the 30-GHz CPW filter.

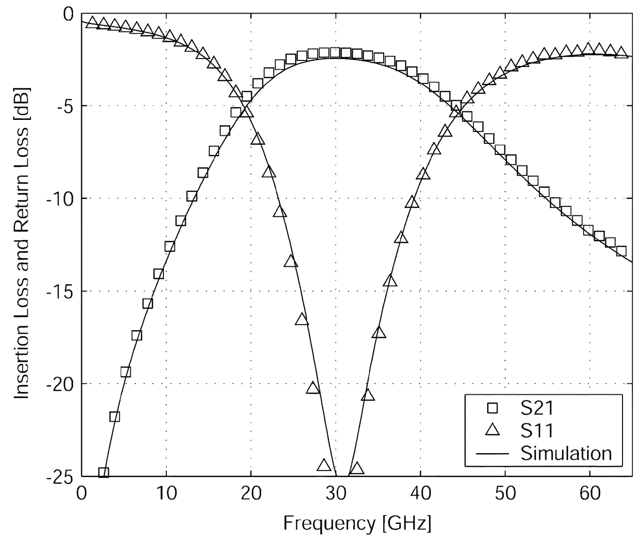


Fig. 12. Measured and simulated results for the 30-GHz CPW filter.

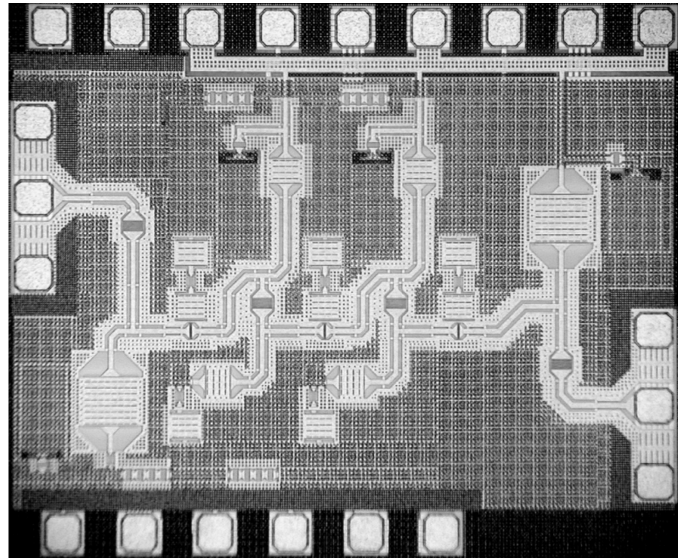


Fig. 13. Chip microphotograph of the 3-stage 40-GHz CPW amplifier.

VIII. Amplifier Design

Two wideband mm-wave amplifiers designed as general-purpose amplifiers operating at 40 GHz and 60 GHz have been fabricated in a 130-nm digital CMOS technology with no special analog or RF options. Figs. 13 and 14 show the die microphotograph of the 40-GHz amplifier, which measures 1.3 mm $\times$ 1.1 mm including pads, and the 60-GHz amplifier, which measures 1.3 mm $\times$ 1.0 mm including pads.

Both amplifiers were designed to have about 25% bandwidth. The topology of the two amplifiers is essentially identical, consisting of three stages of cascode devices with input, output, and interstage reactive matching (Fig. 15). The only significant differences between the two amplifiers are the bias currents and lengths of the transmission lines.

Cascode transistors are used in order to reduce the Miller capacitance and improve stability, and are unconditionally stable above 27 GHz. The cascode transistors for the 40 GHz amplifier are biased at a current density of 100 μA/μm, with a MAG of 8.9 dB at 40 GHz. For the 60-GHz amplifier, the cascodes are biased at 150 μA/μm, and the MAG is 6.0 dB at 60 GHz. The devices are biased from a V_{DD} of 1.5 V for increased headroom and output power. From simulations, all terminal-pair voltages for the individual transistors remain below the rated breakdown voltages.

CPW transmission lines are used extensively in the design for impedance matching, interconnect wiring, and the bias networks. All lines are kept as short as possible to minimize losses and are significantly shorter than $\lambda/4$ ($<$190 μm for the 40-GHz amplifier, $<$82 μm for the 60-GHz amplifier). The T-lines at the gate and drain are used to supply bias and are also incorporated into the matching networks. Meander CPW lines are used throughout the 40-GHz design in order to reduce area. The insertion loss of the interstage matching network is 2.5 dB for the 40-GHz design and 1.8 dB for the 60-GHz amplifier. Interestingly, the losses due to the passives are lower at 60 GHz. Although the conductor loss due to skin effect increases with frequency, the lines needed for the matching networks become shorter.

The input and output of the amplifiers are ac-coupled, and the GSG pads are included as part of the design. Both ports are designed to be matched to 50 Ω. The insertion loss of the input matching network is 1.6 dB and 1.3 dB, and the insertion loss

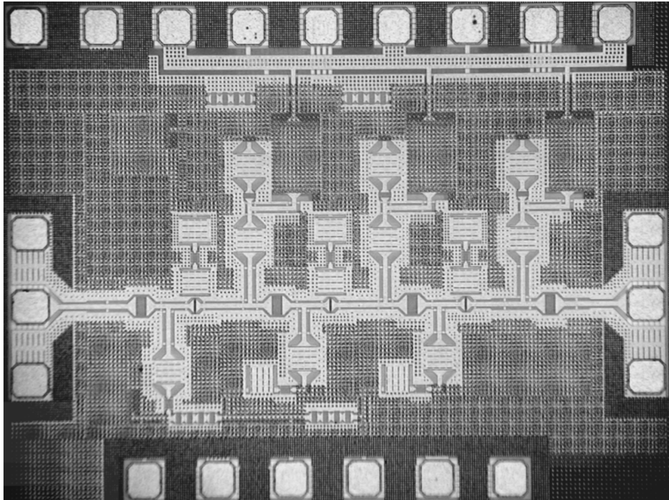


Fig. 14. Chip microphotograph of the 3-stage 60-GHz CPW amplifier.

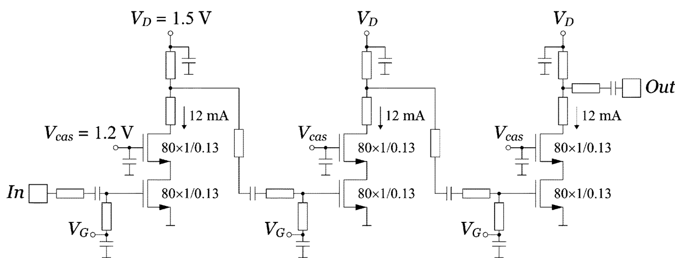


Fig. 15. Simplified schematic of the 60-GHz 3-stage amplifier using CPW transmission lines.

of the output matching network is 2.0 dB and 1.6 dB for the 40-GHz and 60-GHz amplifiers, respectively.

IX. Measurement Results

A. Measurement Setup

On-wafer S-parameter measurements up to 65 GHz were performed using a Cascade Microtech probe station, GSG coplanar probes, and an Anritsu 37 397C VNA. Open-short de-embedding [12] was used to remove the effects of the pads when measuring individual devices (transistors, transmission lines, etc.). Pad removal was not necessary for the filter or amplifiers since the pads were incorporated into the design.

The VNA was also used for single-tone compression measurements up to 65 GHz. A 65-GHz Anritsu SC6230 power sensor and Anritsu 2437A power meter was used to calibrate out the losses of the cabling and probes.

On-wafer vector-corrected two-tone distortion measurements were performed on a custom setup using the procedure described in [20]. The measurement system was limited to 50 GHz, so the intermodulation distortion was characterized only for the 40-GHz amplifier. A reflectometer external to the VNA was mounted on the probe station to maximize dynamic range. Both on-wafer and coaxial calibration standards are needed to obtain the vector-corrected power at the probe tip. A full description of the procedure and algorithm for de-embedding can be found in [20].

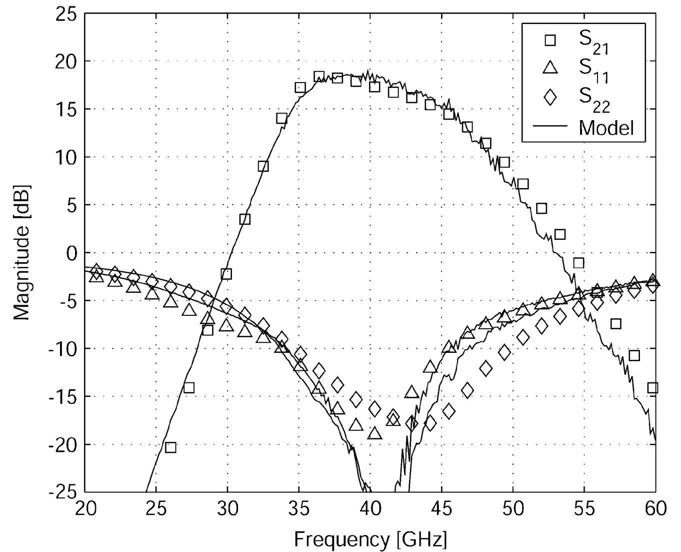


Fig. 16. Measured (markers) and simulated (lines) S-parameters for the 40-GHz amplifier.

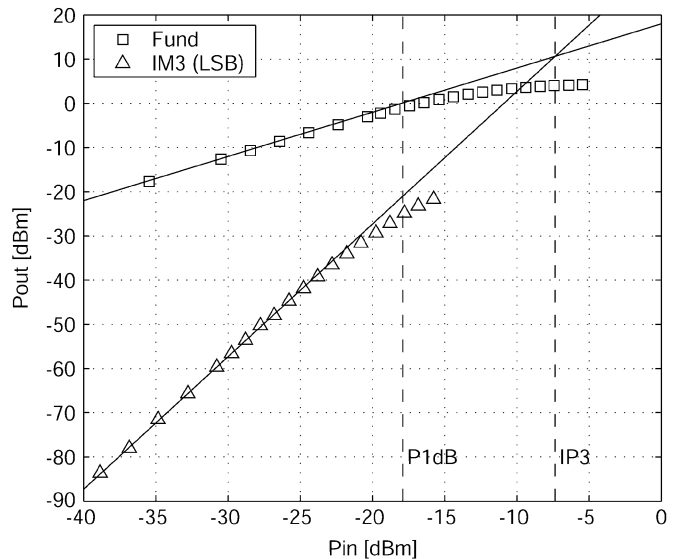


Fig. 17. Measured two-tone (38 GHz and 38.25 GHz) distortion for the 40-GHz amplifier.

Noise figure (NF) measurement of the 60-GHz amplifier was performed using a Millitech WR-15 noise source, WR-15 waveguide probes, output isolator, OML 50–75-GHz DSB downconversion mixer, and an Agilent N8973A NF measurement system.

B. 40-GHz and 60-GHz Amplifier Results

The measured and modeled S-parameters for the 40-GHz amplifier are shown in Fig. 16. The amplifier achieves a peak power gain of 19 dB, and input and output return losses are >15 dB. The 3-dB bandwidth is 34–44 GHz, and the amplifier maintains good return losses across this band. The measured reverse isolation is better than 50 dB up to 65 GHz, indicating that parasitic coupling through the silicon substrate is very small. This isolation is obtained without any special isolation strategy such as triple-well. The two-tone intermodulation distortion

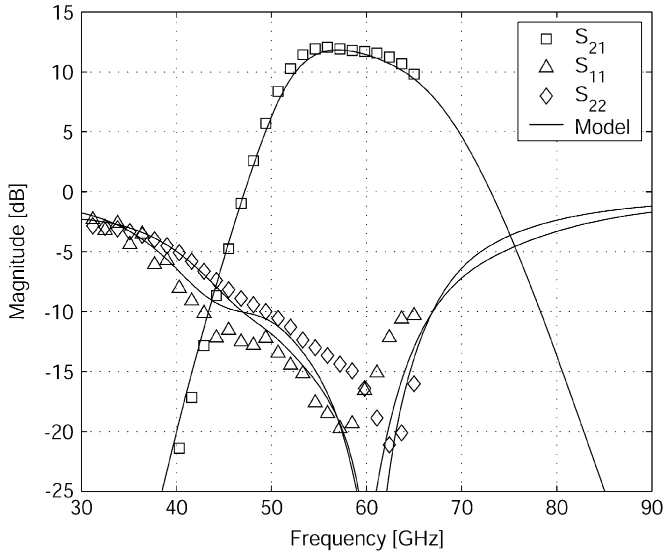


Fig. 18. Measured (markers) and simulated (lines) S-parameters for the 60-GHz amplifier.

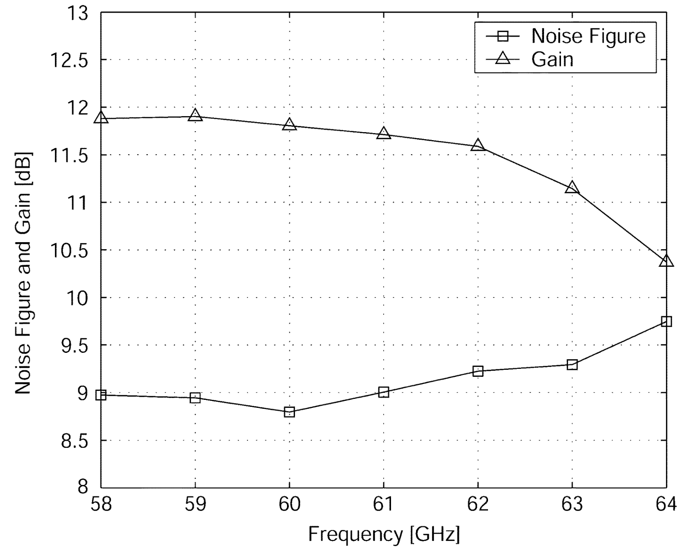


Fig. 20. Measured noise figure and gain for the 60-GHz amplifier.

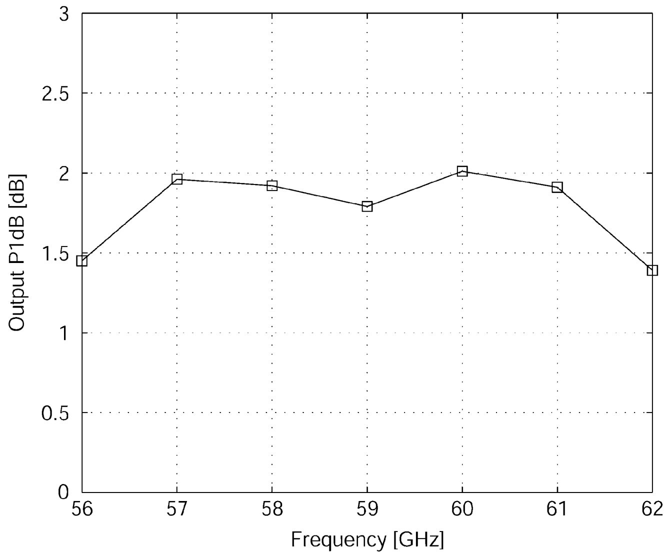


Fig. 19. Measured output 1-dB compression point for the 60-GHz amplifier.

TABLE I
COMPARISON WITH STATE-OF-THE-ART BULK-CMOS AMPLIFIERS

Technology	0.18-μm CMOS	0.18-μm CMOS	0.13-μm CMOS	0.13-μm CMOS
Frequency [GHz]	24	25.7	40	60
Gain [dB]	10	8.9	19	12
3-dB BW [GHz]	-	23–27.5	34–44	51–65
Output P_{1dB} [dBm]	-	–1.3	–0.9	+2.0
IIP3 [dBm]	–3.0	+2.8	–7.4	-
NF [dB]	6.0	6.9	-	8.8
S_{11} [dB]	–14	< –14	< –15	< –15
S_{22} [dB]	-	< –12	< –15	< –15
S_{12} [dB]	-	< –32.5	< –50	< –45
Power dissipation	47 mA	30 mA @ 1.8 V	24 mA @ 1.5 V	36 mA @ 1.5 V
Reference	[5]	[4]	This work	

measurements are shown in Fig. 17. The measured output 1-dB compression point $(P_{1\,dB})$ is -0.9 dBm and IIP3 is -7.4 dBm. Simulations predict an output $P_{1\,dB}$ of -1.1 dBm and IIP3 of -10.2 dBm. The noise figure of this amplifier has not been measured, but the simulations show a NF of 5.4 dB. This amplifier dissipates 24 mA from a 1.5-V supply.

The 60-GHz amplifier has also been characterized, and the S-parameters are shown in Fig. 18. The amplifier achieves a peak power gain of 12 dB, input and output return losses >15 dB, and the 3-dB bandwidth is 51–65 GHz. The measured reverse isolation is better than 45 dB up to 65 GHz. The measured output $P_{1\,dB}$ over frequency is given in Fig. 19. The frequency range is limited by the ability for the VNA to drive the amplifier into saturation. At 60 GHz, the measured output $P_{1\,dB}$ is $+2.0$ dBm, while simulations predict output $P_{1\,dB}$ of $+1.0$ dBm and IIP3 of -0.5 dBm. The output power of the 60-GHz amplifier is higher than the 40-GHz amplifier primarily because of the extra bias current. If the efficiency is kept constant, increasing the current by 50% results in a 3.5 dB increase in output power.

The measured NF of the 60-GHz amplifier is shown in Fig. 20. No de-embedding of any on-chip losses was performed since the pads and input match were part of the design. The NF is 8.8 dB at 60 GHz and remains below 9.3 dB up to 63 GHz. This is higher than the 6.9 dB predicted by simulations because the default BSIM3 noise model that was used does not properly account for effects such as excess short-channel thermal noise or induced gate noise. A more advanced RF transistor model, such as BSIM4 or MOS11, would be required to accurately predict the noise performance. This amplifier dissipates 36 mA from a 1.5-V supply. A summary of the measured results is provided in Table I along with a comparison to state-of-the-art bulk-CMOS amplifiers.

X. Conclusion

A mm-wave design and modeling methodology for CMOS devices and circuits has been presented. Optimization of the active and passive components for mm-wave circuits leads to transistors with finger widths $<1\ \mu$m and the use of CPW transmission lines. A modeling methodology, using relatively simple models that carefully account for the resistive losses, has been described and applied to transistors and transmission lines. The models have been verified to achieve broadband accuracy from dc to 65 GHz. Finally, two wideband mm-wave CMOS amplifiers, operating at 40 GHz and 60 GHz, were fabricated using a 130-nm bulk-CMOS technology. The measured S-parameters correspond extremely well to the simulated models owing to the accurate device modeling. The 40-GHz amplifier achieves 19-dB gain, output $\mathrm{P_{1\,dB}} = -0.9$ dBm, and $\mathrm{IIP3} = -7.4$ dBm, while consuming 36 mW. The 60-GHz amplifier achieves 12-dB gain, output $\mathrm{P_{1\,dB}} = +2.0$ dBm, and $\mathrm{NF} = 8.8$ dB, while consuming 54 mW. These are the first demonstrated amplifiers operating above 30 GHz fabricated using a mainstream bulk-CMOS technology. These amplifiers greatly improve on the current state-of-the-art, owing to the mm-wave CMOS design and modeling approach described in this work.

Acknowledgment

The authors acknowledge STMicroelectronics and B. Gupta for chip fabrication. The authors thank A. Cognata and T. Shirley of Agilent Technologies, Santa Rosa, for help with two-tone and distortion measurements, and H. Tran and K. Fujii of Agilent Technologies, San Jose, for help with noise figure measurements. Useful discussions and feedback from G. Baldwin, A. Vladimirescu, D. Yee, I. O'Donnell, J. Vanderhaegen, J. Weldon, and the BWRC OGRE group were greatly appreciated.

References

[1] L. M. Franca-Neto, R. E. Bishop, and B. A. Bloechel, "64 GHz and 100 GHz VCO's in 90 nm CMOS using optimum pumping method," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2004, pp. 444–445.

[2] R.-C. Liu, H.-Y. Chang, C.-H. Wang, and H. Wang, "A 63 GHz VCO using a standard 0.25 μm CMOS process," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2004, pp. 446–447.

[3] M. Tiebout, H.-D. Wohlmuth, and W. Simbürger, "A 1 V 51 GHz fully-integrated VCO in 0.12 μm CMOS," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2002, pp. 238–239.

[4] K.-W. Yu, Y.-L. Lu, D.-C. Chang, V. Liang, and M. F. Chang, "K-band low-noise amplifiers using 0.18 μm CMOS technology," *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 3, pp. 106–108, Mar. 2004.

[5] L. M. Franca-Neto, B. A. Bloechel, and K. Soumyanath, "17 GHz and 24 GHz LNA designs based on extended-S-parameter with microstrip-on-die in 0.18 μm logic CMOS technology," in *Proc. Eur. Solid-State Circuits Conf.*, Sep. 2003, pp. 149–152.

[6] X. Guan and A. Hajimiri, "A 24 GHz CMOS front-end," in *Proc. Eur. Solid-State Circuits Conf.*, Sep. 2002, pp. 155–158.

[7] M. Madihian, H. Fujii, H. Yoshida, H. Suzuki, and T. Yamazaki, "A 1–10 GHz, 0.18 μm-CMOS chipset for multi-mode wireless applications," in *IEEE MTT-S Int. Microwave Symp. Dig.*, June 2001, pp. 1865–1868.

[8] I. Bahl and P. Bhartia, *Microwave Solid State Circuit Design*, 2nd ed. Hoboken, NJ: Wiley, 2003.

[9] R. Spence, *Linear Active Networks*, London: Wiley, 1970.

[10] C. Enz, "An MOS transistor model for RF IC design valid in all regions of operation," *IEEE Trans. Microw. Theory Tech.*, vol. 50, no. 1, pp. 342–359, Jan. 2002.

[11] S. Emami, C. H. Doan, A. M. Niknejad, and R. W. Brodersen, "Large-signal millimeter-wave CMOS modeling with BSIM3," in *IEEE RFIC Symp. Dig.*, Jun. 2004, pp. 163–166.

[12] M. C. A. M. Koolen, J. A. M. Geelen, and M. P. J. G. Versleijen, "An improved de-embedding technique for on-wafer high-frequency characterization," in *Proc. IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, Sep. 1991, pp. 188–191.

[13] C. P. Yue and S. S. Wong, "On-chip spiral inductors with patterned ground shields for Si-based RF IC's," *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 743–752, May 1998.

[14] F. Ellinger, "26–42 GHz SOI CMOS low noise amplifier," *IEEE J. Solid-State Circuits*, vol. 39, no. 3, pp. 522–528, Mar. 2004.

[15] S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, 3rd ed. New York: Wiley, 1994.

[16] T. C. Edwards and M. B. Steer, *Foundations of Interconnect and Microstrip Design*, 3rd ed. New York: Wiley, 2000.

[17] B. Kleveland, C. H. Diaz, D. Vook, L. Madden, T. H. Lee, and S. S. Wong, "Exploiting CMOS reverse interconnect scaling in multigigahertz amplifier and oscillator design," *IEEE J. Solid-State Circuits*, vol. 36, no. 10, pp. 1480–1488, Oct. 2001.

[18] G. Carchon, W. De Raedt, and B. Nauwelaers, "Novel approach for a design-oriented measurement-based fully scalable coplanar waveguide transmission line model," in *IEE Proc. Microwave, Antennas Propagat.*, vol. 148, Aug. 2001, pp. 227–232.

[19] Agilent IC-CAP 2002 User's Guide [Online]. Available: http://eesof.tm.agilent.com

[20] B. Hughes, A. Ferrero, and A. Cognata, "Accurate on-wafer power and harmonic measurements of mm-wave amplifiers and devices," in *IEEE MTT-S Int. Microwave Symp. Dig.*, Jun. 1992, pp. 1019–1022.

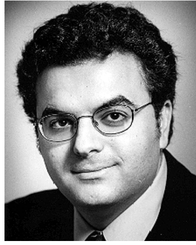
Chinh H. Doan (S'99) received the B.S. degree from the California Institute of Technology, Pasadena, in 1997, and the M.S. degree from the University of California, Berkeley, in 2000, both in electrical engineering. He is currently working toward the Ph.D. degree at the University of California, Berkeley, where he is a member of the Berkeley Wireless Research Center.

He held an internship position at Agilent Technologies, in 2002, where he designed and fabricated broadband monolithic microwave integrated circuit (MMIC) power amplifiers for test and measurement equipment operating up to 50 GHz using GaAs pHEMT technology. His current area of research is focused on the design of millimeter-wave transceiver circuits on CMOS, active and passive device modeling, and de-embedding and verification techniques for on-wafer probing on lossy substrates.

Sohrab Emami received the B.S. degree from the Sharif University of Technology, Tehran, Iran, in 1995, and the M.S. degree from the Tokyo Institute of Technology, in 2001, both in electrical engineering. He is currently working toward the Ph.D. degree at the University of California, Berkeley, where he is a member of the Berkeley Wireless Research Center.

His current research is focused on the design, modeling, and analysis of CMOS frequency-converting blocks such as mixers, frequency multipliers, and dividers, for 60-GHz wireless systems. From 1995 to 1998, he co-founded KARA-Systems Inc. working on control and instrumentation in the Iranian oil and natural gas industries. From 2000 to 2001, he held a part-time position at Hitachi Ltd., Tokyo, Japan, designing low-power CMOS baseband filters for direct-conversion wideband-CDMA receivers. In 2003, he was awarded the Hitachi-UC Berkeley Management of Technology Fellowship, where his responsibilities included the design and implementation of a multi-antenna prototype reader for the Hitachi MuChip RFID.

Mr. Emami also participated in the 21st International Physics Olympiad, Netherlands, and was the Gold Medal winner of the National Physics Olympiad.



Ali M. Niknejad (S'93–M'00) received the B.S.E.E. degree from the University of California, Los Angeles, in 1994, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1997 and 2000, respectively.

From 2000 to 2002, he was with Silicon Laboratories, Austin, TX, where he was involved with the design and research of CMOS RF integrated circuits and devices for wireless communication applications. Presently, he is an Assistant Professor with the Electrical Engineering and Computer Science Department, University of California, Berkeley. He is an active member at the Berkeley Wireless Research Center (BWRC) and he is the Codirector of the BSIM Research Group. His current research interests lie within the area of analog integrated circuits, particularly as applied to wireless and broadband communication circuits. His interests also include device modeling and numerical techniques in electromagnetics.

Dr. Niknejad is currently serving as an Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS.

Robert W. Brodersen (M'76–SM'81–F'82) received the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge, in 1972.

He was then with the Central Research Laboratory at Texas Instruments for three years. Following that, he joined the Electrical Engineering and Computer Science faculty of the University of California at Berkeley, where he is now the John Whinnery Chair Professor and Co-Scientific Director of the Berkeley Wireless Research Center. His research is focused in the areas of low-power design and wireless communications and the CAD tools necessary to support these activities.

Prof. Brodersen has won best paper awards for a number of journal and conference papers in the areas of integrated circuit design, CAD, and communications, including in 1979 the W.G. Baker Award. In 1983, he was co-recipient of the IEEE Morris Liebmann Award. In 1986, he received the Technical Achievement Awards in the IEEE Circuits and Systems Society and in 1991 from the Signal Processing Society. In 1988, he was elected to be member of the National Academy of Engineering. In 1996, he received the IEEE Solid-State Circuits Society Award and in 1999 received an honorary doctorate from the University of Lund in Sweden. In 2000, he received a Millennium Award from the Circuits and Systems Society, the Golden Jubilee Award from the IEEE. In 2001 he was awarded the Lewis Winner Award for outstanding paper at the IEEE International Solid-State Circuits Conference and in 2003 was given an award for being one of the top ten contributors over the 50 years of that conference.