Chapter 3
Design and Modeling of Active and Passive Devices

Ali M. Niknejad, Sohrab Emami, Chinh Doan, Babak Heydari, Mounir Bohsali

3.1 Passive Devices

3.1.1 Transmission Lines

Transmission lines play a critical role at mm-wave frequencies. Due to the relatively small wavelength, significantly long structures such as quarter wave can be realized on-chip. Transmission lines are suitable for high frequencies since there is no ambiguity in how one defines reference planes – since the signal and ground are always co-located, it's easy to connect a transmission line structure at any point in the circuit and predict the resulting reactance. Furthermore, the close physical proximity of the ground return current creates a dipole (multipole) radiation pattern that couples less energy to the substrate, which improves the quality factor of these devices. The well-defined ground return path also significantly reduces magnetic and electric field coupling to adjacent structures[1].

At mm-wave 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, transmission 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 transmission lines.

Four real (or two complex) parameters are needed to completely capture the properties of any quasi-TEM transmission line at a given frequency, ω_0. A transmission line can be characterized by its equivalent frequency-dependent $RLGC$ distributed circuit model (Fig. 3.1), which can be related to the characteristic impedance (Z_0) and complex propagation constant (γ) by [2]

Ali M. Niknejad, Sohrab Emami, Chinh Doan, Babak Heydari, Mounir Bohsali:
Berkeley Wireless Research Center, University of California, Berkeley

[1] Portions of this text are taken from [3], (© IEEE 2005)

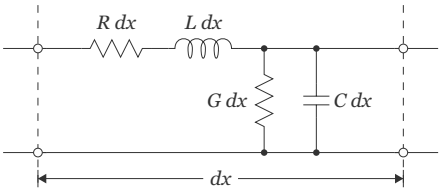


Fig. 3.1 Distributed RLGC lossy transmission line model [3] (© IEEE 2005).

$$Z_0 = \sqrt{\frac{R + j\omega_0 L}{G + j\omega_0 C}} \tag{3.1}$$

$$\gamma = \sqrt{(R + j\omega_0 L)(G + j\omega_0 C)} = \alpha + j\beta \tag{3.2}$$

For low-loss lines, the attenuation and phase constants, α and β, can be approximated as

$$\alpha \approx \frac{R}{2Z_0} + \frac{GZ_0}{2} \tag{3.3}$$

$$\beta = \frac{2\pi}{\lambda_g} = \omega_0\sqrt{LC} \tag{3.4}$$

There is a major deficiency with the standard representation of Eq. 3.1 and 3.2. As seen in Eq. 3.3, the two different loss mechanisms are combined into one parameter , making it difficult to discern the relative importance of R and G. This is particularly relevant for transmission lines implemented on low-resistivity silicon, since they have a non-negligible G. To address this issue, the following four real parameters are proposed to characterize the line [3]

$$Z = \sqrt{L/C} \tag{3.5}$$

$$\lambda = \frac{2\pi}{\omega_0\sqrt{LC}} \tag{3.6}$$

$$Q_L = \omega_0 L/R \tag{3.7}$$

$$Q_C = \omega_0 C/G \tag{3.8}$$

Notice that the two loss mechanisms are completely decoupled in Eq. 3.7 and 3.8. The first-order Taylor series expansions can be used to relate the characteristic impedance in Eq. 3.1 to the quantities in Eq. 3.5 - 3.8.

$$Z_0 \approx Z\left(1 + \frac{j}{2}\left(\frac{1}{Q_C} - \frac{1}{Q_L}\right)\right) \tag{3.9}$$

From Eq. 3.9, the sign of $\Im(Z_0)$ reveals which loss mechanism (Q_L or Q_C) is dominant for low-loss lines. If Z_0 is assumed to be real, this implicitly requires that $Q_L = Q_C$, which is not generally true.

Unlike transmission lines implemented on GaAs, where G is essentially zero, transmission lines implemented on low-resistivity silicon often have low capacitive quality-factors (Q_C) due to the substrate coupling. For transmission 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.

3.1.1.1 Quality Factor Metric

Analogous to the case for lumped inductors [4], there are many different definitions for Q of a transmission line, where the applicability depends upon the intended function of the transmission line circuit. The most commonly used definition is the Q of the line when used as a resonator (Q_{res}),

$$Q_{res} = \omega_0 \frac{\text{avg. energy stored}}{\text{avg. power loss}} = \frac{\omega_0(W_m + W_e)}{P_L} \tag{3.10}$$

where ω_0 is the resonance frequency, W_m and W_e are the average magnetic and electric energy stored, and P_L is the average power dissipated in the line. Q_{res} can be related to the quantities introduced in Eq. 3.3 and 3.4 by [5]

$$Q_{res} \approx \frac{\beta}{2\alpha} = \frac{\pi}{\alpha\lambda_g} = \frac{\pi\sqrt{\epsilon_{eff}}}{\alpha\lambda_0} \tag{3.11}$$

where λ_0 is the free-space wavelength and ϵ_{eff} is the effective dielectric constant. Since ϵ_{eff} is determined mostly by the dielectric properties, and not the transmission line structure or dimensions, maximizing Q_{res} is roughly equivalent to minimizing the attenuation constant α. Using Eq. 3.11 along with Eq. 3.3 and 3.4, Q_{res} can be related very simply to the parameters in Eq. 3.7 and 3.8 by

$$\frac{1}{Q_{res}} \approx \frac{1}{Q_L} + \frac{1}{Q_C} \tag{3.12}$$

If the transmission line is forming a resonator, this is indeed a good definition to use. The Q_{res} quality factor is related to the half-power bandwidth of a quarter/half wave resonator. This is in direct analogy with lumped parallel/series resonant circuit, which is a good model for the resonator near resonance. But if the transmission line is employed for matching or interconnect, a different defintion or metric is needed.

Another common metric is to simply compare the propagation loss versus frequency. While this is a fast and simple way to look at a transmission line, it ignores the fact that different lines can have a different rate of phase accumulation (such as slow wave structure) and a different characteristic impedance, which is important

in the design of matching networks. To first degree, then, the propagation constant should be compared versus β, rather than to frequency. Since β is proportional to frequency

$$\beta = \frac{2\pi}{\lambda_0} = \frac{\omega_0}{v} \tag{3.13}$$

the plot of α versus ω has the same shape with a frequency scaling factor related to the phase propagation velocity $v \leq c_0$. This definition is useful in cases where we desire a certain phase shift through a transmission line, and we are comparing two different transmission lines to see which one is more suitable.

When transmission lines are employed as inductors, or equivalently if $\ell \ll \lambda_0$, then we can approximate the impedance of a shorted section of a short transmission line

$$Z_i = Z_0 \tanh(\gamma\ell) \approx \gamma\ell Z_0 = (\alpha + j\beta)\ell Z_0 \tag{3.14}$$

which shows the line impedance is proportional to the characteristic impedance Z_0. Assuming low-loss conditions so that $Z_0 \approx Z$ is real, then we have

$$Q = \frac{\Im(Z_i)}{\Re(Z_i)} \approx \frac{\beta}{\alpha} \tag{3.15}$$

For lumped circuits, a common metric is to compare the product of Q factor and the inductance value, since it's increasingly difficult to realize large reactance with high Q

$$Q \times L \approx \frac{\beta}{\alpha}\frac{\ell\beta Z}{\omega} \tag{3.16}$$

since β is proportional to frequency, we have

$$Q \times L \approx \frac{\beta}{\alpha}\frac{\ell}{v}Z = \frac{\beta}{\alpha}t_d Z \tag{3.17}$$

which shows that in addition to a large value of β/α, we also desire a transmission line with a large value of Z and a slow phase velocity, or equivalently a larger inductance per unit length. This argument can be further refined by defining more precisely an inductance quality factor.

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 [4]. Thus, for matching networks, the most meaningful metric is

$$Q = 2\omega_0 \frac{\text{net energy stored}}{\text{avg. power loss}} = \frac{2\omega_0(W_m - W_e)}{P_R + P_G} \tag{3.18}$$

where W_m and W_e are the average magnetic and electric energy stored, and P_R represent the "series" resistive losses and P_G the "shunt" conductive losses. Defining the series inductive and shunt capacitive Q we have [7]

$$Q_L = 2\omega_0 \frac{W_m}{P_R} \tag{3.19}$$

$$Q_C = 2\omega_0 \frac{W_e}{P_G} \tag{3.20}$$

we can express the overall Q as

$$\frac{1}{Q} = \frac{1}{\eta_L Q_L} + \frac{1}{\eta_C Q_C} \tag{3.21}$$

where

$$\eta_L = 1 - \frac{W_e}{W_m} \tag{3.22}$$

and

$$\eta_C = \frac{W_m}{W_e} - 1 \tag{3.23}$$

For a shorted transmission line, under the assumption of low loss, one can show that

$$W_m \approx \frac{1}{2} \frac{LV^{+^2}\ell}{Z_0^2} \left(1 + \text{sinc}\left(\frac{4\pi\ell}{\lambda}\right)\right) \tag{3.24}$$

and

$$W_e \approx \frac{1}{2} CV^{+^2}\ell \left(1 - \text{sinc}\left(\frac{4\pi\ell}{\lambda}\right)\right) \tag{3.25}$$

Thus we have

$$\frac{1}{\eta_L} = \frac{1}{2\,\text{sinc}(\frac{4\pi\ell}{\lambda})} + \frac{1}{2} \tag{3.26}$$

and

$$\frac{1}{\eta_C} = \frac{1}{2\,\text{sinc}(\frac{4\pi\ell}{\lambda})} - \frac{1}{2} \tag{3.27}$$

If the line is inductive (i.e., $W_m \gg W_e$), then $\eta_C \gg \eta_L$ and $Q_{net} \approx \eta_L Q_L$. The loss in the line is therefore almost completely determined by Q_L. For example, consider a shorted transmission line with $\ell < 0.1\lambda$. 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 transmission lines on silicon, where the low-resistivity substrate causes Q_C to be non-negligible.

3.1.1.2 Microstrip and Coplanar Waveguide

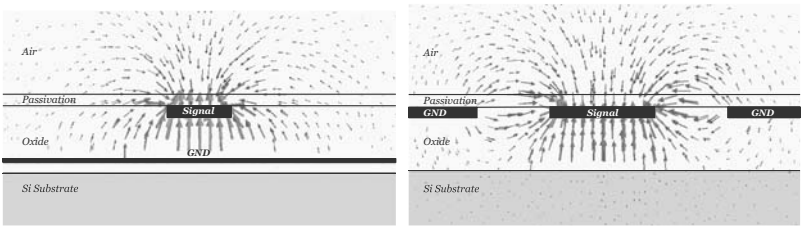


Fig. 3.2 Electric field distributions from 3-D EM simulations of (a) microstrip and (b) coplanar waveguide transmission lines [3] (© IEEE 2005).

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. 3.2a 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. 3.3b). The biggest drawback to microstrip lines on standard CMOS is the close proximity of the ground plane to the signal line ($\sim 4\,\mu\mathrm{m}$), yielding very small distributed inductance, L. This significantly degrades the inductive quality factor, Q_L (Fig. 3.3a).

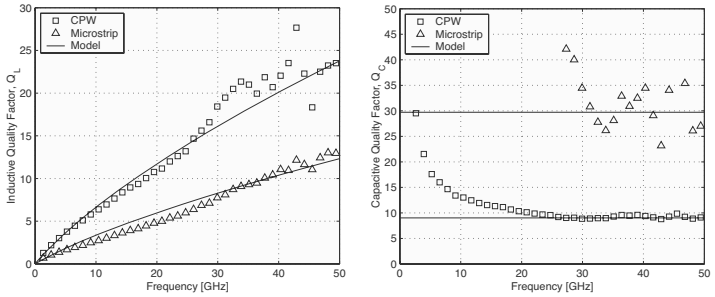


Fig. 3.3 Measured (markers) Q_L and Q_C for a coplanar waveguide and a microstrip line. The solid lines are for an empirical first-order model where the R variation is due only to skin effect and G is caused by a constant dielectric loss tangent [3] (© IEEE 2005).

Another option for on-chip transmission lines is the use of coplanar waveguides (CPWs) [10], [13], which are implemented with one signal line surrounded by two adjacent grounds (Fig. 3.2b). 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 $W = 10\,\mu\text{m}$ and $S = 7\,\mu\text{m}$ has a Z_0 of $59\,\Omega$ and a Q_L measured to be about double that of the microstrip (Fig. 3.3a). Therefore, CPW transmission lines are preferred in design of amplifiers (tuning and matching networks) due to their considerably higher Q_L compared to microstrip lines.

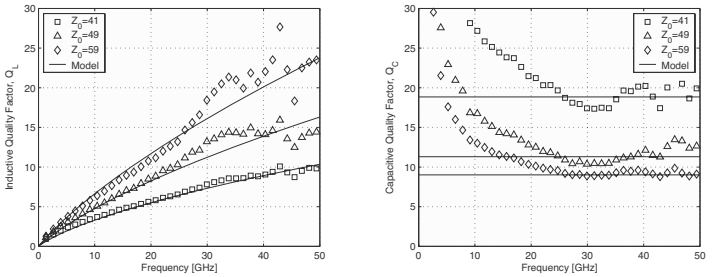


Fig. 3.4 Measured (markers) Q_L and Q_C for coplanar waveguides with varying geometries. The solid lines are for an empirical first-order model where the R variation is due only to skin effect and G is caused by a constant dielectric loss tangent [3] (© IEEE 2005).

By varying the signal-to-ground spacing, it is possible to design CPW lines to have either large Q_L and high-impedance ($S = 7\,\mu\text{m}$) or large Q_C and low-impedance ($S = 2\,\mu\text{m}$) (Fig. 3.4). On the other hand microstrip lines have, to first-order, constant Q_L and Q_C regardless of geometry. The larger the gap spacing S, the more fields penetrate into the substrate and cause additional shunt losses which lower Q_C. To first order, though, the series losses which contribute to Q_L remain relatively constant since these losses are dominated by the resistivity of the signal and ground plane. Second order effects, such as eddy current induced losses, change since the proximity of the ground plane determines the magnetic vector potential in the substrate, and hence the induced current magnitude. In a microstrip, to change the Z_0, thinner lines must be used, which increases the series losses substantially.

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 [10]. 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.

3.1.1.3 Transmission Line Modeling

Transmission lines are essentially two-dimensional structures, and TEM fields can in theory be predicted from static solutions in the transverse plane. In practice, though, intentional non-uniformity in the structure is introduced to dampen the presence of higher order TE/TM modes. Furthermore, the dielectric is non-uniform with different

low-K dielectric inter-layers in the typical stack-up. The abundance of metal layers also allows much more complicated structures to be realized, and this begs the question of the optimal structure that one should realize. For these reasons, prediction of performance requires extensive 3D simulation.

To achieve the highest level of accuracy for the transmission line models, a design-oriented modeling methodology, similar to [14], 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 Fig. 3.3, 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 transmission line models assume no coupling to adjacent structures. This assumption is justified since well-defined ground return path 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.

Full-wave electromagnetic simulation of transmission lines is useful for the design of new structures, especially when the line properties are not uniform with length. This occurs, for instance, when dummy fills are incorporated into the air structure or if regular bridging layers are used. Slow wave structures also employ non-uniform filling with filaments. In these cases full-wave 3D simulation is very time consuming. Accurate simulations are also difficult to obtain due to the high frequency material properties which are not typically characterized by the foundry. This includes dielectric loss of the various layers of dielectrics used in a modern CMOS process, the metal resistivity and roughness, and the precise substrate profile. The designer should strive to control the layout as much as possible by placing dummy fill blocking layers around the transmission lines and by carefully controlling the layout where the electromagnetic fields have an appreciable magnitude. Attempts to simulate these structures using commercial tools directly using only knowledge of the substrate conductivity and assuming lossless dielectrics has been unsuccessful. In order to match measurements and simulations, a custom tailored substrate profile with various dielectric loss parameters is needed in order to fit a wide array of measured data. Given a predictive simulation setup allows much greater flexibility in the design of mm-wave passive structures.

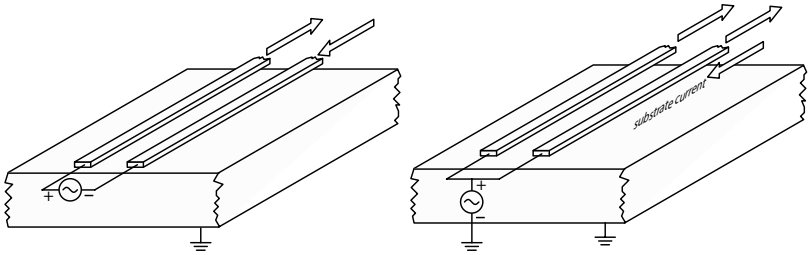


Fig. 3.5 A differential transmission line excited (a) in odd mode and (b) in even mode. Arrows indicate direction of current flow.

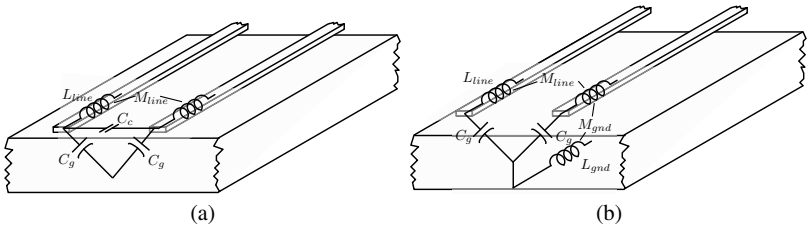


Fig. 3.6 Distributed circuit models for the (a) odd and (b) even modes of propagation.

3.1.1.4 Differential Lines

Since differential circuits are common in ICs, differential transmission lines are very convenient structures. As shown in Fig. 3.5, a differential pair can excite equal and opposite currents on a differential two-wire transmission line. Clearly there are two modes of propagation on the line, the “odd" mode which is usually the desired mode, where the voltage on each line is out of phase with the other voltage, and the “even" mode, where the voltages have equal magnitude. In this case the voltage is excited with reference to the substrate potential and the return current path flows through the substrate. We may therefore think of the even mode as a microstrip occurring primarily between the two signal lines (effectively connected in parallel) and the substrate ground. The two modes have different characteristic impedance and different phase velocity. The characteristic impedance of the odd mode is given by analyzing the equivalent half circuit shown in Fig. 3.6a

$$Z_{oo} = \sqrt{\frac{L_o}{2C_c + C_g}} \tag{3.28}$$

where L_o is the inductance per unit length defined by integrating the magnetic field between the lines, and C_c and C_g are the ground and mutual capacitance per unit

length. The odd mode inductance can also be obtained from the partial inductance of each line (assumed identical) by including the mutual inductance between the lines, $L_o = L_{line} - M_{lines}$, which subtracts due to the opposite direction of current flow. We are implicitly assuming that no current flows in the ground plane for this configuration due to the balanced excitation. Even though induced currents flow in the ground, the net ground current is zero. The capacitance C_c is doubled due to Miller effect. For the even mode, the return current flows in the substrate so the inductance per unit length L_e is calculated by integrating in the vertical plane from the signal line to the substrate. Since the two lines are at the same potential, the capacitance C_c has no effect and the equivalent half circuit can be calculated as shown in Fig. 3.6b. The characteristic impedance is given by

$$Z_{oe} = \sqrt{\frac{L_e}{C_g}} \tag{3.29}$$

where L_e is the even mode inductance per unit length or calculated through the partial inductance terms by $L_e = L_{line} + M_{line} + L_{gnd} - M_{gnd}$. We see that if $L_e \sim L_o$, the common mode impedance is higher. In fact, we implicitly assumed that the return current flows through the substrate, which is not necessarily the case. To understand this point, imagine that there is an alternative path to ground through metallization between the source and load, such as other ground planes, interconnect, or the substrate. Since current flows the path of least impedance, at low frequencies the current may flow through the distant return path increasing the even mode inductance substantially, or $L_e \gg L_o$, which makes the even mode impedance large. At high frequencies, though, the current will flow underneath the structure – even though the path is more resistive, it has lower inductance. This makes the even mode very lossy as well, which is reflected in a higher even mode propagation loss $\alpha_e > \alpha_o$.

The phase velocity of the two modes can differ due to the different field configurations. In the odd mode the fields reside substantially between the conductors with very little leakage into the substrate and air, which results in $\epsilon_o \approx \epsilon_{oxide}$. On the other hand, for the even mode, there is significant leakage into the substrate, which decreases the phase velocity (since $\epsilon_{\rm Si} \approx 3\epsilon_{\rm ox}$).

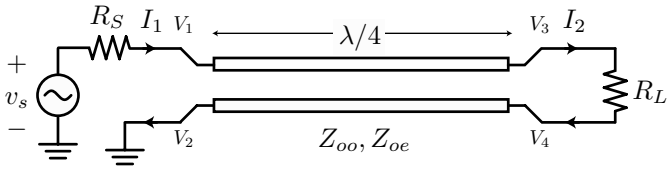


Fig. 3.7 A transmission line converts an unbalanced signal to a balanced signal.

One particularly good application for a differential transmission line is to convert a single-ended signal into a differential signal. The circuit shown in Fig. 3.7, is known as a transmission line . Nominally the circuit should be a quarter wavelength.

Intuitively we can see that the single ended port excites an equal component of the even and odd modes. To see this suppose that $V_e = V_o = V_0/2$, then

$$V_1 = (V_e + V_o) = V_0 \tag{3.30}$$

and

$$V_2 = V_e - V_o = 0 \tag{3.31}$$

At the load end of the line, we wish to have a fully balanced signal, or $V_e(\ell) = 0$. A balanced load is an open circuit boundary condition for the even mode, which means the line is highly imbalanced. Since the line is excited with a common mode voltage at the source end, a standing wave pattern emerges on the line due to the open circuit reflection. Since the line is exactly $\lambda/4$, the voltage profile on the line is monotonic. Given that the source is driven with a non-zero common mode, the common mode voltage must decay to zero at the load, resulting in a perfect transformer. To propagate all the energy of the source into the the odd mode, we should terminate the line in the characteristic impedance of the odd mode $Z_L = Z_{oo}$ to avoid reflections.

3.1.1.5 Slow Wave Structures

One way to avoid substrate leakage in a transmission line is to place a solid shield under the conductors. For the differential mode, the return current flows through the the conductors, and not the ground, but due to the changing magnetic field, eddy currents are induced on the ground plane which lowers the inductance per unit length. The net current in the ground plane is zero since the induced currents have opposite direction, and in essence we see the ground current flows in a loop (typical of "edddy currents"). Commensurate with the decrease in inductance, the capacitance per unit length also increases due to the increase in C_g, and in fact this increase must exactly balance the decrease in inductance to maintain constant phase velocity.

If we break the shield in such a way as to limit current flow along the length of the line, then we see that the inductance per unit length remains intact while the capacitance per unit length increases. This is a "slow wave" transmission line, since the phase velocity is lowered. One way to accomplish a slow wave structure shown in Fig. 3.8. Here, conductive strips are placed beneath the two wires to increase the capacitance per unit length without altering the inductance per unit length. In a modern CMOS process, a very high density of filaments can be used to increase the capacitance significantly. As a result, the quality factor of a resonator improves since the dielectric losses are minimized in the transmission line. Researchers have demonstrated quality factors as high as 40 for such integrated resonators in CMOS [28].

Another slow wave structure is a varactor loaded line, where MOS capacitors are placed periodically along the line. The transmission line characteristic impedance and phase velocity are now a function of the bias voltage, which has application in phase shifters or dynamic tuning elements.

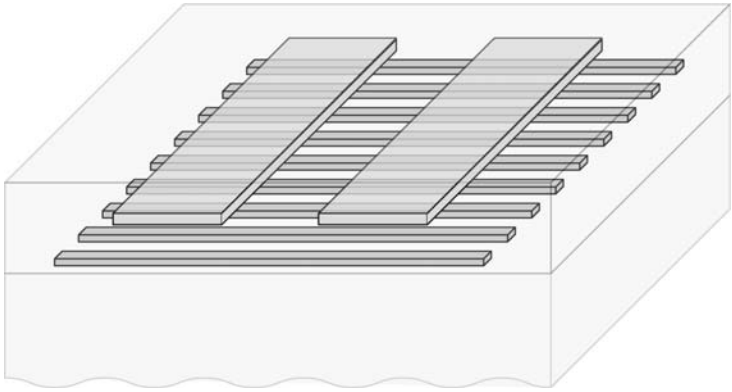


Fig. 3.8 A balanced slow wave two-conductor transmission line.

3.1.2 Inductors

Inductors are used extensively to tune out parasitics and for matching at RF frequencies and take the form of spiral inductors or short sections of transmission lines. While lumped inductors are generally smaller than distributed circuit equivalents, the advantages of distributed circuit elements include better prediction, scalability, and improved isolation. The value of the inductance of a shorted transmission line depends only on the length, Z_0, and propagation constant $\gamma = \alpha + j\beta$. If a transmission line is well characterized, then any arbitrary inductance can be synthesized by varying the length of the line. Furthermore, there are no "leads" that add parasitic inductance. A common example is the inductance degeneration employed in amplifiers, which places an inductor from the source to ground. The value of the degeneration inductance is ambiguous, though, as the terminals of the inductor are widely separated. The "return path" flows through ground and through the gate (or base) elements, forming a series resonant circuit. The proximity of the ground and the placement of the gate (base) inductors therefore have an appreciable impact in the value of inductance, requiring careful analysis. Since ground return currents flow intrinsically in the transmission line, there is no ambiguity as to how currents flow in the structure and the proximity of the ground plane minimizes induced currents in the substrate.

In many situations, though, a lumped inductor can be realized with higher quality factor. For typically small inductors with $L \sim 100\text{pH}$, simple ring inductors, as shown in the inset of Fig. 3.10, are sufficient. In fact, we can draw an analogy between a ring inductor and a differential transmission line. To realize high Z_0 transmission line, we increase the gap spacing to increase the inductance per unit length while reducing the odd mode capacitance per unit length. But in order to short the line with low inductance, we prefer to bend the end of the line and reduce the gap spacing. Furthermore, to present close leads to connect to a capacitor or transistor, we prefer

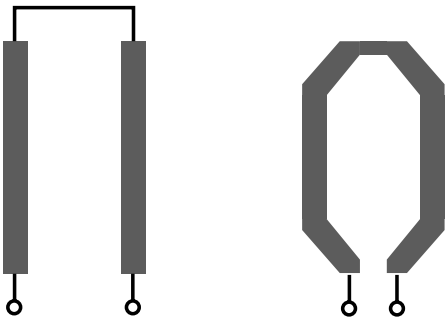


Fig. 3.9 (a) A high Z_0 line is formed by using a large gap spacing in the differential line. (b) A differential transmission takes the form of a loop when practical leads are formed.

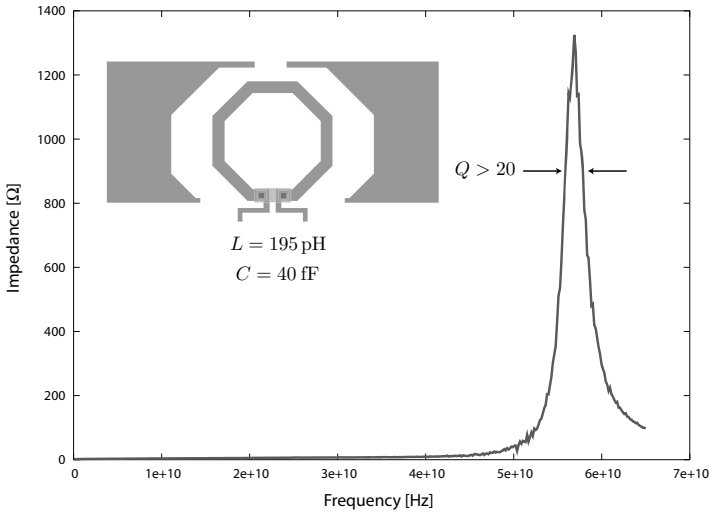


Fig. 3.10 Measured resonance of a mm-wave tank formed by a ring inductor and MIM capacitors.

to also bend the front of the line, as shown in Fig. 3.9b. Clearly this structure very much resembles a ring inductor, showing a close connection between transmission lines and ring inductors.

A ring inductor has been designed and fabricated in a Si RF process. To accurately measure the Q factor, a parallel resonant circuit is formed with MIM capacitors. In this way, the Q factor is insensitive to small lead inductance and resistance since at resonance the circuit forms a "open" or large impedance circuit. The measured resonance curve is shown in Fig. 3.10. The resonator occupies an area of $150^2 \mu\mathrm{m}^2$ and has a loaded measured quality factor over 20. The inductor has been simulated and measured separately and the Q factor is between 30-40, which shows that the MIM capacitors have comparable Q. An electric shield surrounds the structure to minimize the coupling to Si substrate. This "open" shield structure is used to minimize eddy current flow.

3.1.3 Capacitors

In mm-wave circuits, capacitors in matching networks and resonators are usually realized as transmission lines whereas lumped capacitors are employed for AC coupling and DC bypass. Because lumped capacitors reside in the signal path, optimizing their design and layout as well as modeling their behavior accurately over a broad frequency range is of a crucial importance. For both AC coupling and DC bypass, an ideal structure would have infinite impedance at DC and zero impedance at the frequency of interest. Thus, large high quality factor capacitors with self-resonance frequency situated above the frequency of interest are desirable. Even if the capacitor self-resonates below the frequency of interest, as long as its impedance at the frequency of interest is sufficiently low (even if inductive), it can still be used effectively for both AC coupling and DC bypass. If the impedance is not sufficiently low, then it becomes important to accurately model the behavior so that the parasitics are absorbed into the matching network.

3.1.3.1 Coupling Capacitors

"Finger" (comb) capacitors, sometimes referred to as "MOM" (metal-oxide-metal) capacitors, can be easily fabricated in any modern CMOS process due to the abundance and high packing density of the metal layers. Finger capacitors are composed of a large number of parallel fingers connected to either port of the device as shown in Fig. 3.11. Each finger consists of all available metal layers stacked on top of each other in order to decrease its resistance which increasing both its quality factor and the self-resonance frequency. The metal layers are connected together through the maximum number of vias possible in order to take advantage of the additional via-to-via capacitance. If a lower parasitic substrate capacitance is

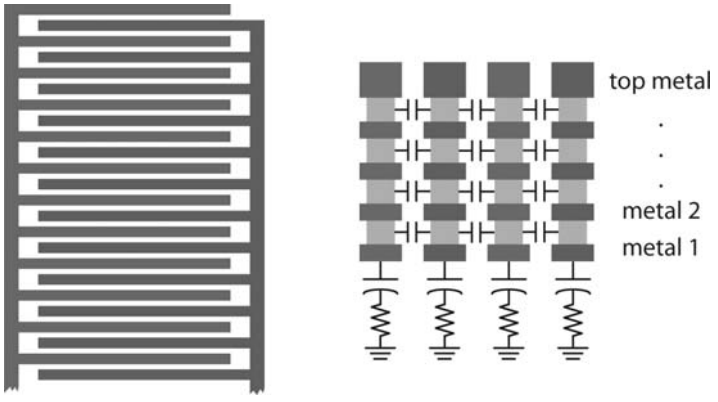


Fig. 3.11 Layout geometry of a "finger" or "comb" capacitor structure.

desired, the lowest metal layer can be omitted, but this negatively affects the quality factor of the structure. At each port, a wide slab consisting of a small 45° taper is used to connect the fingers. In order to realize a given capacitance, either a large number of short fingers in parallel or a small number of long fingers can be used. The former structure results in smaller finger resistance and inductance but larger lead parasitics due to the transition from the signal lines to the capacitor plates. The latter structure exhibits the opposite tradeoff. This tradeoff and the optimum geometry are dependant on the overall capacitor size.

Given that no new layers are needed in the fabrication of MOM capacitors, they are less expensive than Metal-Insulator-Metal (MIM) capacitors, which employ a sandwich structure with a thin high-K dielectric. If the density of MIM capacitors are sufficiently larger than"finger" structures, then the area savings may outweigh the additional fabrication costs.

3.1.3.2 Modeling

The symmetric capacitor equivalent circuit shown in Fig. 3.12 is used to model AC coupling capacitors over a broad frequency range (from DC to 100 GHz). R_1, C_1, and L_1 model the finger capacitor while C_2 models the oxide, and C_3 and R_2 model the substrate.

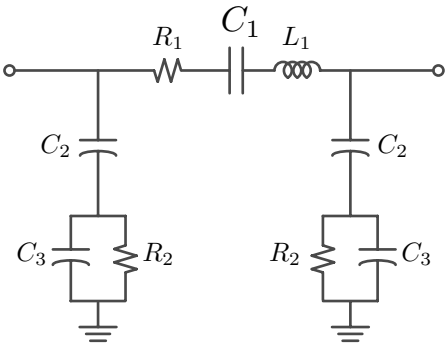


Fig. 3.12 Symmetric layout AC coupling capacitor model.

The circuit components values are optimized in order to match the model simulated network parameters to the measured network parameters. If only the S-parameters are used in the optimization, it is very difficult to extract the value of R_1. In an S-parameter measurements, the $50\,\Omega$ resistance is connected in series with each port, overshadowing R_1 which is usually in the order of hundreds of $\mathrm{m}\,\Omega$'s. For this reason, Y and Z parameters are derived from the measured S parameters and used simultaneously in the optimization. Because the branch composed of R_1, C_1, and L_1 has a much lower impedance than the branches composed of $C_{2,3}$ and R_2, the

former dominates the Y parameters and the latter dominate the Z parameters. Thus, a simultaneous Y and Z parameter optimization is used.

3.1.3.3 De-Coupling Capacitors

Very large capacitors (e.g. 2-5 pF) are desirable for DC bypass applications. However, scaling the AC coupling capacitors described above would result in considerably lower self-resonance frequencies that render the structure unusable as a DC decoupling capacitor. In a bypass capacitor, one port is always connected to ground (or supply) and therefore the entire bottom plate of the capacitor can be "shorted" and used as a low inductance path to ground. Having a low impedance ground plane considerably lowers the series inductance and thus increases the self-resonance frequency of the capacitor enabling large capacitors to be functional at mm-wave frequencies.

Other techniques in the layout of the MOM capacitor have appeared in the literature. The layout incorporating a mesh structure is shown in Fig. 3.13 uses fingers at right angles, resulting in both a fringe and overlap capacitance. The layers are interconnected multi-layer meshes that are interlocked together.

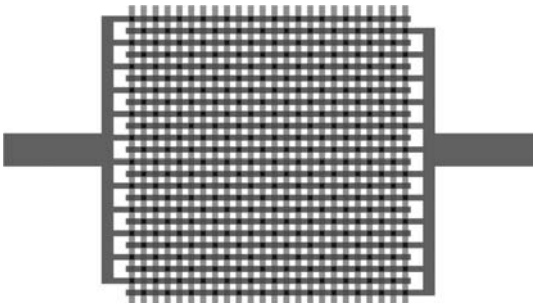


Fig. 3.13 The layout of a "mesh" capacitor structure.

3.1.3.4 Varactors

Varactors are key building blocks in many circuits, particularly in voltage-controlled oscillators. The variation in capacitance versus tuning voltage and the quality factor are important metrics. The quality factor of capacitors drops inversely with frequency due to any series resistance, and this is a severe limitation for on-chip structures. In Fig. 3.14 we plot the quality factor of a MOS varactor. Evidentally the Q factor is very low in the mm-wave frequencies. To first order the MOS varactor Q factor is independent of the device size since the channel resistance and capacitance scale with the W of the structure. In inversion, the maximum capacitance is given by the effective oxide thickness

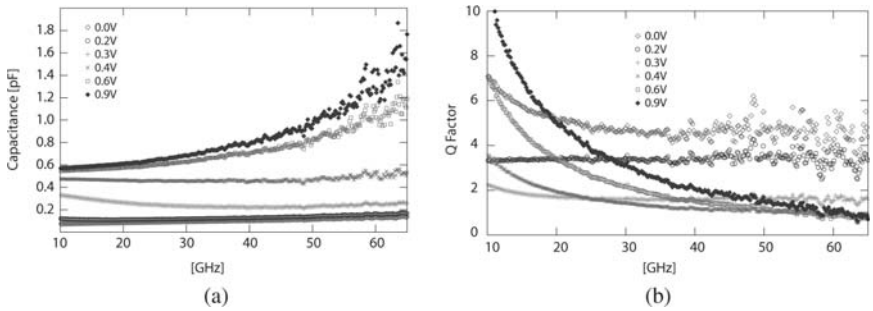


Fig. 3.14 Measured (a) capacitance and (b) quality factor of a MOS varactor ($N = 5 \times W = 40$) as a function of bias voltage.

$$C_v = W \cdot L \cdot C_{ox} \tag{3.32}$$

whereas the resistance is proportional to the gate and channel resistance. The channel conductance contribution is given by

$$G_{ch} \propto \frac{W}{L} \mu C_{ox} (V_{GS} - V_T) \tag{3.33}$$

so that the Q is independent of W and C_{ox}

$$Q = \frac{G_{ch}}{\omega C_v} \propto \frac{\mu (V_{GS} - V_T)}{L^2} \tag{3.34}$$

necessitating one to employ the shortest channel length and highest mobility possible. For this reason electrons are the preferred carriers in the channel of a MOS varactor. The low Q factor of mm-wave varactors is a major limitation since a poor quality VCO tank directly translates into poor phase noise performance unless tuning range is sacrificed. The difficulty in producing a wide tuning VCO in the mm-wave band favors transceiver architectures with lower frequency VCOs and frequency multipliers.

3.1.4 Transformers

Transformers have found many unique applications in RF CMOS and SiGe circuits. For mm-wave circuits, they can be used for driving balanced circuits, such as a Gilbert cell mixer, for AC coupling, and for impedance matching. Their compact size, the ability to AC couple and provide DC voltages, makes these devices very convenient in the design of mm-wave circuits.

Two coupled inductors, for instance two ring inductors shown in Fig. 3.15a, form a simple transformer. The measured insertion loss of the structure is shown in Fig. 3.15b. The structure exhibits about 1 dB of loss when the device is loaded

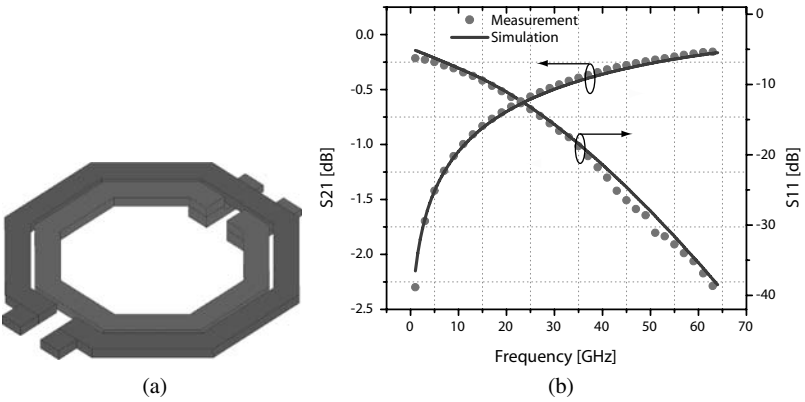


Fig. 3.15 (a) The layout of a simple 1:1 transformer. (b) The measured insertion loss and matching of the structure.

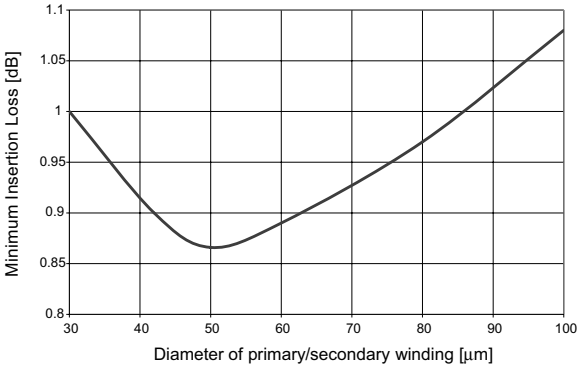


Fig. 3.16 The simulated insertion loss of a transformer as a function of outer diameter.

with the optimal source/load impedance. In many applications, the load and source capacitance can be absorbed into the transformer. More complicated geometries are possible for the realization of higher turn ratios [29]. Due to the high frequency of operation, the area of these structures must be scaled down to realize high self-resonant frequencies. For a given frequency of operation, one can determine the optimal size through EM simulation, as shown in Fig. 3.16, where the outer radius is swept to find the optimal radius. Clearly a larger inductor if beneficial since the winding resistance can be thought of as a shunt resistor to ground of value ωLQ, which should be as large as possible. A larger inductor has lower series resistance but higher substrate losses. There is thus an optimal radius that produces the best $Q \times L$ product for the windings. An example design will be highlighted in the next chapter, where the LO port of a Gilbert cell mixer is driven with an on-chip balun.

3.1.5 Resonators

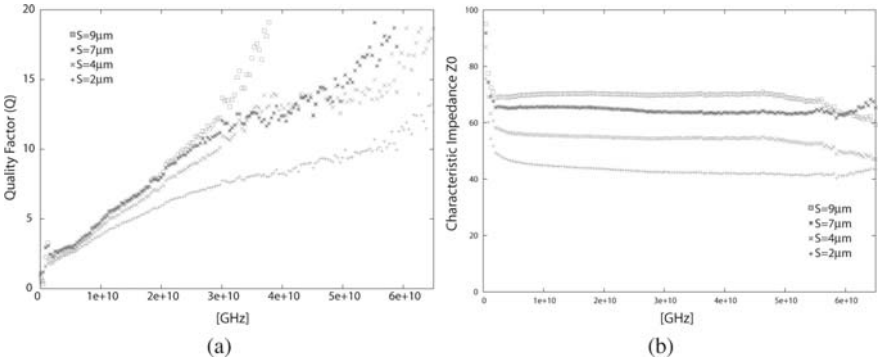


Fig. 3.17 The measured (a) quality factor $Q = \beta/2\alpha$ and (b) characteristic impedance of a CPW transmission line resonators with varying gap spacing: $S = 2\,\mu\text{m}, 4\,\mu\text{m}, 7\,\mu\text{m}, 9\,\mu\text{m}$.

Resonators are useful building blocks in filters and in oscillators. We have already demonstrated that simple LC tanks can be used as resonators in the mm-wave regime, and the behavior is closely captured by a lumped equivalent circut. As long as the structure is physically small as to keep the self-resonant frequency larger than the mm-wave band of interest, there is no problem in thinking of a small ring or spiral as a lumped inductor.

An alternative to the ring is a resonant transmission line. For instance a shorted section of transmission line will self-resonate at every odd multiple of the quarter wave frequency. As already noted, the quality factor of the line is given by $Q = \beta/2\alpha$. The Q of several CPW transmission line resonators is plotted in Fig. 3.17a based on the measured S-parameters of the transmission lines. The various electrical and magnetic loss mechanisms have already been discussed but it is noteworthy that at resonance the electric and magnetic field energy are equal, and so both contributions of loss are important. It is clear that using a wide gap spacing increases the Q, but the improvement in Q is limited since the electric field leaks into the substrate and introduces significant loss at higher frequencies. The characteristic impedance Z_0 is shown in Fig. 3.17b, where the Z_0 increases in step with the gap spacing as more magnetic energy is stored while the electrical energy storage drops. It is noteworthy again that the improvement in Z_0 is limited by the substrate, especially above 40 GHz. We often desire to alter the self-resonant frequency of a resonator, in particular in the design of voltage-controlled oscillators (VCOs). We may use a varactor loaded LC tank or even a slow wave varactor loaded transmission line. The tank inductor can be realized by a transmission line shorter than $\lambda/4$.

It is interesting to note that the optimal quarter wave line is not a uniform line [15], since the voltage/current profile and the corresponding electrical/magnetic energy

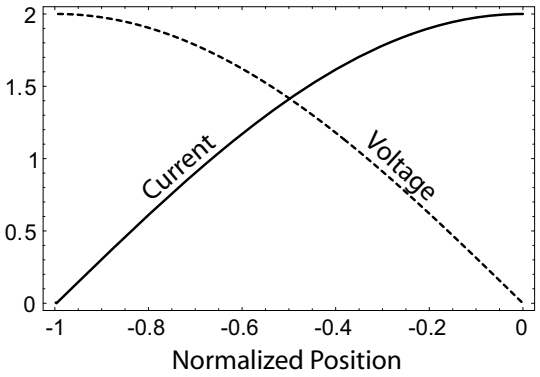


Fig. 3.18 The standing wave pattern for voltage and current along a quarter wave short-circuited line in resonance.

storage varies along the line. A differential stripline shorted at one end exhibits the standing wave mode shown in Fig. 3.18 at resonance. At the shorted end of the line, the voltage is at a minimum and the current at a maximum, so the losses at this point come only from the series resistance of the metal line. Conversely, at the driven end, the voltage is at a maximum and the current at a minimum, so the losses at this point come only from the shunt conductance between the differential lines. This phenomenon can be exploited to lower the losses of the resonator and thus raise the quality factor [15]. At the shorted end of the line, the metal conductors can be made wider to reduce the series resistance without having to worry about the increase in shunt conductance. The gap spacing can also be made wide since the voltage difference between the lines is small. Similarly, at the driven end of the line, the conductors can be made very narrow, lowering the conductance, without having to worry about the increase in series resistance. The gap spacing should be small in order to minimize field leakage into the substrate, since the fields are strongest at this end of the line. Along the rest of the line, shunt conductance and series resistance can be traded off to achieve a tapered line with much lower losses than the optimum uniform line. The optimum taper shape would thus end up looking similar to Fig. 3.19a. From this analysis we draw an important conclusion that the optimal resonator is in fact very similar to an LC circuit. To see this notice that circuit can be partitioned into a low Z_0 transmission line in series with a “loop" inductor, as shown in FIg. 3.19b. It’s clear that the low Z_o section is mainly contributing capacitance to the tank and can be replaced by a more energy efficient MIM or MOM structure described earlier. With this substitution, we clearly identify this circuit as a an optimal LC tank.

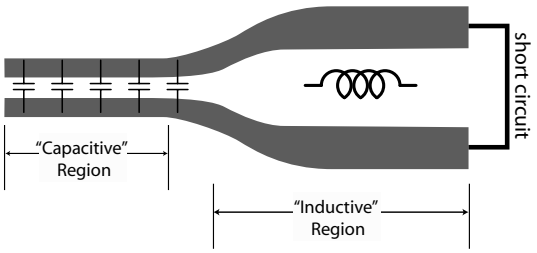


Fig. 3.19 The optimal profile for a transmission line tapers from a low Z_0 capacitive section to a high Z_0 inductive region. This taper profile resembles a lumped LC tank.

3.2 Active Devices

3.2.1 Modeling

Circuit designers typically use "compact" models for design and verification of Si integrated circuits. Compact models are the interface between the technology and the design. A circuit designer learns about a process by experimenting with the compact model, rather than running expensive and time-consuming experiments. Therefore compact models should be scalable with geometry and accurate across a wide temperature and bias voltage range.

Several good compact models have been developed for digital, analog, and RF applications [16] [17] [18] [19]. These models use a combination of physical and empirical methods to develop general equations, usually a large number of them, to describe the behavior of the device. Several parameters are embedded in each equation in order to customize it for the desired technology or device. These parameters are necessarily determined through complicated curve fitting procedures (parameter extraction) and shape the familiar model card for circuit designers. Most compact models have the advantage of describing the behavior of the device in all regions of operation at the same time. Furthermore, they provide small and large signal analysis as well as noise analysis. They also operate over a fair range for geometry, width and length of the device, over which the extracted parameters are valid. This generality however comes with an accuracy penalty if the model is used over a bias or geometry range outside of the extraction process. Moreover, the core equations in most compact models have been derived under quasi-static assumptions. This, together with the fact that most of available extracted parameters are done so for low-GHz frequency purposes, make these compact models less desirable and inaccurate for millimeter wave applications.

There are two main reasons for this inaccuracy: First of all, as mentioned before, the fact that the parameter extraction has been done in lower frequencies makes the extrapolation to mm-wave frequencies problematic. Some of device mechanisms that are not well captured at low frequencies, and naturally not modeled properly, have considerable effect on the performance of the device in higher frequencies, resulting

in some inaccuracy. The substrate network including capacitances and resistances is an example of such an effect. The inaccuracy due to this effect could be addressed by increasing the frequency range for parameter extraction process.

The second reason for the error in modeling is due to the important role of the device layout and this is more difficult to address. The device interconnections to the outside world introduce small inductors, resistors and capacitors to the model. These small components are generally negligible at lower frequencies making the device model more or less independent of layout. These components however change and in fact dominate the performance of the device as the frequency increases. This makes it crucial to include these parasitics into the model. An accurate prediction of these parasitic requires full-wave electromagnetic simulation, which is difficult and expensive to perform in a compact model, and therefore it is best to capture these parasitic through experimental means. Existing compact models can be used as the core for a hybrid customized mm-wave model. In essence, each small finger of the transistor is modeled with the "intrinsic" transistor model based on the quasi-static equations whereas the interconnection between the fingers and the interconnect are captured by electromagnetic simulation and experimental techniques.

Given the difficulties in modeling the device, one may be tempted to work directly with measured data. In traditional microwave design the common approach is to use measured S-parameter data for a specific device and treat to the transistor as a black box. This approach is very accurate in nature and accounts for all parasitics and distributed effects associated with the device and the layout. While this method is sufficient for small-signal circuit design applications, the accuracy of the S-parameter data hinges on reliable measurements of the device and de-embedding structures. As a result, the accuracy of the method may deteriorate for very high frequencies, both due to limited accuracy of test equipment and due de-embedding errors. Besides, this method is not suitable for simulation of any non-linear circuit such as mixers or oscillators or the assessment of the dynamic range of amplifiers. Moreover, because the transistor is treated as a black-box, there is no physical insight for improving the device performance or layout. Due to these issues, for mm-wave application, a combination of RF methodology and traditional microwave method is preferred even for small-signal applications.

3.2.2 Active Device Design

As was mentioned in the previous section, device performance in mm-wave frequencies is impacted by the device layout. Unlike low frequency circuit design in which the device design is absolutely in the realm of process engineers, here the circuit designer could – and should – alter the device performance drastically by changing the device layout. This enables the designer to layout the device based on the performance metric which is more important in a specific application. f_{max}, for instance, which is an indicator of the speed of the transistor, has been reported to vary from

80 GHz to up to 280 GHz for an identical CMOS 90nm process, mainly due to differences in the layout [20][21][2].

The most common and important performance metrics for mm-wave devices include the f_T, f_{max}, Maximum Stable Gain (MSG), Maximum Unilateral Gain (U), output power (P_{out}), drain efficiency (η_d) and minimum noise figure (F_{min}). The selection of the appropriate metric as an optimization target depends on the specific application of the device.

If a transistor is modeled as a linear two-port network and represented with y-parameters, a common figure-of-merit used to characterize the active device is Mason's unilateral gain [6],

$$U = \frac{1}{4}\frac{|y_{21} - y_{12}|^2}{g_{11}g_{22} - g_{12}g_{21}} \tag{3.35}$$

where $g_{ij} = \Re(y_{ij})$. One property of the unilateral gain, as proved by Mason, is that the addition of lossless capacitors or inductors around the device will not change U. The conditions for a generic two-port to be active are complex [22], but for nearly all CMOS transistors, a simpler test for activity can be used, viz.

$$U > 1 \tag{3.36}$$

For most practical cases, U is a monotonically decreasing function of frequency, and the device becomes passive where $U = 1$. Since f_{max} is often beyond the frequency capabilities of the measurement system, 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 f_{max}, U often drops at a rate much faster than 20 dB/decade (Fig. 3.20). Therefore, it is imperative that U is measured and modeled as closely to the targeted operating frequency of the circuit as possible to minimize errors associated with extrapolation.

Transistor Layout

Consider a multi-finger transistor composed of N_F identical devices in parallel. If ideal connections between all of the devices are assumed, the port admittance matrix of the composite device is $Y_{TOT} = N_F Y_{Finger}$. Since all of the admittances are scaled by N_F, it is easy to see from Eq. 3.35 that

$$U_{TOT} = U_{Finger} \tag{3.37}$$

Therefore, f_{max} of the multi-finger transistor is identical to f_{max} of the individual fingers. Even if non-ideal interconnect is modeled, the dominant effect for transistors much smaller than a wavelength is additional low-loss parasitic capacitance and inductance, which have a negligible effect on U, and thus f_{max} is also not greatly

[2] Portions of this text are taken from [3], (© IEEE 2005)

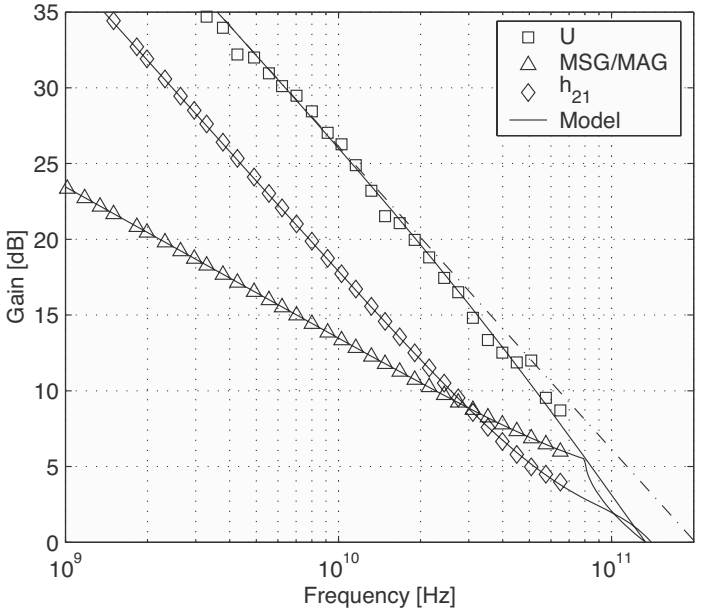


Fig. 3.20 Measured (markers) and modeled (solid lines) unilateral gain U, maximum stable gain MSG, maximum available gain MAG, and current gain h_{21}, for a typical NMOS device ($W/L = 100 \times 1\,\mu\text{m}/0.13\,\mu\text{m}$, $I_{DS}/W = 300\,\mu\text{A}/\,\mu\text{m}$, $V_{DS} = 1.2\,\text{V}$). The maximum frequency of oscillation, based upon the device circuit model, is $f_{max} = 135\,\text{GHz}$. This is much lower than the value of 200 GHz attained if a 20 dB/decade slope is assumed [3] (© IEEE 2005).

affected. In the next section we will revisit this issue when we consider transistor layouts that become appreciably large to invalidate this assumption.

Therefore, for now it is sufficient to only consider the optimal layout for a single finger. The physical layout of a single finger is shown in Fig. 3.21, 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), non-quasi-static channel resistance (r_{nqs}), and resistive substrate network (R_{sb}, R_{db} and R_{bb}) [23]. The series source and drain resistances are dominated by the intrinsic spreading resistance and are less sensitive to layout changes (such as by increasing the length of the source/drain regions). The non-quasi-static channel resistance models the effective increase in gate resistance due to the finite channel charging time, and can be shown to be inversely proportional to g_m [24]. The gate resistance, R_G, accounts for the distributed RC nature of the polysilicon gate, and can be approximated using a lumped resistor with value [8] [9]

$$R_G = \frac{R_{poly} W_F}{3n^2 L} \tag{3.38}$$

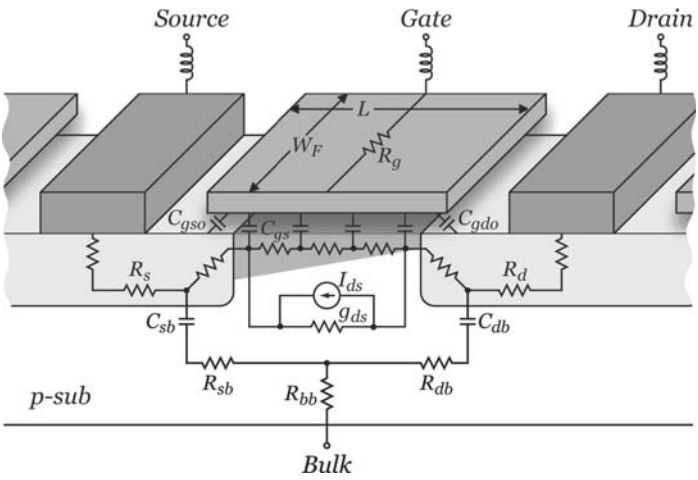


Fig. 3.21 Simplified physical model for one finger of an NMOS device [3] (© IEEE 2005).

where R_{poly} is the polysilicon gate sheet resistance, W_F is the finger width, L is the channel length, and $n = 1, 2$ is determined by the number of gate contacts.

Although using two gate contacts reduces the gate resistance by a factor of 4, one can alternatively use fingers with width $W_F/2$ and double the number of fingers to achieve the same effect. Using two gate contacts requires more complicated routing of the gate line, and often comes at the expense of increasing other parasitics, such as the gate inductance, the source resistance and inductance, and substrate resistance. Therefore, for an in-line transistor layout the narrowest possible gate fingers should be used.

To gain insight into the effect of layout on f_{max}, consider the first-order approximation [11] [12] (neglecting R_D and substrate losses),

$$f_{max} \approx \frac{f_t}{2\sqrt{R_g\left(g_m C_{gd}/C_{gg}\right) + \left(R_g + r_{ch} + R_S\right) g_{ds}}} \tag{3.39}$$

where $C_{gg} = C_{gs} + C_{gd}$ is the total gate capacitance. It is well-known that in order to maximize f_{max}, the fingers should be narrow. For a fixed I_{DS}/W (i.e., constant V_{GS}), reducing the finger width results in R_G being decreased, while all other resistances (per finger) are increased. For very narrow finger widths, the second term in the denominator of Eq. 3.39 will dominate, R_G will become negligible, and f_{max} will approach a value that is independent of W_F. The polysilicon gate sheet resistance only affects how narrow the fingers must be made. Although the dependence of f_{max}

on the complete model with all parasitics is complicated and cannot be expressed in closed-form, Eq. 3.39 possesses the correct qualitative dependence of f_{max} on W_F. Furthermore, minimizing W_F to optimize f_{max} allows the substrate contacts to be placed more closely to the device, thus minimizing the losses due to the network. 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.

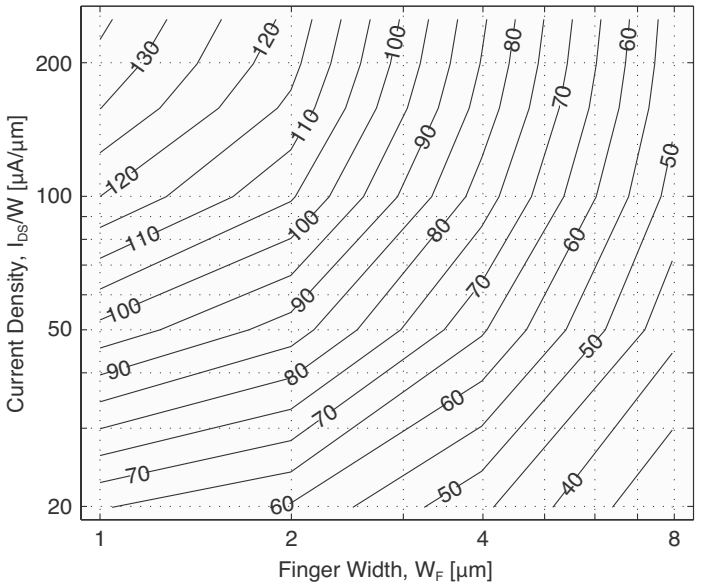


Fig. 3.22 Measured f_{max} [GHz] for minimum channel length ($L = 0.13\,\mu$m) NMOS transistors. The constant f_{max} contour lines are linearly interpolated between the measurement data. The peak measured f_{max} for a $100 \times 1\,\mu\text{m}/0.13\,\mu\text{m}$ device biased at $I_{DS}/W = 250\,\mu\text{A}/\mu\text{m}$ is 135 GHz [3] (© IEEE 2005).

The optimal transistor finger width for a commercial 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. 3.22. Nine devices with $W_F = 1\text{-}8\,\mu\text{m}$ and $N_F = 40\text{-}100$ in common-source configuration with the bulk and source grounded and the gate contacted on one side were fabricated. For six bias points ($20 - 300\,\mu\text{A}/\mu\text{m}$) per device, a transistor model was extracted from the measured data in order to find f_{max}. The constant f_{max} contours shown in Fig. 3.22 were linearly interpolated between the measured data points.

For a constant current density, the device f_T remains fixed (e.g., $100\,\mu\text{A}/\mu\text{m}$, f_T is 70 GHz). It is clear from Fig. 3.22 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 $1\,\mu\mathrm{m}$ each). This is in stark contrast to GaAs FETs with metal gates, where relatively few fingers of wide devices ($30-75\,\mu\mathrm{m}$) are typically used [25]. Furthermore, the device must be biased well into strong inversion (around $100-300\,\mu\mathrm{A}/\mu\mathrm{m}$) 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.

Transistor Wiring

To determine the effect of the wiring layout on device parameters, a physical small signal model is used in order to ascertain the effect of each parasitic on the desired performance metrics of the device. The small-signal model of the device is not necessarily unique and different combination of lumped element values could satisfy the required matching between the measurement and the simulation result. As a result, in order to make the model physical, the values of these parasitics were partly determined through 3D EM simulation (Ansoft HFSS) and were set as the initial value for optimization [3].

The developed physical small-signal model helps determine the effect of each parasitic element on the performance of the transistor. Ultimately this insight can be used to determine the optimal transistor wiring and layout. For example, the layout and wiring of a common source device has been optimized for f_{max} in a commercial 90nm technology process. Using the previously discussed approach, the starting point for this procedure was a $N_F = 80$ finger $80\,\mu\mathrm{m}/90\mathrm{nm}$ sized device with an MSG of 7.5 dB at 60 GHz and the extrapolated f_{max} of 143 GHz. Compared to 130 nm technology, this device consumes less current for the same gain, but it only has a slightly higher gain at 60 GHz and only a marginally larger f_{max}. A sensitivity analysis was performed using a small-signal model and the variation of maximum unilateral gain U and maximum stable gain together with maximum frequency of operation, f_{max}, were determined.

As expected, the gate to drain capacitance and the gate resistance have the largest impact on f_{max}, and thus layout wiring should minimize them as much as possible. It is worth noting that MSG does not change with a reduction in gate and drain series resistances when the transistor is conditionally stable ($K < 1$). This makes sense since to render the device stable in this region, we have to add resistances to the input and output ports. The source resistance, however, changes the MSG since it changes the effective transconductance through its local series feedback effect.

The NMOS structure was modified based on these findings. Mainly the shape of gate and drain tapers, number of gate vias, and width of connections and gate/drain overlap regions were changed. Fig. 3.23 shows a layout comparison of the structure before and after modification. The measured performance of the initial device and that of the modified device is shown in Fig. 3.24. The f_{max} for the improved structure is up to 178 GHz as was predicted by the analysis. The maximum stable gain of

[3] Portions of this text are taken from [27], (© IEEE 2007)

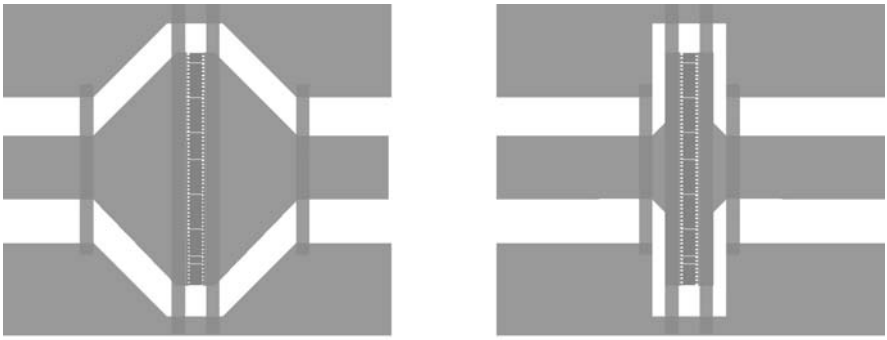


Fig. 3.23 Initial and improved layout for an $80\,\mu\mathrm{m}/90\,\mathrm{nm}$ NMOS device. The improved layout also includes more substrate contacts and a higher density of gate vias.

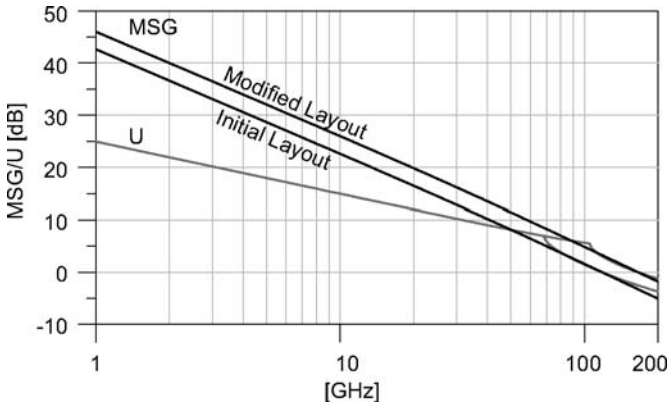


Fig. 3.24 The effect of layout improvement over Mason's Unilateral Gain U and MSG [27] (© IEEE 2007).

the device is intact however since the device is in the conditionally stable region and changing series gate and drain resistances does not change the maximum stable gain.

Round-Table Structure

Working with a simple in-line transistor layout, our previous research showed that the optimal multi-finger layout and wiring of an NMOS device could increase the f_{max} up to 20%, but increasing the performance required further innovation. This is particularly true of the available gain in conditionally stable frequencies. In order to improve the performance of the device even further, a new structure for the device

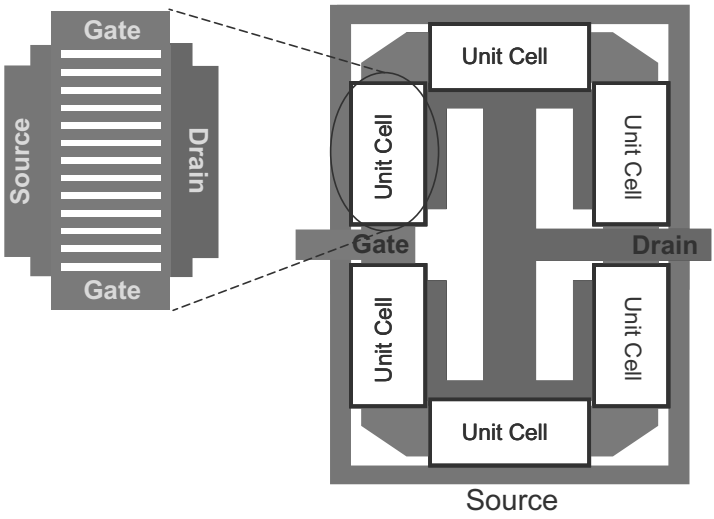


Fig. 3.25 Layout of the “Round Table" device [27] (© IEEE 2007).

was proposed [26]. The idea is to reduce the parasitic losses by using a modular approach in device design.

The building block is a standard $W = 10\,\mu\mathrm{m}$ devcie with $N_F = 10$ fingers, using double-gate contacts in order to decreases the finger resistance of the device as much as possible. Since each finger of the device forms a distributed RC network, double contact reduces the resistance of each finger by a factor of four [29]. These cells are then connected in a matrix or circular fashion depending on the desired size of the final transistor. Fig. 3.25 shows a $W = 60\,\mu\mathrm{m}$ NMOS using a circular connection, hence the name “Round Table". This structure uses external double-contacts (between cells) and multi-path connections between sources and drain of the sub cells.

Several dimensions of these devices were fabricated in 90nm CMOS process. Measurements were carried out up to 65 GHz and probing pads were de-embedded from the devices. Fig. 3.26 shows MSG, Mason’s gain (U) and H_{21} of a $W = 40\,\mu\mathrm{m}$ round-table NMOS. The f_{max} is calculated by extrapolation of the Mason’s gain U for frequencies between 20 GHz to 50 GHz, a frequency range where the most reliable data occurs. As evident, measurements suggest significant improvement in both the speed and the desired gain of these devices as compared to in-line layout transistors with the same number of fingers. Even though f_T remains almost constant (100 GHz), f_{max} improved by almost two fold, or to about 300 GHz. This is of course the extrapolated f_{max} since the device introduces new high frequency poles after 100 GHz, rendering the linear approximation of the U curve inaccurate beyond 100 GHz. The model extrapolated f_{max} is 200 GHz. Unlike the improvement of the in-line multi-finger device presented in the previous section, the MSG of the round-table device increases even at frequencies in which the device is conditionally

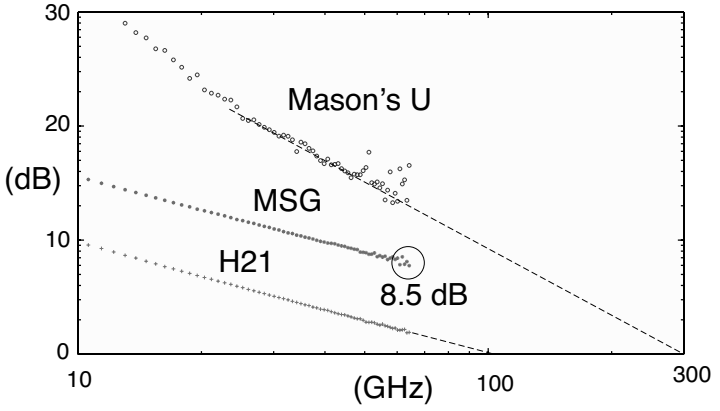


Fig. 3.26 Measured h_{21}, Mason's Unilateral Gain (U) and Maximum Stable Gain (MSG) for a $40\,\mu\mathrm{m}/90\,\mathrm{nm}$ Round Table device [26] (© IEEE 2007).

stable. The MSG at 60 GHz is 8.5 dB from the value of 7.5 dB (regular NMOS) for $I_D = 28\,\mu\mathrm{A}/\mu\mathrm{m}$. The ratio, f_{max}/f_t, a measure of the optimality of the physical structure of the device, is greater than 2 and as large as 3 (when using the 20 dB/dec extrapolation), the highest reported for CMOS. The improvement of the maximum stable gain is the result of decreased source resistance and parasitic drain to gate capacitance that both act as internal series and shunt feedback. The improvement of f_{max} was mostly due to reduction in the gate and drain resistances.

Table 3.1 A comparison of the small-signal equivalent circuit parameters for an in-line transistor and a Round Table transistor [27] (© IEEE 2007).

	In-Line Device	**Round Table Device**
$R_g\ (\Omega)$	4.46	2.23
$R_d\ (\Omega)$	3.54	2.42
$R_s\ (\mathrm{m}\Omega)$	672	438
$C_{gs}\ (\mathrm{fF})$	35.7	57.1
$C_{gd}\ (\mathrm{fF})$	21.3	17.2

Table. 3.1 compares the result of extracted small-signal parameters of a round-table $W = 40\,\mu\mathrm{m}$ device to a regular optimized multi-finger $40\,\mu\mathrm{m}$ transistor. All the resistive losses have been reduced considerably as shown in the accompanying table. The parasitic gate-source capacitance of the device is increased. This is mainly due to the increased overlap capacitance between source and gate in order to reduce the gate and source resistances. This is a good trade-off since the C_{gs} does not directly affect the f_{max} and MSG as much as the device loss.

3.2.3 Small-Signal Model

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 [23].

Considering the small margins for modeling errors, the following modeling methodology for active devices is suggested 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 should be extracted only for fixed layouts.
- The transistors in the circuit have the identical layout as the devices used for the model extraction.
- 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 [1].

The physical model depicting the significant high-frequency parasitics was shown in Fig. 3.21, and Fig. 3.27 shows the corresponding extended circuit model. The core device is modeled using either a lumped small-signal model (Fig. 3.27) 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_{gd}, 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 [30]. S-parameters for the simulated small-signal model and measured data up to 65 GHz are shown in Fig. 3.28 for a $100 \times 1\,\mu\mathrm{m}/0.13\,\mu\mathrm{m}$ NMOS transistor biased at $V_{GS} = 0.65$ V and $V_{DS} = 1.2$ 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.

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. 3.20. The accurate 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.

For a common-source 90nm CMOS device, a small-signal model as shown in Fig. 3.29 is used. This lumped equivalent circuit also adds parasitic resistances and inductances as well as a substrate network to the well-known hybrid-π model of the

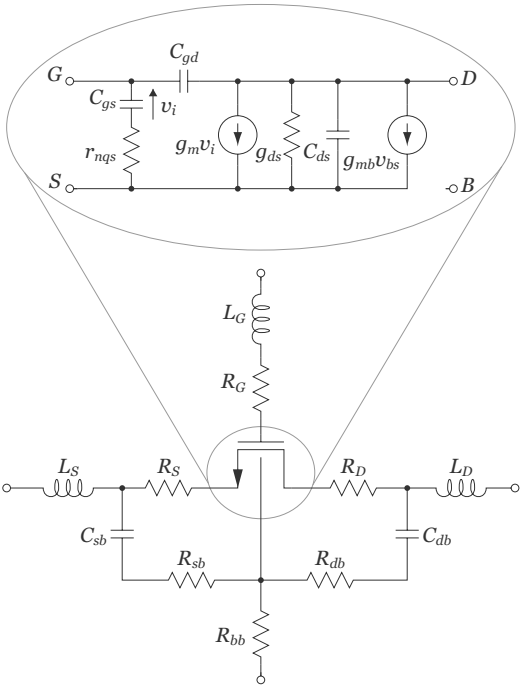


Fig. 3.27 Small-signal transistor model for an NMOS device showing the important parasitic elements [3] (© IEEE 2005).

core transistor. Figure 3.30 shows the result of this modeling technique by comparing S-parameters measurement of a sample common-source transistor to its extracted model. The model shows a good match with the measurement up to 100 GHz as evident in this figure.

3.2.4 Large-Signal Model

Although ac models are often sufficient for the design of linear circuit blocks, the optimization of nonlinear circuits such as mixers, power amplifiers, oscillators, and frequency multipliers, requires precise knowledge of the nonlinear characteristics of the active devices over a wide operation range[4].

Large-signal device models have evolved into two basic categories: table-based [32] and physical [33][34]. At mm-wave frequencies, GaAs FET models commonly employ table-based models derived from bias-dependent linear measurements. The large-signal accuracy of table-based models is limited by the existence of discontinu-

[4] Portions of this text are taken from [1], (© IEEE 2004)

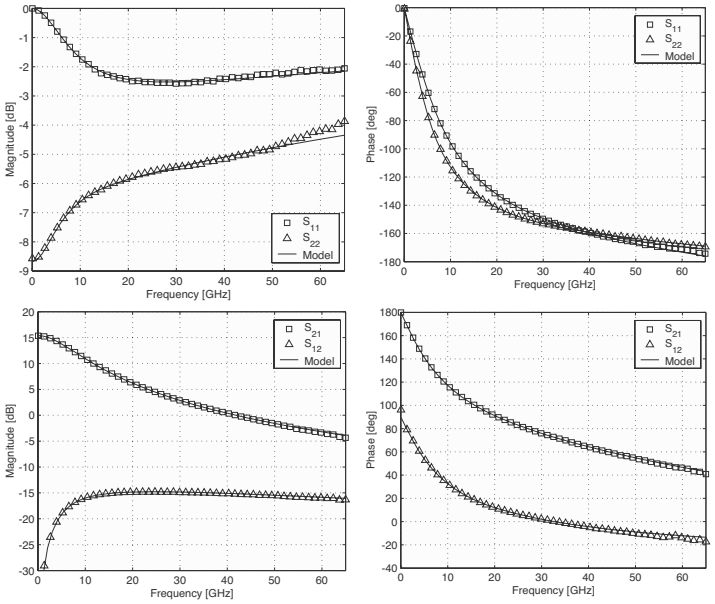


Fig. 3.28 Measured (markers) and simulated (solid lines) S-parameters for a typical NMOS device ($W/L = 100 \times 1\,\mu\text{m}/0.13\,\mu\text{m}$, $I_{DS}/W = 300\,\mu\text{A}/\mu\text{m}$, $V_{DS} = 1.2\,\text{V}$) [3] (© IEEE 2005).

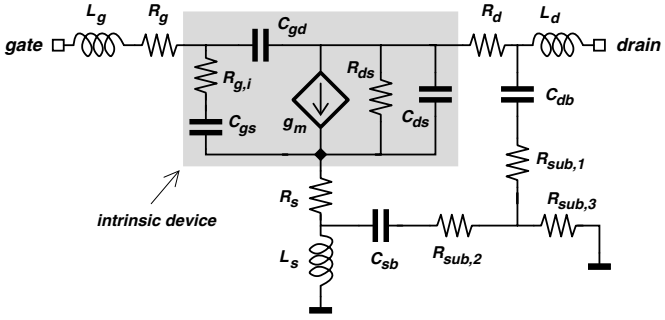


Fig. 3.29 Small-signal equivalent circuit of a common source CMOS device modeled in 90 nm technology [31] (© IEEE 2007).

ities in the model elements and nonlinearities in their interpolation due to imperfect measurement data [35]. Recent efforts to model the large-signal distortion performance of CMOS transistors with compact models have provided accurate results, but have been verified only at frequencies below 2 GHz [34][36].

In this section, we introduce an extension to the standard BSIM3 modeling procedure, and provide experimental validation of small-signal and large-signal accuracy up to 65 GHz.

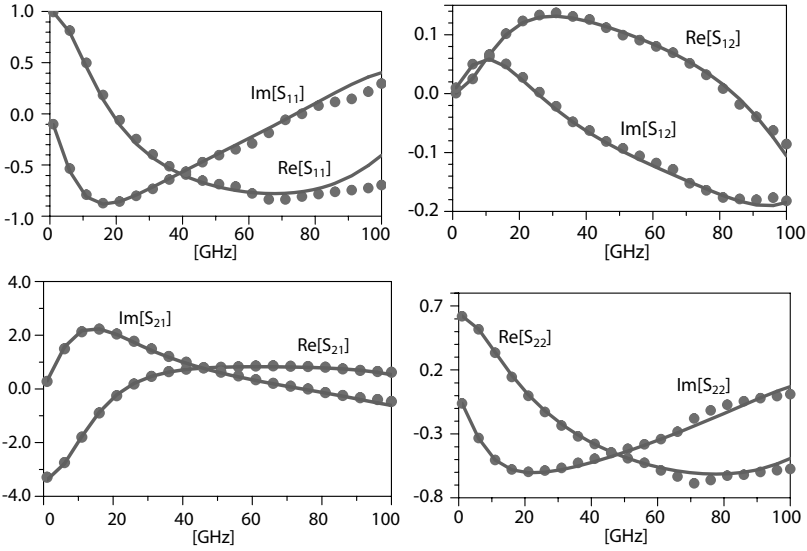


Fig. 3.30 A comparison of the measurement (marker) and modeled (line) transistor S-parameters up to 100 GHz for an NMOS device in 90 nm technology [27] (© IEEE 2007).

Modeling Methodology

The proposed modeling methodology is based on the quasi-static assumption that the mm-wave large-signal performance of a transistor is primarily governed by its dc nonlinearities, while the dynamic performance can be modeled with the addition of extrinsic parasitics to capture loss and inductive effects, which become particularly important at mm-wave frequencies.

The proposed modeling methodology uses a core BSIM model for the intrinsic transistor, augmented with extrinsic parasitics as depicted in Fig. 3.31. The series source and drain resistances are dominated by the intrinsic spreading resistance in the source-drain extension (SDE) region near the channel. The gate resistance, R_G, accounts for the distributed RC nature of the polysilicon gate, and can be approximated using a single lumped resistor[5]. Additionally, series inductors must be added to all terminals, L_G, L_D, L_S, to properly model the delay effects associated with interconnect wiring. Layout-dependent interconnect capacitances are also added around the intrinsic device, and junction diodes account for the voltage-dependence of C_{db} and C_{sb}.

[5] The bias dependence due to the channel inversion level is ignored.

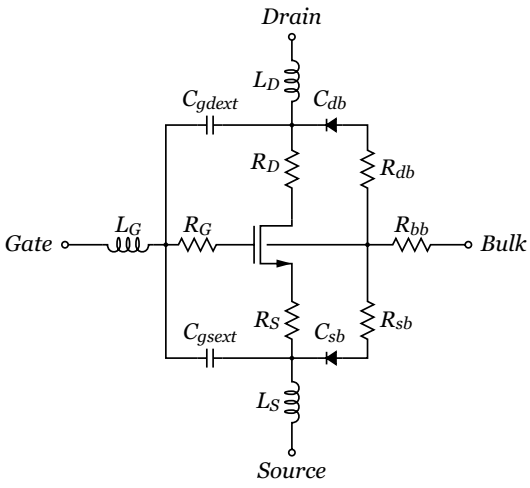


Fig. 3.31 BSIM3 core model with extrinsic parasitics [1] (© IEEE 2004).

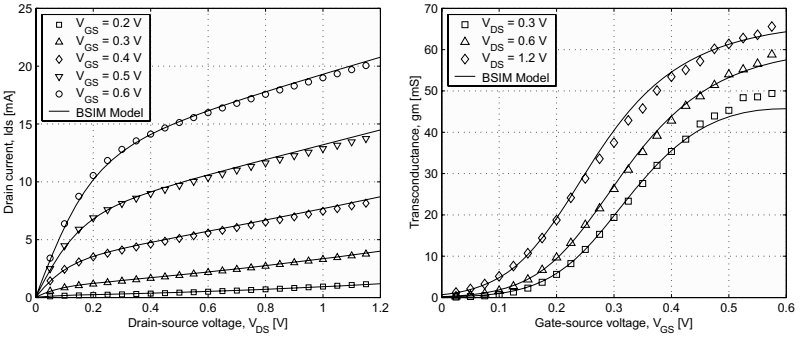
Model Extraction

A $80 \times 1\,\mu\mathrm{m}/0.13\,\mu\mathrm{m}$ NMOS transistor, optimized for operation at mm-wave frequencies and fabricated on a standard 0.13-μm CMOS process, is used as a demonstration vehicle in the large signal modeling. De-embedding structures were used to remove the effects of the pads from the small-signal device characterization.

For dc large-signal characterization, I-V measurements were used to extract the core BSIM parameters of the fabricated common-source NMOS. As shown in Figs. 3.32a and Fig. 3.32b, a good agreement between measured and modeled dc curves can be achieved for this device. Next, over a wide bias range, extensive on-wafer S-parameter measurements to 65 GHz were performed on a Cascade Microtech probe station using an Anritsu 37397C VNA and Cascade Microtech GSG coplanar probes. The external parasitic component values for the model were extracted using a hybrid optimization algorithm in Agilent IC-CAP. S-parameters for the simulated model and measured data up to 65 GHz are shown in Fig. 3.33 for a typical bias sweep over V_{GS} for the de-embedded transistor. The broadband accuracy of the model verifies that using lumped parasitics (Fig. 3.31) is suitable well into the mm-wave region.

Large-Signal Measurement Setup

The nonlinear model described in Section 3.2.4 needs experimental verification. Two common approaches to characterize transistor mm-wave nonlinearity are with mm-



(a) (b)

Fig. 3.32 (a) Measured and modeled I_{DS} vs. V_{DS}. (b) Measured and modeled g_m vs. V_{GS} [1] (© IEEE 2004).

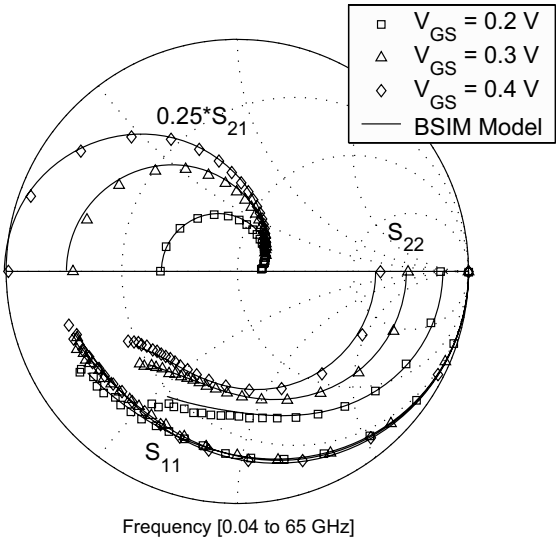


Fig. 3.33 Measured and modeled S-parameters for $V_{DS} = 1.2$V, and $V_{GS} = 0.2.0.4$V [1] (© IEEE 2004).

wave load-pull measurements and power spectrum analysis [37]. While the former requires automated tuners, the latter, which was chosen in this work, can be performed with only a synthesizer, VNA, and power meter (Fig. 3.34). The fabricated device is driven over a wide range of input power and bias conditions, while the output powers at the fundamental and harmonic frequencies are measured. In the harmonics power measurement setup of Fig. 3.34, the VNA is configured as a receiver in the

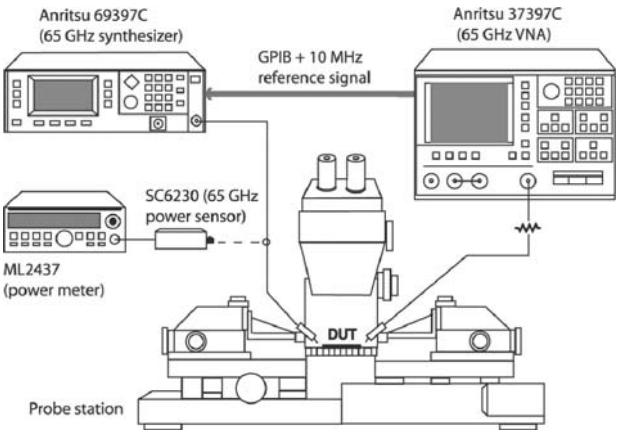


Fig. 3.34 Test setup for mm-wave harmonics measurements [1] (© IEEE 2004).

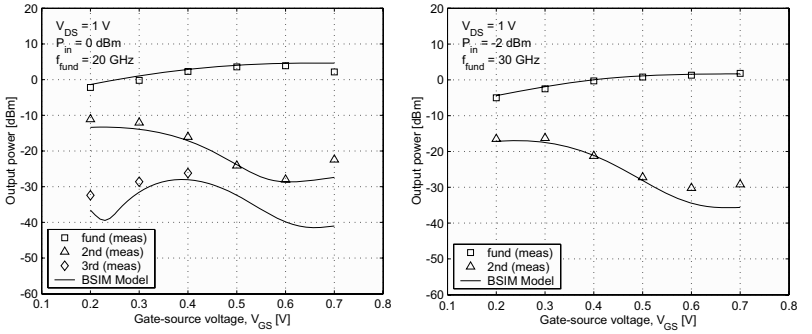
Set-On mode. In this mode of operation, the source lock circuitry of the 37397C is completely by-passed which allows all of the 37397C samplers to operate over their full dynamic range. The 65 GHz synthesizer and 37397C are locked to the same 10 MHz reference, enabling coherent reception at the harmonic frequencies. A 65 GHz high dynamic-range power sensor is used to de-embed the insertion loss of the cables, probes, adaptors, etc. from the measurements. Port2 of the VNA requires a 10-dB attenuator to avoid compression when the fundamental power is strong, thus limiting the sensitivity of the power measurement at 60 GHz to approximately -35 dBm.

Measured And Simulated Results

Fig. 3.35a shows the measured and modeled fundamental, second, and third harmonics at the output when the device is driven by a 20 GHz 0-dBm signal with a bias of $V_{DS} = 1$V and variable V_{GS}. The output powers at the fundamental and second harmonics agree very well, whereas the agreement at the third harmonic is reasonable, but not as good. The measurement of the third harmonic for $V_{GS} > 0.4$V is limited due to the dynamic range of the VNA samplers at 60 GHz.

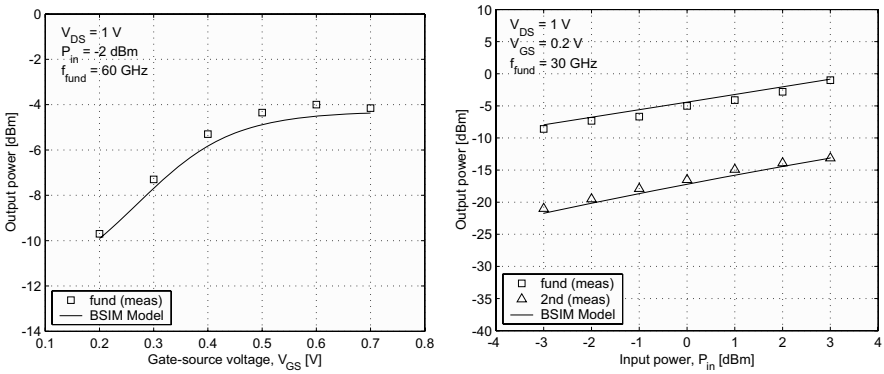
Similar experiments were performed at a different input power (-2 dBm) and signal frequency (30 GHz). Fig. 3.35b shows the measured and modeled fundamental and second harmonics of the output signal for the same bias conditions. Good agreement is observed at both the fundamental and second harmonic. Since the second harmonic at 60 GHz is strong, it is within the dynamic range of the measurement setup.

Fig. 3.36a displays the measured and modeled 60 GHz large-signal amplification characteristics of the device. The model closely predicts the fundamental output



(a) (b)

Fig. 3.35 (a) Fundamental, second, and third harmonics vs. the gate bias (V_{GS}) for $V_{DS} = 1$V, $P_{in} = 0$dBm, $f_{fund} = 20$ GHz. (b) Fundamental and second harmonic vs. the gate bias (V_{GS}) for $V_{DS} = 1$V, $P_{in} = -2$dBm, $f_{fund} = 30$ GHz [1] (© IEEE 2004).



(a) (b)

Fig. 3.36 (a) Large-signal gain vs. gate bais (V_{GS}) for $V_{DS} = 1$V, $P_{in} = -2$dBm, $f_{fund} = 60$ GHz. (b) Class AB power sweep curves [1] (© IEEE 2004).

power for the device over a large bias range from weak inversion to very strong inversion.

Many mm-wave circuits, such as mixers and frequency multipliers, operate by exploiting the strong nonlinearity of transistors biased near the threshold voltage, effectively using the device as a rectifier. To determine the accuracy of the model for these applications, the transistor was biased at constant $V_{GS} = 0.2$V, and the input power was swept from $P_{in} = -3$dBm to $+3$dBm at a fundamental of 30 GHz. The results shown in Fig. 3.36b show that the extended BSIM model provides good accuracy for class AB operation over a wide range of input signal power.

The results validate that with accurate dc nonlinearity fitting, a simple extended compact model can predict the harmonic distortion behavior of a device up to 60 GHz. The ability to model the strongly nonlinear case of class AB operation is also verified, which suggests that the model is viable for use in the design of nonlinear mm-wave blocks such as mixers, power amplifiers, and frequency multipliers.

3.2.5 FET Noise Model

An accurate noise model is crucial in designing low noise amplifiers. Typically foundry compact noise models have been optimized to fit measurements up to 20 GHz, and typically the mismatch between measurement and the model is significant at higher frequencies due to inadequacies of compact models. This inadequacy arises from the distributed nature of the device near the f_T and uncertainty to the cause of the excess noise in short channel devices. Traditional microwave design, on the other hand, is based directly on two-port noise parameters. Although this method is the most accurate, its applicability is limited due to the difficulty in noise parameter measurement and the lack of robust de-embedding for Si devices.

The noise figure of a two-port system based on the four noise parameters is given by [2]

$$F = F_{min} + \frac{R_n}{G_s}\left|Y_s - Y_{opt}\right|^2 \tag{3.40}$$

in which F_{min} is the minimum achievable noise figure, R_n is the noise sensitivity resistance, and Y_{opt} is the optimal source noise admittance, and $Y_s = G_s + jB_s$ is the source admittance.

As technology has advanced, CMOS technology minimum noise figure F_{min} has dropped significantly, approximately like (f/f_t)

$$F_{min} = 1 + 2\left(\frac{f}{f_t}\right)\sqrt{g_m R_g \frac{\gamma}{\alpha}}$$

The above is calculated based on Pospiezalski's noise model [38] (see next section), which has resulted in a good match with CMOS devices in mm-wave frequencies. In the limit that we use short transistor fingers to minimize the gate resistance R_g, the noise is bounded by the NQS resistance seen at the gate $R_g = 1/5g_m$ [24], which makes the fundamental noise just dependent on the technology parameters

$$F_{min} > 1 + 2\left(\frac{f}{f_t}\right)\sqrt{\frac{1}{5}\frac{\gamma}{\alpha}}$$

Based on this simplistic prediction, a plot of F_{min} is shown in Fig. 3.37, which shows that F_{min} is as low as 2.5 dB at 60 GHz in the 90nm node. Actual measurements show that this lower bound is reached to within $1\,\mathrm{dB}$ in practice, which is very encouraging. For example, measurement results at mm-wave frequencies (60 GHz) on the 130 nm transistor also reveal that the minimum achievable noise figure is between 3-4 dB. This level of performance is sufficient for many applications. The remaining challenge is to build an amplifier that can actually match for low noise.

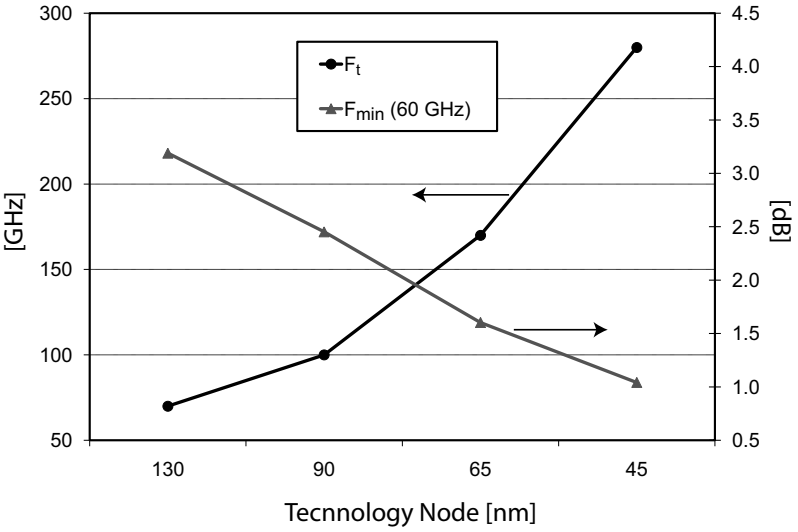


Fig. 3.37 Lower bound for minimum achievable noise figure F_{min} as a function of technology node.

3.2.5.1 Review of Noise Models

The classical Van Der Ziel RF noise model [39][40] is well known. As shown in Fig. 3.38, the high frequency noise of the device is described by two correlated

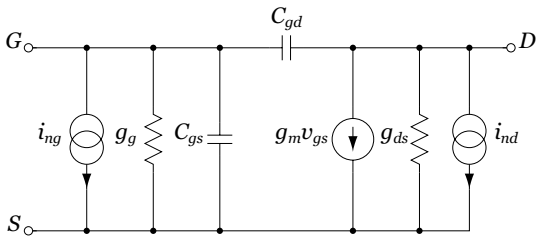


Fig. 3.38 The van der Ziel RF noise model.

current noise sources, namely the induced gate noise (i_{ng}) and the short-channel drain current noise (i_{nd}). The current noise sources are given by

$$\frac{\overline{i_{nd}^2}}{\Delta f} = 4kT\left(\frac{\gamma}{\alpha}\right) g_m \tag{3.41}$$

$$\frac{\overline{i_{ng}^2}}{\Delta f} = 4kT\delta g_g \tag{3.42}$$

$$g_g = \frac{\alpha\omega^2 C_{gs}^2}{5g_m} \tag{3.43}$$

$$c \equiv \frac{\overline{i_{ng} \cdot i_{nd}^*}}{\sqrt{\overline{i_{ng}^2}\,\overline{i_{nd}^2}}} \approx j0.395 \tag{3.44}$$

where c is the correlation factor between the two current noise sources. The induced gate noise increases with the square of frequency, playing an important role at microwave and mm-wave frequency. The short-channel drain current noise is proportional to γ, which increases with shorter channel length processes [41]. Flicker noise, usually considered a low frequency noise source, can play an important role at mm-wave frequencies due to the noise translation in non-linear and time-varying circuits.

Unfortunately correlated noise sources are not easily incorporated into standard based compact models. Even though it is mathematically trivial to add correlated noise sources to the simulators, due to the historical absence of such elements, other approaches have been taken to model the noise. A procedure to implement correlated noise into Verilog-A is reported in [42]. In the holistic thermal noise model of BSIM [16], shown in Fig. 3.39, the noise is partitioned into a gate noise and a drain noise. The induced gate noise is captured due to the source noise voltage, but unfortunately the phase of the correlation is not imaginary as predicted by the Van Der Ziel model. The Philips MOS11 model, shown in Fig. 3.39b, captures the gate noise automatically since the transistor is broken into a pseudo-distributed circuit [43]. Measurements show that in practice five sections is sufficient to accurately capture the induced noise.

On the opposite end of the spectrum, one can characterize noise based on measurement data, as shown in Fig. 3.40. In this approach, the transistor is modeled as a two-port noiseless black box with two external correlated input noise generators, namely the equivalent input noise voltage and the equivalent input noise current.

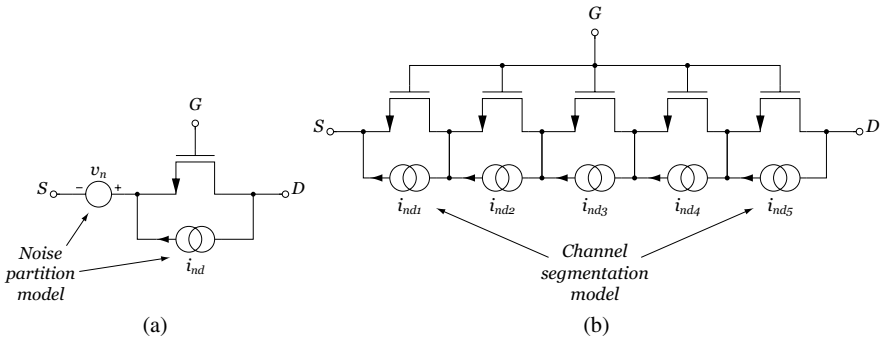


Fig. 3.39 (a) BSIM4 noise model. (b) Philips MOS 11 noise model.

The drawback is that this model is valid only for a fixed bias point and at a single frequency.

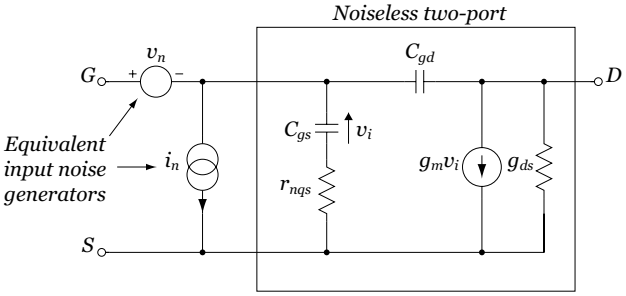


Fig. 3.40 Measurement-based noise model.

Equivalent-circuit noise models are essentially based on the Van Der Ziel model. The PRC model shown in Fig. 3.41a, models the induced gate noise and the drain current noise through the following equations

$$\frac{\overline{i_{nd}^2}}{\Delta f} = 4kTg_m \cdot P \tag{3.45}$$

$$\frac{\overline{i_{ng}^2}}{\Delta f} = 4kT\frac{\omega^2 C_{gs}^2}{g_m} \cdot R \tag{3.46}$$

$$\frac{\overline{i_{ng} \cdot i_{nd}^*}}{\sqrt{\overline{i_{ng}^2}\,\overline{i_{nd}^2}}} \approx jC \tag{3.47}$$

and the parameters P, R, and C are extracted from measurement data. On the other hand, the Pospieszalski model [38], shown in Fig. 3.41b, assumes the gate noise voltage and the drain noise current are uncorrelated, and it models the induced gate noise source as thermal noise contributed by the NQS resistance. It thus requires the extraction of only one parameter. Due to the simplicity of this noise model, and its ability to predict the device noise (shown in the next section), we have adopted this model in our mm-wave research.

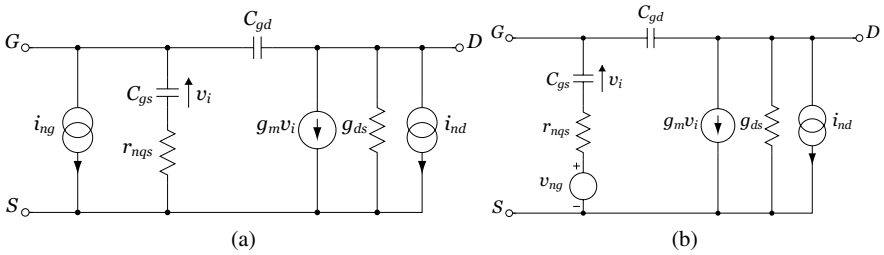


Fig. 3.41 (a) PRC noise model. (b) Pospieszalski noise model.

The basic small-signal model can include noise if one can correctly partition the gate resistance into an intrinsic and extrinsic portion. Since the extrinsic resistance is bias independent, one can use several bias points to derive the intrinsic gate resistance. Extraction of the short-channel drain noise current yields a value for γ of 1.4. The Pospieszalski model was chosen because it yields reasonably good results while being simple and requiring the extraction of only one variable, which is beneficial because high frequency measurements are usually quite difficult.

In essence the gate r_{nqs} and channel resistance (R_{ds}) are noisy. The former is kept at the gate temperature which is close to the room temperature while the equivalent temperature for the latter could reach several thousands degrees. Based on the value of noise parameter, $\gamma = 1.4$, the drain temperature is $3640°$C.

Noise Measurement and Simulation Results

Noise measurement in the mm-wave regime requires a semi-custom setup due to the unavailability of commercial solutions. A typical setup is shown in Fig. 3.42, which incorporates a calibrated noise source, a source tuner to vary the source impedance, a low noise amplifier and low noise mixer to down-convert the signal to an IF for noise measurement [44]. Most noise figure meters allow an external down-converter to be used for exactly this purpose. By varying the source impedance and measuring

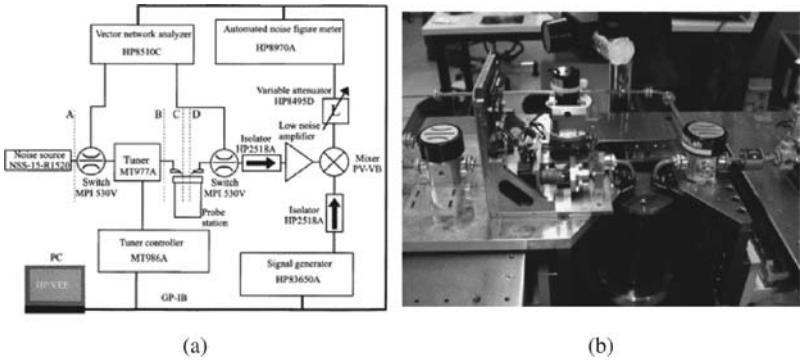


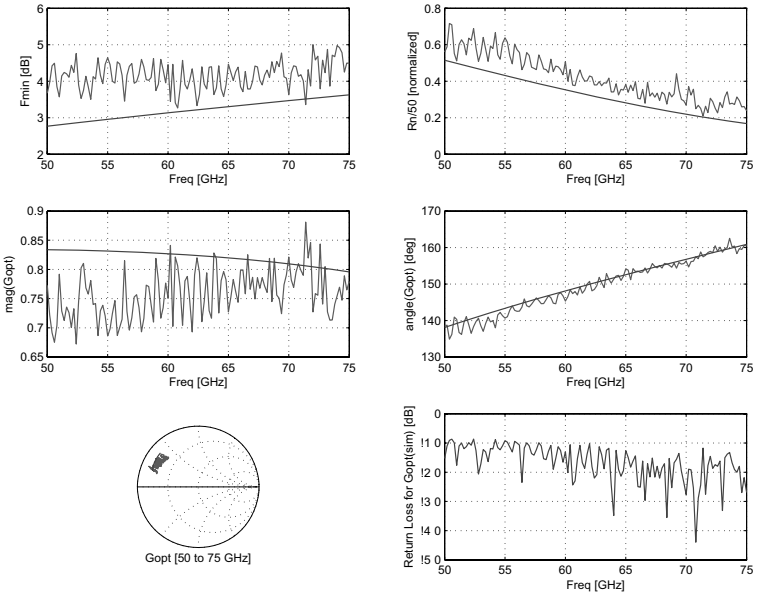
Fig. 3.42 (a) Noise measurement setup at mm-wave frequencies. (b) Photo of noise measurement setup. (Courtesy of VTT)

the noise figure of the device, we can find the noise parameters of the transistor or circuit under measurement.

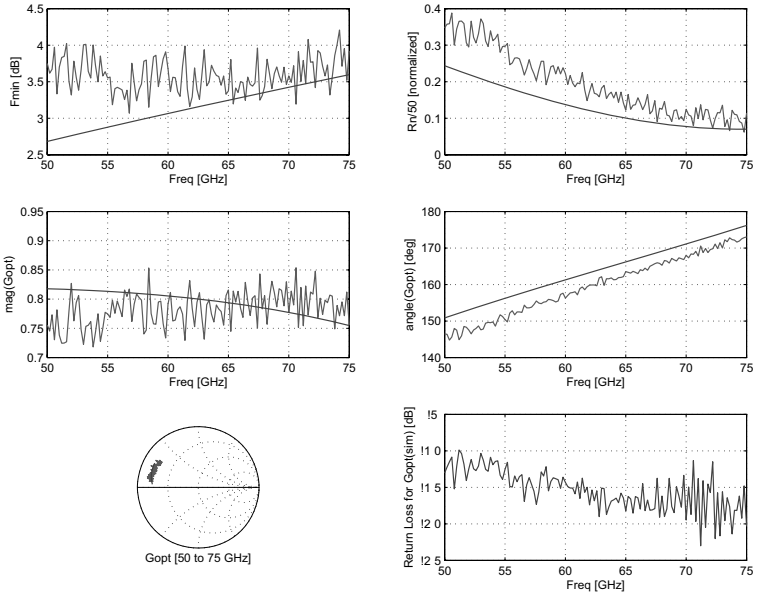
In Fig. 3.43-3.44 we compare the measured and simulated minimum noise figure (F_{min}), noise resistance (R_n), and optimal impedance (Y_{opt}) using the extended transistor modeling approach for a commercial CMOS 130nm process. All simulations results use $\gamma = 1.4$. A good match is observed in many cases, especially in the trends of the noise parameters versus bias and frequency. It's interesting to note that the noise of a cascade device is substantially higher than a single transistor, a result that differs at low RF frequencies. This is because the cascade device contributes negligible noise at low frequency due to the large degeneration impedance at the cascade node. At high frequency, the capacitance at this node presents less degeneration to the casocde device, increasing the niose. Equivalently, the drain noise of the cascode circulates within the device at low frequencies but flows through the node capacitance at high frequency, which flows through the output.

The minimum noise figure and noise sensitivity of a $W = 40\,\mu\mathrm{m}$ round-table device in a 90 nm commercial CMOS is simulated based on the proposed model and shown in Fig. 3.45. The F_{min} increases nearly linearly with frequency. However, the noise sensitivity R_n decreases in the millimeter wave region. This translates into a smaller penalty when deviating from the optimal source impedance. Another key point in designing low noise amplifiers is the trade-off between gain and noise optimization. In this case the optimal gain and noise circles for the same device are close to each other on the Smith Chart. This together with the smaller noise sensitivity means that a respectable noise figure could be achieved even by designing the amplifier only for gain performance.

In Fig. 3.46 we compare the measurement and simulated results for the 60 GHz 3-stage amplifier discussed in [3]. Measurement and simulation results match reasonably well. Most importantly, the trend with frequency and bias is well predicted, allowing the circuit designer to optimize the performance of an amplifier.

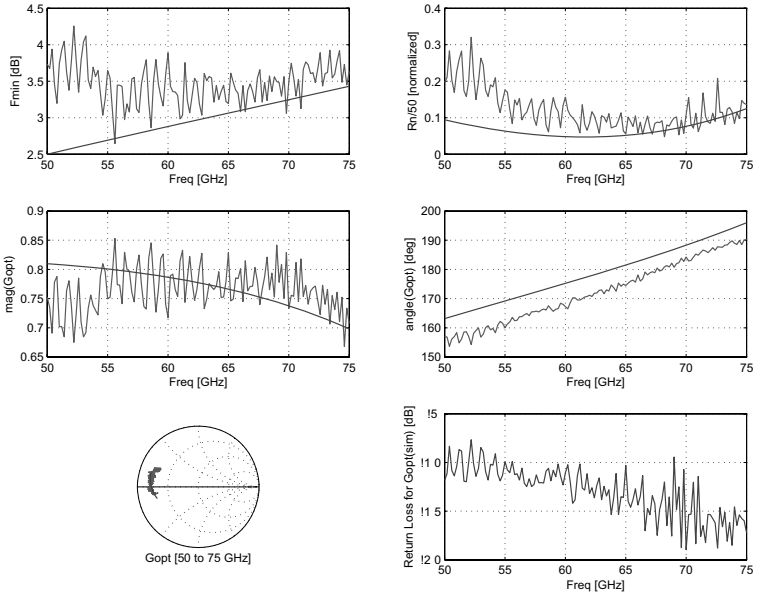


(a)

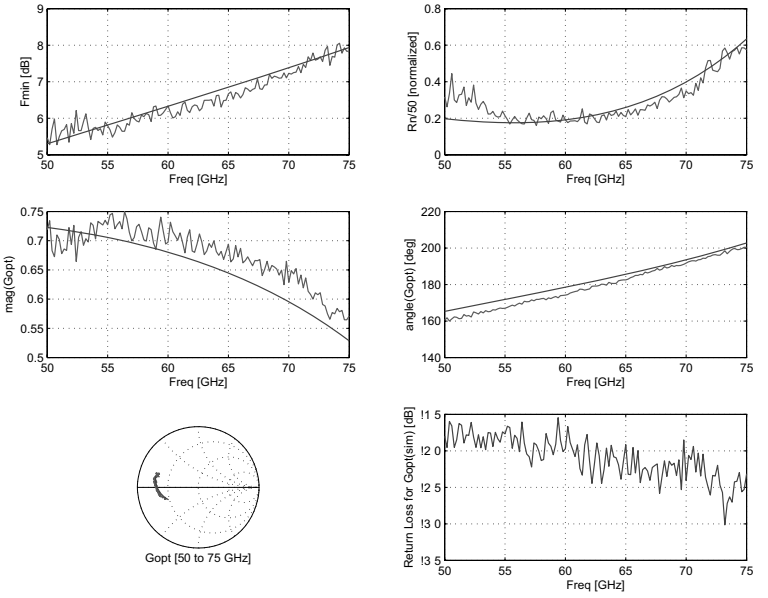


(b)

Fig. 3.43 Measurement and simulation results for the minimum noise figure (F_{min}), the noise resistance (R_n), and the optimal impedance (Y_{opt}) using the extended transistor modeling approach for the devices biased at $I_{DS}/W = 150\,\mu\text{A}/\,\mu\text{m}$. NMOS devices sized at: (a) $40 \times 1\,\mu$ (b) $60 \times 1\,\mu$.



(a)



(b)

Fig. 3.44 Measurement and simulation results for the minimum noise figure (F_{min}), the noise resistance (R_n), and the optimal impedance (Y_{opt}) using the extended transistor modeling approach for the devices biased at $I_{DS}/W = 150\,\mu\text{A}/\,\mu\text{m}$. NMOS devices sized at: (a) $80 \times 1\,\mu$ (b) Cascode $40 \times 2\,\mu$.

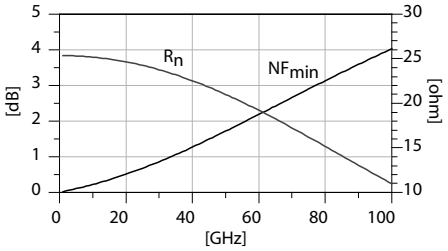


Fig. 3.45 Simulated NF_{min} and noise sensitivity parameter R_n vs. frequency for a Round Table device.

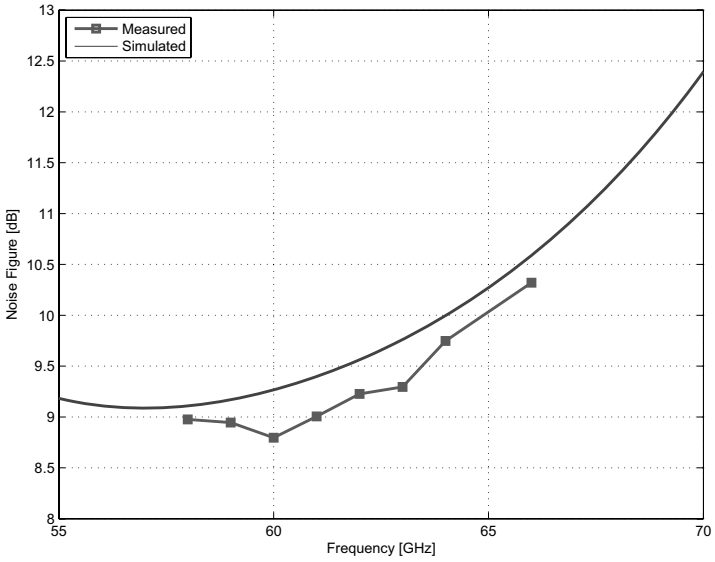


Fig. 3.46 Measurement and simulation results of noise figure for a 60 GHz amplifier [3] (© IEEE 2005).

3.3 Conclusion

In this chapter we have highlighted the design and modeling of passive and active components for mm-wave applications. The design and modeling approach is a result of several years of experiments and characterizations performed on various technology nodes (180nm, 130nm, and 90nm). Custom layout is integral for achieving performance near the limits of activity of the process. In a modern CMOS or

SiGe process, there is a rich variety of devices available to the designer, and yet the designer should not shy away from customizing these structures for mm-wave application requirements. We have demonstrated that custom tailored transmission lines, transistors, and capacitors can be used for higher quality factors, higher gain, and higher self-resonant frequency. The accuracy of the design approach and the models is borne out by measurements and by circuit building blocks described in Ch. 4. By building a library of active and passive components, and by carefully modeling the components into the mm-wave regime, one forms the foundation for a predictable and robust mm-wave design methodology. We have focused our attention on CMOS technology, but certainly with passive devices almost all observations and comments can be applied directly to SiGe or any other technology incorporating a semi-conductive substrate. The design methodology for active devices is highly dependent on the underlying device physics. We have shown, however, that the intrinsic transistor model remains valid well into the mm-wave regime, only requiring careful measurements and modeling of the extrinsic device. This simplifies Si based design tremendously, since active devices are carefully characterized and modeled over process and temperature.

References

1. S. Emami, C. H. Doan, A. M. Niknejad, R. W. Brodersen, "Large-signal millimeter-wave CMOS modeling with BSIM3," *RFIC Digest of Papers*, pp. 163-166, June 2004.
2. D.M. Pozar*Microwave engineering*, Wiley, New York, 1998.
3. C. H. Doan, S. Emami, A. M. Niknejad, and R. W. Brodersen, "Millimeter-wave CMOS design," *IEEE Journal of Solid-State Circuits*, vol. 40, pp. 144-155, Jan. 2005.
4. 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, pp. 743–752, May 1998.
5. S. Ramo, J. R. Whinnery, and T. Van Duzer, *Fields and Waves in Communication Electronics*, 3rd ed. New York: Wiley, 1994.
6. S. J. Mason, "Power gain in feedback amplifiers," *IRE Trans. Circuit Theory*, vol. CT-1, pp. 20–25, June 1954.
7. C. Doan, Ph.D. Dissertation in Preparation, U.C. Berkeley.
8. C. Enz, "An MOS transistor model for RF IC design valid in all regions of operation," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 342–359, Jan. 2002.
9. B. Razavi, R.-H. Yan, and K. F. Lee, "Impact of distributed gate resistance on the performance of MOS devices," *IEEE Trans. Circuits Syst. I*, vol. 41, pp. 750–754, Nov. 1994.
10. T. C. Edwards and M. B. Steer, *Foundations of Interconnect and Microstrip Design*, 3rd ed. New York: Wiley, 2000.
11. S. P. Voinigescu, S. W. Tarasewicz, T. MacElwee, and J. Ilowski, "An assessment of the state-of-the-art 0.5μm bulk CMOS technology for RF applications," *in IEDM Tech. Dig.*, Dec. 1995, pp. 721–724.
12. J. N. Burghartz, M. Hargrove, C. Webster, R. Groves, M. Keene, K. Jenkins, D. Edelstein, R. Logan, and E. Nowak, "RF potential of a 0.18-μm CMOS logic technology," in *IEDM Tech. Dig.*, pp. 853–856, Dec. 1999.
13. 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, pp.1480–1488, Oct. 2001.
14. G. Carchon, W. De Raedt, and B. Nauwelaers, "Novel approach for a design-oriented measurementbased fully scalable coplanar waveguide transmission line model," *IEE Proc. Microwave, Antennas Propagat.*, vol. 148, pp. 227–232, Aug. 2001.
15. W. F. Andress and D. Ham, "Standing wave oscillators utilizing wave-adaptive tapered transmission lines," *IEEE J. Solid-State Circuits*, vol. 40, no. 3, pp. 638-651, Mar. 2005.
16. The BSIM 4.4 Manual: http://www-device.eecs.berkeley.edu/bsim3.
17. MOS11, http://www.semiconductors.philips.com/Philips_Models/mos_models/model11/.
18. EKV model website, http://legwww.epfl.ch/ekv/.
19. PSP model website, http://www.nxp.com/Philips_Models/mos_models/psp/.
20. J. C. Guo, W.Y. Lien, M.C. Hung, C.C. Liu, C.W. Chen, C.M. Wu, Y.C. Sun, and Ping Yang, "Low-K/Cu CMOS Logic Based SoC Technology for 10Gb Transceiver with 115GHz fT, 80GHz fMAX RF CMOS, High-Q MiM Capacitor and Spiral Cu Inductor," *VLSI Digest of Technical Papers*, pp. 39-40, June 2003.
21. L. F. Tiemeijer et al., "Record RF performance of standard 90 nm CMOS technology," *IEDM Technical Digest*, pp. 441-444, Dec. 2004.
22. R. Spence, *Linear Active Networks*, London: Wiley, 1970.
23. C. Enz, "An MOS transistor model for RF IC design valid in all regions of operation," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 342–359, Jan. 2002.
24. Tsividis, Y., *Operation and Modeling of the MOS Transistor*, 2nd ed. Boston : WCB/McGraw-Hill, c1999.
25. I. Bahl and P. Bhartia, *Microwave Solid State Circuit Design*, 2nd ed. Hoboken, NJ: Wiley, 2003.
26. B. Heydari, M.Bohsali, E.Adabi, A. M. Niknejad, "Low-Power mm-Wave Components up to 104GHz in 90nm CMOS," *ISSCC Dig. Tech. Papers*, pp. 200-201, Feb. 2007.

27. B. Heydari, M. Bohsali, E. Adabi, A.M. Niknejad, “mm-Wave devices and circuit blocks up to 104 GHz in 90nm CMOS,” to appear in *IEEE J. Solid-State Circuits*.
28. D. Huang, W Hant, N.-Yi Wang, T.W. Ku, Q. Gu, R. Wong, M.-C.F. Chang, “A 60GHz CMOS VCO using on-chip resonator with embedded artificial dielectric for size, loss and noise reduction,” *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, pp. 1218-1227, Feb. 2006.
29. A. M. Niknejad, *Electromagnetics for High-Speed Analog and Digital Communication Circuits*, 1st Edition. Cambridge University Press, 2007.
30. Agilent IC-CAP 2002 User’s Guide, http://eesof.tm.agilent.com.
31. B. Heydari, P. Reynaert, E. Adabi, M. Bohsali, B. Afshar, M. A. Arbabian and A. M. Niknejad, “A 60-GHz 90-nm CMOS cascode amplier with interstage matching,” to be presetned at *EU Microwave Conference (EuMic)*, 2007.
32. J. Wood and D. E. Root, “Bias-dependent linear scalable millimeter-wave FET model," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 2352-2360, Dec. 2000.
33. BSIM3, http://www-device.eecs.berkeley.edu/∼bsim3
34. R. Van Langevelde, L. F. Tiemeijer, R. J. Havens, M. J. Knitel, R. F. M. Ores, P. H. Woerlee, and D.B.M. Klaassen, “RF-distortion in deep-submicron CMOS technologies," in *Electron Devices Meeting*, pp. 807-810, Dec. 2000.
35. K. Koh, H.-M. Park, and S. Hong, “A spline large-signal FET model based on bias-dependent pulsed I-V measurement," *IEEE Trans. Microwave Theory Tech.*, vol. 50, pp. 2589-2603, Nov. 2002.
36. T. Y. Lee and Y. Cheng, “MOSFET HF distortion behavior and modeling for RF IC design," in *CICC Conf. Dig. Tech. Papers*, Sep. 2003, pp. 87-91.
37. I. Angelov, H. Zirath, and N. Rorsman, “Validation of a nonlinear transistor model by power spectrum characteristics of HEMT.s and MESFET.s," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 1046-1052, May 1995.
38. M.W. Pospieszalski, “Modeling of Noise parameters of MESFETs and MODFETs and Their Frequency and Temperature Dependence," *IEEE Trans. on Microwave Theory and Techniques*, vol. 37, pp.1340-1350, Sept. 1989.
39. A. Van Der Ziel, *Noise in Solid State Devices and Circuits*, Wiley, New York, 1986.
40. A. Van Der Ziel, “Thermal Noise in Field Effect Transistors," *Proc. IEEE*, pp.1801-12, Aug. 1962.
41. A. Abidi, “High-frequency noise measurements on FETs with small dimensions," *IEEE Trans. Electron Devices*, vol. 33, pp. 1801-1805, Nov. 1986.
42. C. McAndrew, G. Coram, A. Blaum and O. Pilloud, “Correlated Noise Modeling and Simulation," *Workshop on Compact Modeling*, pp. 40-45, 2005.
43. A. J. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. A. Zegers-van Duijnhoven, V. C. Venezia, “Noise Modeling for RF CMOS Circuit Simulation," *IEEE Trans. on Electron Devices*, vol. 50, pp.618-632, March 2003.
44. http://www.vtt.fi/.