7

Compact MOSFET Models for RF Applications

7.1 Introduction

For analog and radio frequency (RF) applications of metal-oxide-semiconductor field-effect transistors (MOSFETs), it is critical to understand the behavior of these devices under applied signal. In Chapter 6, the analytical expressions are derived for modeling the terminal charges and capacitances of a MOSFET device based on the quasistatic (QS) assumption. For analog/RF applications of MOSFETs, the effect of noise, non-quasistatic (NQS) effect, and resistive network of these devices must be properly characterized and modeled for circuit CAD (computer-aided design). This chapter presents the basic understanding of modeling device noise, high-frequency characteristics such as NQS effect, and the resistive network of the devices.

7.2 MOSFET Noise Models

A careful observation of the drain current of a MOSFET reveals minute random fluctuations, referred to as *noise*. Noise is inherently present in electronic devices with or without the presence of an applied external signal. Noise could be in output current as well as in output voltage. In an MOS (metal-oxide-semiconductor) analog circuits, noise in MOSFET devices can interfere with weak signals and significantly impact the accurate characterization of circuit performance. Therefore, it is very important to model and characterize noise in devices [1–3].

The amount of noise depends on the bandwidth of the characterizing system. A common noise characterization technique involves a very narrow bandwidth, centered on a frequency *f*. The current noise spectral components within this bandwidth have a certain mean square value. The ratio of this value to the bandwidth, as the latter is allowed to approach zero,

tends to what is called the power spectral density (PSD) of the current noise, denoted by $S_i(f)$. This quantity has units of square amperes per hertz (A^2/Hz). Often the square root of the PSD is used instead, given by $A/\sqrt{Hz}$. For a noise voltage v_n, one can similarly define a PSD $S_v(f)$ as V^2/Hz or its square root in $V/\sqrt{Hz}$ [4].

The total mean square noise current within an arbitrary bandwidth extending from $f = f_1$ to $f = f_2$ can be found by summing the mean square values of the individual components within each sub-bandwidth Δf. More precisely, using the PSD concept, we have

$$\overline{i_n^2} = \int_{f1}^{f2} S_i(f)df \tag{7.1}$$

A similar result can be obtained for voltage noise. Detailed characterization and modeling techniques of low- [5] and high-frequency [6] noise in advanced MOSFET devices are available in the literature. In this chapter, we have presented the basic understanding of modeling noise in MOSFET devices.

7.2.1 Fundamental Sources of Noise

Random fluctuations in the current (or voltage) in a device are generated by some fundamental processes in the device. The various types of noise present in an electronic device include (1) thermal (Johnson/Nyquist) noise, (2) shot noise, (3) generation–recombination noise, (4) random telegraph signal (burst/popcorn) noise, and (5) $1/f$ or flicker noise. The detailed description of these sources can be found in the review articles [6,7]. This chapter presents only the basic models of the thermal and flicker noise in MOSFET devices.

7.2.2 Thermal Noise

7.2.2.1 Physical Mechanism of Thermal Noise

Thermal noise arises from the random thermal motion of electrons in a material. When an electron gets scattered, its velocity is randomized. Thus, at a particular instant, the number of electrons moving in a certain direction may be more than that in another direction and small net current flows. This current fluctuates in magnitude and direction, but the average over a long time is always zero. The PSD of thermal noise current in a material of resistance R and temperature T is not white or flat up to infinitely high frequencies. It exists in every resistive medium and is unavoidable. However, it may be minimized by proper circuit design technique. For instance, input matching techniques using reactive elements can be used to lower the noise

in amplifiers since reactive elements do not generate thermal noise. Also, system bandwidth should be kept as small as possible to pass the desired signal since unused portions of the bandwidth cause unnecessary noise.

In order to understand thermal noise in a MOSFET, we will discuss first the thermal noise model of a resistor. It is known that the thermal noise of a resistor is directly proportional to temperature *T*. The spectral noise power density $S_i(f)$ (mean square value of current per frequency bandwidth) of a resistor, *R*, can be given by [8]

$$S_i(f) = \frac{\overline{i^2}}{\Delta f} 4kT \frac{1}{R} \tag{7.2}$$

where:

k is the Boltzmann's constant

The equivalent circuit of the thermal noise can be represented by a shunt current source $\overline{i^2}$, as shown in Figure 7.1.

The thermal noise characteristics in a MOSFET operating in strong inversion region have been studied for over two decades. The origin of thermal noise in a MOSFET has been found to be related to the random thermal motion of carriers in the channel of the device [9]. Depending on this understanding, noise models have been developed and implemented in circuit simulators [4]. Even though using the thermal noise model of a resistor can qualitatively explain the thermal noise in a MOSFET, it is not quantitatively accurate even at low drain bias [10,11]. Furthermore, as the moderate inversion region becomes important for low power applications, there is an increasing need for accurate noise modeling in this region. Therefore, the noise behavior of a transistor should be well modeled from strong inversion through moderate inversion, into weak inversion.

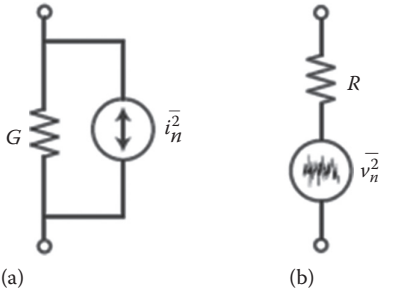

FIGURE 7.1
Equivalent circuit for the thermal noise of a resistor: (a) noise power as a current source in shunt with mean square value $\overline{i_n^2}$ and (b) noise power as a voltage source with mean square value $\overline{v_n^2}$.

7.2.2.2 Thermal Noise Model

Basic Model: The thermal noise model originally implemented in SPICE2 (Simulation Program with Integrated Circuit Emphasis) [12] is given by

$$Si_d(f) = \frac{8kT}{3} g_m \tag{7.3}$$

where:
g_m is the gate transconductance of the device

Equation 7.3 is found to be inadequate in the linear region, especially when $V_{ds} = 0$, where the transconductance is zero so that the calculated noise density is zero. However, in reality, the noise power density is not zero. To resolve this problem, the SPICE2 noise model is modified to the following form:

$$Si_d(f) = \frac{8kT}{3}\left(g_m + g_{ds} + g_{mb}\right) \tag{7.4}$$

where:
g_{ds} and g_{mb} are the output conductance and the bulk transconductance, respectively

Advanced Thermal Noise Model: An advanced thermal noise model is used in the industry standard compact modeling tools. We will derive the thermal noise model following the steps described by Tsividis and McAndrew [4]. It is well known that PSD of the noise voltage, generated across a resistor of value R, is $4kTR$ [2]. If a small element in the MOSFET channel has a resistance, ΔR, the noise voltage power of this element is

$$\overline{\left(\Delta v_t\right)^2} = 4kT\Delta R\Delta f \tag{7.5}$$

Assuming the length of the small element of the channel is Δy, then ΔR is

$$\Delta R = \frac{\Delta y}{\mu W_{eff} Q_i} \tag{7.6}$$

where:
W_{eff} is the effective channel width
μ is the electron mobility
Q_i is the channel charge per unit area

Substituting Equation 7.6 into Equation 7.5 gives

$$\overline{\left(\Delta v_t\right)^2} = \frac{4kT\Delta f\Delta y}{\mu W_{eff} Q_i} \tag{7.7}$$

The current change caused by the voltage change Δv_t in a device of effective channel length, L_{eff} is given by

$$\Delta i_t = \frac{W_{eff}}{L_{eff}} \mu Q_i \Delta v_t \tag{7.8}$$

Then the mean square value of Δi_t is

$$\overline{(\Delta i_t)^2} = \left(\frac{W_{eff}}{L_{eff}} \mu Q_i \right)^2 \overline{(\Delta v_t)^2} \tag{7.9}$$

Substituting Equation 7.7 into Equation 7.9, we have

$$\overline{(\Delta i_t)^2} = 4kT \frac{W_{eff}}{L_{eff}^2} \mu Q_i \Delta y \Delta f \tag{7.10}$$

The total noise current power in a bandwidth Δf can be obtained by integrating Equation 7.10 along the channel

$$\begin{aligned} \overline{(\Delta i_t)^2} &= 4kT\Delta f \frac{\mu}{L_{eff}^2} \int_0^{L_{eff}} Q_i W_{eff} dy \\ &= 4kT \frac{\mu}{L_{eff}^2} Q_I \Delta f \end{aligned} \tag{7.11}$$

where:

Q_I is the total inversion layer charge in the channel

Then from Equation 7.11, the PSD of thermal noise in a MOSFET can be expressed as

$$Si_d(f) = \frac{\overline{\Delta i_t^2}}{\Delta f} = 4kT \frac{\mu}{L_{eff}^2} Q_I \tag{7.12}$$

Equation 7.12 can be used for modeling thermal noise in MOSFETs with model-specific computation of the total inversion charge Q_I. Equation 7.12 shows that the channel thermal noise PSD is independent of frequency where the assumption of QS behavior is valid. It is observed from Equation 7.12 that the thermal noise increases with the increasing gate voltage V_{gs} since Q_I increases with V_{gs} and it increases with the decreasing channel lengths [6]. However, the experimental data show that thermal noise depends weakly on the drain voltage V_{ds}. This indicates that the noise contribution from the velocity saturation region of the channel is negligible. This is theoretically justified from the thermal noise model since Q_I saturates in the velocity saturation region.

7.2.3 Flicker Noise

7.2.3.1 Physical Mechanism of Flicker Noise

The low-frequency noise, commonly referred to as the *flicker noise* or 1/*f noise*, is characterized by 1/*f* dependency of its spectral density. There has been a continuous effort to understand the physical origin of flicker noise [13–18], leading to three different theories: (a) carrier density fluctuation model [19], (b) mobility fluctuation model [20], and (c) correlated carrier and mobility fluctuation model or the unified theory [21]. In the *carrier density fluctuation* model, the noise is explained by the fluctuation of channel free carriers due to the random capture and emission of carriers by interface traps at the $Si\text{-}SiO_2$ interface. According to this model, the input noise is independent of the gate bias, and the magnitude of the noise spectrum is proportional to the density of the interface traps. A 1/*f* noise spectrum is predicted if the trap density is uniform in the oxide. The experimental results show a $1/f^{\gamma}$ spectrum where the value of γ is in the range of $0.7 < \gamma < 1.2$ [17,22]. Experimental data also show that γ decreases with increasing gate bias in *p*-channel MOSFETs [23]. These experimental results are explained using modified charge density fluctuation model whereas the technology and the gate bias dependence of γ are explained assuming nonuniform spatial distribution of active traps in the oxide [19,23]. In the *mobility fluctuation model*, the flicker noise is considered to be the result of fluctuations in the carrier mobility given by Hooge's empirical relation for the spectral density of the flicker noise in a homogeneous device [24]. It has been proposed that the fluctuation of the bulk mobility in MOSFETs is induced by changes in phonon population [25]. The mobility fluctuation models predict a gate bias–dependent noise. However, they cannot always account for the magnitude of the noise [26]. In the *unified theory*, the origin of 1/*f* noise is assumed to be due to the capture and emission of carriers by the interface traps causing fluctuation in both the carrier number and mobility [21]. The unified theory can explain most of the experimental data and has been implemented in most compact-model extraction and circuit CAD tools [27–29].

7.2.3.2 Flicker Noise Model

The basic flicker noise model implemented in SPICE2 is given by

$$S_{id}(f) = \frac{K_F \cdot I_{ds}^{AF}}{f^{EF} C_{ox} L^2} \tag{7.13}$$

where:

S_{id} is the drain current noise PSD
I_{ds} is the drain current
AF is the flicker noise exponent
EF is the flicker noise frequency coefficient
K_F is the flicker noise coefficient

The basic model given by Equation 7.13 is not adequate for the characterization of noise in advanced MOSFET devices. Therefore, in this section, the unified flicker noise model that explains most of the observed behavior of low-frequency noise is presented.

Let us assume that y is the distance along the direction of channel length, z is the distance along the direction of channel width, and x is the coordinate along the direction of oxide thickness perpendicular to both the y and z directions.

For a section of channel width W_{eff} and length Δy in a MOSFET, fluctuations in the amount of trapped interface charge will introduce correlated fluctuations in the channel carrier concentration and mobility. The resulting fractional change in the local drain current can be expressed as [29]

$$\frac{\delta I_{ds}}{I_{ds}} = \left[\frac{1}{\Delta N}\frac{\delta \Delta N}{\delta \Delta N_t} \pm \frac{\delta \mu_{eff}}{\delta \Delta N_t}\right] \cdot \delta \Delta N_t \tag{7.14}$$

where:

$\Delta N = NW_{eff}\Delta y$

$\Delta N_t = N_t W_{eff}\Delta y$

N is the number of channel carriers per unit area

N_t is the number of occupied traps per unit area

The sign in front of the mobility term in Equation 7.14 is dependent on whether the trap is neutral or charged when filled [29]. The ratio of the fluctuations in the carrier number to the fluctuations in occupied trap number, $R_n = \delta\Delta N/\delta\Delta N_t$, is close to unity at strong inversion but assumes smaller values at other bias conditions [30]. A general expression for R_n is

$$R_n = \frac{\delta \Delta N}{\delta \Delta N_t} = -\frac{C_{inv}}{C_{ox} + C_{inv} + C_{dep} + C_{it}} \tag{7.15}$$

where:

C_{inv} is the inversion layer capacitance

C_{dep} is the depletion layer capacitance

C_{it} is the interface trap capacitance

The relationship between C_{inv} and N can be approximated as

$$C_{inv} \cong \frac{q}{v_{kT}} N \tag{7.16}$$

where:

v_{kT} is the thermal voltage

Thus, Equation 7.15 can be rewritten as,

$$R_n = -\frac{N}{N + N^*} \tag{7.17}$$

where

$$N^* = \frac{v_{kT}}{q}\left(C_{ox} + C_{dep} + C_{it}\right) \tag{7.18}$$

In order to evaluate $\delta\Delta\mu_{eff}/\delta\Delta N_t$, Matthiessen's rule is used so that

$$\begin{aligned}\frac{1}{\mu_{eff}} &= \frac{1}{\mu_n} + \frac{1}{\mu_{ox}} \\ &= \frac{1}{\mu_n} + \alpha N_t\end{aligned} \tag{7.19}$$

where:

μ_{ox} is the mobility limited by oxide charge scattering

α is the scattering coefficient and is a function of the local carrier density due to channel charge screening effect [31]

It can be shown from Equation 7.19 that

$$\frac{\delta\mu_{eff}}{\delta\Delta N_t} = -\frac{\alpha\mu_{eff}^2}{W_{eff}\Delta y} \tag{7.20}$$

Substituting Equations 7.15 and 7.20 in Equation 7.14 yields

$$\frac{\delta I_{ds}}{I_{ds}} = \left[\frac{R_n}{N} \pm \alpha\mu_{eff}\right].\frac{\delta\Delta N_t}{W_{eff}\Delta y} \tag{7.21}$$

Thus, PSD of the local current fluctuations is

$$S_{\Delta Id}(y,f) = \left(\frac{I_{ds}}{W_{eff}\Delta y}\right)^2\left(\frac{R_n}{N} \pm \alpha\mu_{eff}\right)^2 S_{\Delta Nt}(y,f) \tag{7.22}$$

where:

$S_{\Delta Nt}(y,f)$ is the PSD of the fluctuations in the number of the occupied traps over the area $W_{eff}\Delta y$

and is given by

$$S_{\Delta Nt}(y,f) = \int_{E_v}^{E_c}\int_0^{W_{eff}}\int_0^{T_{ox}} 4N_t(E,x,y,z)\Delta y f_t\left(1-f_t\right)\cdot\frac{\tau(E,x,y,z)}{1+\omega^2\tau(E,x,y,z)}dx\cdot dz\cdot dE \tag{7.23}$$

where:

$N_t(E,x,y,z)$ is the distribution of the traps in the oxide and over the energy band

$\tau(E,x,y,z)$ is the trapping time constant

$f_t = \left\{1 - \exp\left[\left(E - E_{fn}\right)/kT\right]\right\}^{-1}$ is the trap occupancy function where E_{fn} is the electron quasi-Fermi level
$\omega = 2\pi f$ is the angular frequency
T_{ox} is the oxide thickness
$E_c - E_v$ is the silicon energy gap

In order to evaluate the integral in Equation 7.23, the following assumptions are used:

1. The oxide traps have a uniform spatial distribution near the interface, that is, $N_t(E,x,y,z) = N_t(E)$
2. The probability of an electron penetrating into the oxide decreases exponentially with the distance from the interface

   As a result the trapping time constant is given by

$$\tau = \tau_0(E)\exp\left(\gamma \cdot x\right) \tag{7.24}$$

   where:

   $\tau_0(E)$ is the time constant at the interface
   γ is the attenuation coefficient of the electron wave function in the oxide

Since $f_t\left(1 - f_t\right)$ in Equation 7.23 behaves like a delta function around the quasi-Fermi level, the major contribution to the integral is from the trap level around E_{fn}. Thus, N_t(E) can be approximated by $N_t(E_{fn})$ and taken out of the integral. Replacing $f_t\left(1 - f_t\right)$ in Equation 7.23 by $kT\left(df_t/dE\right)$ and carrying out the integration yields

$$S_{\Delta Id}(y,f) = N_t\left(E_{fn}\right)\frac{kTW_{eff}\Delta y}{\gamma \cdot f} \tag{7.25}$$

The total drain current noise power spectrum density can be derived as

$$\begin{aligned} S_i(f) &= \frac{1}{L_{eff}^2}\int_0^{L_{eff}} S_{\Delta I_{ds}}(y,f)\Delta y dy \\ &= \frac{kTI_{ds}^2}{\gamma \cdot fW_{eff}L_{eff}^2}\int N_t\left(E_{fn}\right)\left[\frac{R_n}{N(y)} + \alpha\mu_{eff}\right]^2 dy \\ &= \frac{qkTI_{ds}\mu_{eff}}{\gamma \cdot fL_{eff}^2}\int_0^{V_{ds}} N_t\left(E_{fn}\right)\left[1 \pm \frac{\alpha\mu_{eff}N}{R_n}\right]^2 \frac{R_n^2}{N}dV \end{aligned} \tag{7.26}$$

Since α and μ_{eff} are functions of the local carrier density N, Equation 7.26 can be expressed as

$$S_i(f) = \frac{qkTI_{ds}\mu_{eff}}{\gamma \cdot f \cdot L_{eff}^2} \int_0^{V_{ds}} N_t^*\left(E_{fn}\right)\frac{R_n^2}{N}dV \tag{7.27}$$

where:

$N_t^*\left(E_{fn}\right)$ is the equivalent oxide-trap density that produces the same noise power if there were no contributions from the mobility fluctuations

In the present model $N_t^*\left(E_{fn}\right)$ is approximated as a three-parameter function of the channel carrier density such as

$$N_t^*\left(E_{fn}\right) = A + BN + CN^2 \tag{7.28}$$

where:

A, B, and C are technology-dependent model parameters

Using Equation 7.27, the flicker noise power spectrum density in the different operation regions of MOSFETs can be found.

1. Linear region in strong inversion ($V_{gs} > V_{th}$ and $V_{ds} < V_{dsat}$)

   In the strong inversion region, the charge density of carrier is given by

   $$qN(y) = C_{ox}\left[V_{gs} - V_{th} - \alpha V(y)\right] \tag{7.29}$$

   Thus, we have

   $$qN_0 = qN(0) = C_{ox}\left[V_{gs} - V_{th}\right] \tag{7.30}$$

   $$qN_L = qN\left(L_{eff}\right) = C_{ox}\left[V_{gs} - V_{th} - \alpha V_{ds}\right] \tag{7.31}$$

   where:

   N_0 and N_L are carrier densities at the source and drain ends of the channel, respectively

   Now, using Equations 7.30 and 7.31, Equation 7.27 can be rearranged as

   $$S_i(f) = \frac{q^2kTI_{ds}\mu_{eff}}{\gamma \cdot f \cdot L_{eff}^2 C_{ox}} \int_{N_L}^{N_0} N_t^*\left(E_{fn}\right)\frac{R_n^2}{N}dN \tag{7.32}$$

   Substituting Equations 7.15 and 7.28 into Equation 7.32 and performing the integration, we can show

   $$S_i(f) = \frac{q^2kTI_{ds}\mu_{eff}}{\alpha\gamma \cdot f \cdot L_{eff}^2 C_{ox}}\left[A\ln\left(\frac{N_0 + N^*}{N_L + N^*}\right) + B\left(N_0 - N_L\right) + \frac{1}{2}C\left(N_0^2 - N_L^2\right)\right] \tag{7.33}$$

2. Saturation region in strong inversion ($V_{gs} > V_{th}$ and $V_{ds} \geq V_{dsat}$)
   In the saturation region, the channel can be divided into two parts: one part is from the source to the velocity-saturation or *pinched-off* point discussed in Chapter 4 and the other part is from the velocity-saturation point to the drain end given by the length l_i. Then the total flicker noise PSD in the saturation region can be shown as

$$S_i(f) = \frac{q^2 kTI_{ds}\mu_{eff}}{\alpha\cdot\gamma\cdot f\cdot L_{eff}^2 C_{ox}}\cdot\left[A\ln\left(\frac{N_0+N^*}{N_L+N^*}\right)+B\left(N_0-N_L\right)+\frac{1}{2}C\left(N_0^2-N_L^2\right)\right] + \frac{kTI_{ds}^2\mu_{eff}}{\gamma\cdot f\cdot L_{eff}^2 W_{eff}}\frac{A+BN_L+CN_L^2}{\left(N_L+N^*\right)^2}l_i \tag{7.34}$$

   where:

$$N_L = \frac{1}{q}C_{ox}\left(V_{gs}-V_{th}-\alpha V_{dsat}\right)$$

3. Subthreshold region ($V_{gs} < V_{th}$)
   From Equations 4.112 and 4.117, we can show that the channel charge density in the subthreshold region is given by

$$qN(V) = C_d v_{kT}\exp\left[\frac{\left(V_{gs}-V_{th}\right)}{nv_{kT}}-\frac{V}{v_{kT}}\right] \tag{7.35}$$

   where:
   n is the subthreshold swing factor
   V_{th} is the voltage when the surface potential is equal to $2\phi_B$

   Substituting Equation 7.35 into Equation 7.27 and rearranging, we get the following expression for the spectral flicker noise power density in the subthreshold region:

$$Si(f) = \frac{q^2 v_{kT}^2 I_{ds}^2\mu_0}{\gamma\cdot f\cdot L_{eff}^2}\int_{N_L}^{N_0}\frac{N_t^*(E_{fn})}{\left(N+N^*\right)^2}dN \tag{7.36}$$

   where

$$qN_0 = C_d v_{kT}\exp\left(\frac{V_{gs}-V_{th}}{nv_{kT}}\right)$$
$$qN_L = qN_0\left[1-\exp\left(-\frac{V_{ds}}{v_{kT}}\right)\right] \tag{7.37}$$

In the subthreshold region, it is reasonable to assume that $N << N^*$ and $N^*_t = A + BN + CN^2 \cong A$. Thus, the flicker noise power in the subthreshold region can be simplified to

$$Si(f) = \frac{Aqv_{kT}I_{ds}^2}{\gamma \cdot f \cdot W_{eff}L_{eff}N^{*2}} \tag{7.38}$$

However, the flicker noise model of *p*-channel MOSFETs is not clear, especially in the weak inversion region [32].

Figure 7.2 shows the measured and simulated low-frequency noise behavior for both nMOSFET and pMOSFET devices of a typical 130 nm CMOS (complementary metal-oxide-semiconductor) technology.

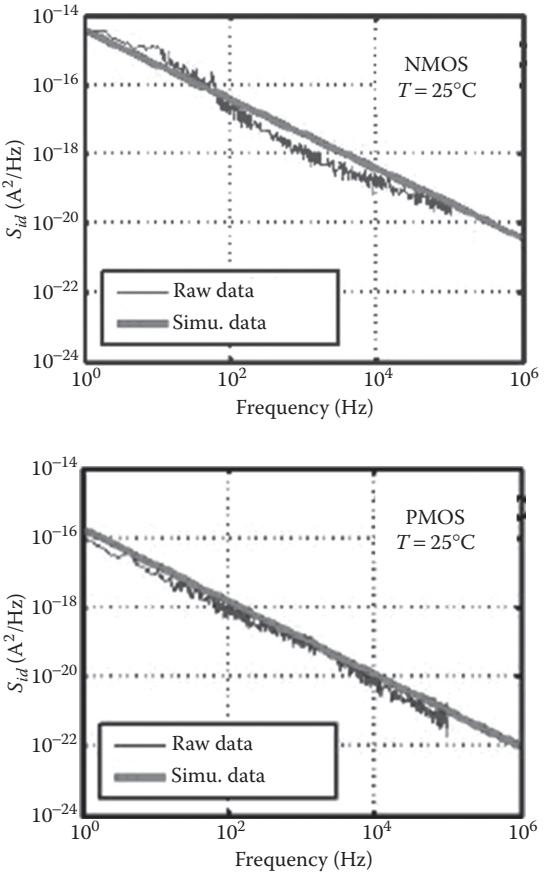

FIGURE 7.2
Simulated and measured low-frequency noise characteristics of nMOSFET and pMOSFET devices as a function of frequency with L = 0.28 μm and W= 10 μm; the data are obtained under the biasing condition V_{gs} = 1.25 V and V_{ds} = 2.5 V at room temperature.

Figure 7.3 shows noise PSD as a function of frequency. The plot shows thermal noise and flicker noise for both nMOSFETs and pMOSFETs.

Figure 7.4 shows the small-signal equivalent circuit of a MOSFET device with different noise sources and parasitic resistance-capacitance components.

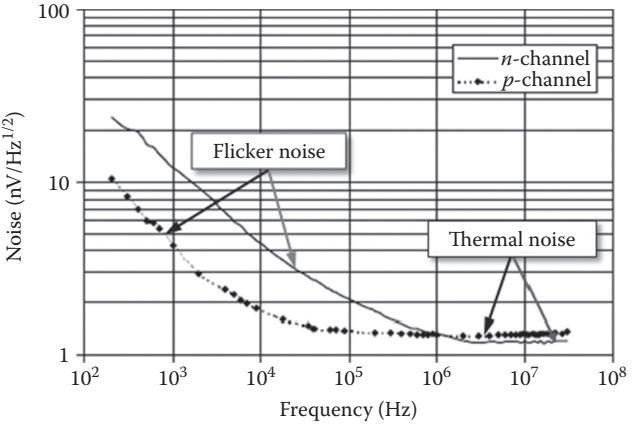

FIGURE 7.3
Noise spectra of an nMOSFET and a pMOSFET device measured at I_{ds} = 500 μA at room temperature; dimensions of both devices are W = 2000 μm and L = 0.5 μm.

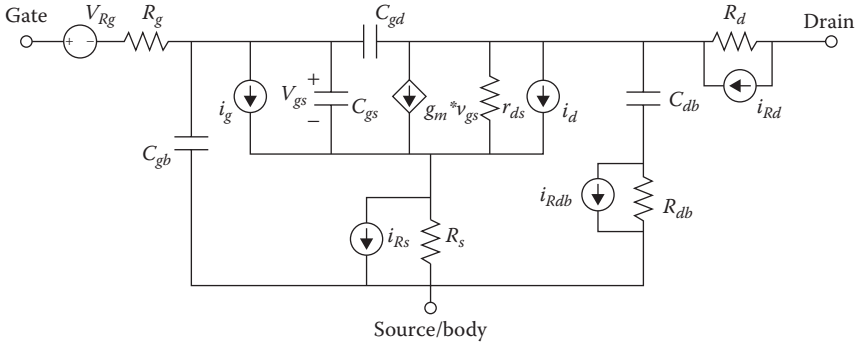

FIGURE 7.4
Small signal equivalent circuit model of a MOSFET suitable for RF and microwave frequencies including the noise sources ignoring body effect; v_{Rg}, i_{Rs}, i_{Rd}, and i_{Rdb} are the noise sources that model thermal noise from parasitic resistaces R_g, R_s, R_d, and R_{db}, respectively; i_g and i_d are the noise sources that model the gate current noise and channel noise (thermal and flicker noise), respectively; C_{gs}, C_{gd}, C_{gb}, and C_{db} are the gate-source, gate-drain, gate-body, and drain-body capacitances, respectively; and i_g, i_d, g_m, and v_{gs} are small-signal parameters. (Data from M.J. Deen et al., *IEEE Trans. Electron Dev.*, 53, 2062–2081, 2006.)

7.3 NQS Effect

In Chapter 6, the charge and C–V models are derived based on the QS approximation; that is, the inversion charge responds to the changes in the applied signal instantaneously. However, the QS approximation breaks down when signal changes occur on a timescale comparable to the device transit time.

As very-large-scale-integrated (VLSI) circuits become more performance-driven, it is sometimes necessary to predict the device performance for operation near the device transit time. However, as discussed in Chapter 5, most models available in SPICE use the QS approximation [33]. In a QS model, the channel charge is assumed to be a unique function of the instantaneous biases; that is, the charge has to respond to change in voltages with infinite speed. Thus, the finite charging time of the carriers in the inversion layer is ignored. In reality, the carriers in the channel do not respond to the signal immediately; thus, the channel charge is not a unique function of the instantaneous terminal voltages (QS) but a function of the history of the voltages (NQS). This problem may become pronounced in the RF applications, or when V_{gs} is close to V_{th}, or when long channel devices coexist with deep submicron devices as in many mixed signal circuits. In these circuits, the input signals may have rise or fall times comparable to or even smaller than the channel transit time. For long channel devices, the channel transit time is approximately inversely proportional to $(V_{gs} - V_{th})$ and directly proportional to L^2. Since the carriers in these devices cannot follow the changes of the applied signal, the QS models may give inaccurate or anomalous simulation results that cannot be used to guide circuit design. Two-dimensional (2D) numerical simulation results show that the most common QS model that uses 40/60 drain/source charge partitioning [34] results in an unrealistic large drain current spike during a fast turn-on [35].

Besides affecting the accuracy of simulation, the nonphysical results can also cause oscillation and convergence problems in the numerical iterations in circuit CAD. It is common among circuit designers to circumvent the convergence problem by using a 0/100 drain/source charge partitioning ratio [36], which attributes all transient charges to the source side. However, the numerical device simulation results show that this nonphysical solution merely shifts the current-spike problem to the source current; thus it only works when the source is grounded.

Moreover, none of these QS models can be used to accurately predict the high-frequency transadmittance of a MOSFET as pointed out by Tsividis and Masetti [37]. It is a common practice in high-frequency circuit designs to break a long channel MOSFET into N equal parts in series (N-lumped model) due to the lack of NQS models. The accuracy increases with N, at the expense of simulation time [38]. However, this method becomes

impractical when the device channel length is small because the short channel effects in the subtransistors may be activated. As an example, it is found that the results of modeling a 200 μm long MOSFET device in strong inversion (saturation) is completely different from modeling two 100 μm long MOSFETs in series.

It has been found that for RF applications the NQS model is necessary to fit the measured high-frequency characteristics of devices with even short channel length where the operation frequency is above 1 GHz [39,40]. Thus, NQS model is desirable in some mixed signal IC (integrated circuit) and RF applications. Therefore, a compact model that accounts for the NQS effect is highly desirable. Some NQS models based on solving the current continuity equation have been proposed [41–43]. They are complex and require long simulation time, making them unattractive for use in circuit simulation. In this chapter, an NQS model based on the Elmore-equivalent resistance-capacitance (RC) circuit is described [35]. It uses a physical relaxation time approach to account for the finite channel charging time. This NQS model applicable for both the large signal transient and small signal AC analysis is discussed in the next section.

7.3.1 Modeling NQS Effect in MOSFETs

Typically, the channel of a MOSFET is analogous to a bias-dependent RC-distributed transmission line [44]. In QS approach, the gate capacitors are lumped with the intrinsic source and drain nodes [35]. This ignores the fact that the charge build-up in the center portion of the channel does not follow a change in V_g as readily as it does at the source or drain edge of the channel. Breaking the transistor into N devices in series offers a good approximation for the RC network but has the disadvantages discussed in Section 7.3. A physical and efficient approach to model the NQS effect would be to formulate an estimate for the delay time through the channel RC network, and incorporate this time constant into the model equations.

One of the most widely used methods to approximate the RC delay was proposed by Elmore [45], considering the mean, or the first moment, of the impulse response. Utilizing Elmore's approach, the RC distributed channel can be approximated by a simple RC equivalent circuit that retains the lowest frequency pole of the original RC network. The new equivalent circuit is shown in Figure 7.5. The Elmore resistance (R_{Elmore}) in strong inversion is calculated from the channel resistance and is given by [35].

$$\begin{aligned} R_{Elmore} &= \frac{L_{eff}}{E_{LM}\mu W_{eff} Q_i} \\ &= \frac{L_{eff}}{E_{LM}\mu W_{eff} C_{ox}\left(V_{gs} - V_{th}\right)} \end{aligned} \tag{7.39}$$

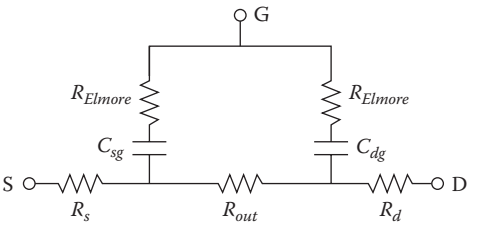

FIGURE 7.5
Transisent and small signal equivalent circuit model for a MOSFET device. (Data from M. Chan et al., *IEEE Trans. on Electron Dev.*, 45, 834–841, 1998.)

where:

Q_i is the amount of channel inversion charge per unit area

E_{LM} is a parameter, referred to as the Elmore constant (=5) that is used to match the lowest frequency pole

It is reported that the value of E_{LM} required to match the output of different possible equivalent circuits is about 3 and it is invariant with respect to W and L [35]. The comparison of the time and frequency domain responses of the Elmore equivalent network shown in Figure 7.5 with the conventional distributed channel representation of a device shows a reasonable match between the Elmore's equivalent circuit and the distributed RC network [35]. However, direct implementation of the model shown in Figure 7.5 requires two additional nodes that increase the computational time. In addition, the change in the device topology may require modifications of the existing compact model formulations. Therefore, to improve the computational efficiency within the framework of an existing compact model, simplifying assumptions are made to develop a simplified NQS model. For example, we assume that the bulk charging current is negligibly small. Then the gate, drain, and source terminal currents can be described by the expression

$$I_j(t) = I_j(t)\big|_{DC} + j_{xpart}\frac{dQ_i(t)}{dt}; \quad \text{where } j = G, D, S \tag{7.40}$$

where:

$I_j(t)$ represents the total gate, drain, and source currents

$I_j(t)\big|_{DC}$ represents the DC gate, drain, and source currents

$Q_i(t)$ is the actual channel charge at any given time t

j_{xpart} represents the channel charge partitioning ratios [46,47] for the gate (G_{xpart}), drain (D_{xpart}), and source (S_{expart})

so that

$$D_{xpart} + S_{xpart} = -G_{xpart} = 1 \tag{7.41}$$

In the 40/60 partitioning scheme, D_{xpart} varies from 0.5 at $V_d = 0$ to 0.4 in the saturation region, and S_{xpart} varies from 0.5 to 0.6 [33,48]. However, since the 0.4/0.6 scheme covers a wide range of voltage and the error introduced by using a constant $D_{xpart} = 0.4$ and $S_{xpart} = 0.6$ is less than 5%, these values can be adopted to simplify the model.

In the QS approach, it is assumed that

$$\begin{aligned}\frac{dQ_i(t)}{dt} &= \frac{dQ_{iq}(t)}{dt} \\ &= \frac{dQ_{iq}}{dV}\frac{dV}{dt}\end{aligned} \tag{7.42}$$

where:

$Q_{iq}(t)$ is the equilibrium, or QS, channel charge under the instantaneous bias at any time t

The assumption of equilibrium at all times gives an error in calculating the NQS currents. To account for the NQS current, a new state variable Q_{def} is introduced to keep track of the amount of deficit (or surplus) channel charge relative to the QS charge at a given time.

$$Q_{def}(t) = Q_{iq}(t) - Q_i(t) \tag{7.43}$$

and

$$\frac{dQ_{def}(t)}{dt} = \frac{dQ_{iq}(t)}{dt} - \frac{dQ_i(t)}{dt} \tag{7.44}$$

Q_{def} is allowed to decay exponentially to zero after a step change in bias with a bias-dependent NQS relaxation time, τ. Thus, the charging current can be approximated by

$$\frac{dQ_i(t)}{dt} \cong \frac{dQ_{def}(t)}{\tau} \tag{7.45}$$

$Q_{def}(t)$ can be calculated from Equation 7.44, and a subcircuit, shown in Figure 7.6, has been introduced to obtain the solution. The subcircuit is a direct translation from Equation 7.44. The node voltage gives the value of $Q_{def}(t)$. The total charging current is given by the current going through the resistor of value τ. With this approach, only one additional node is needed and the topology of the original transistor model is not affected.

The value of the channel relaxation time constant τ is composed of the terms related to the diffusion and drift currents (calculated from the RC Elmore equivalent circuit discussed earlier). The components of τ are given by

$$\tau_{diffusion} = \frac{\left(L_{eff}/4\right)^2}{\mu v_{kT}} \tag{7.46}$$

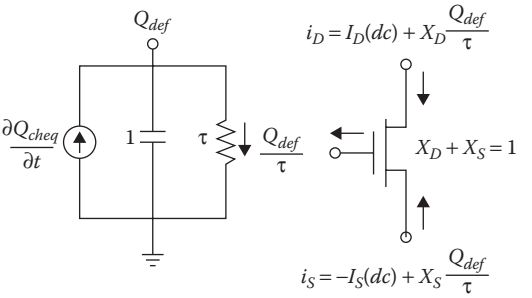

FIGURE 7.6
A simplified representation of Elmore's equivalent circuit for modeling NQS effect in MOSFETs. (Data from M. Chan et al., *IEEE Trans. on Electron Dev.*, 45, 834–841, 1998.) In text, $Q_{cheq} \equiv Q_{iq}$.

$$\tau_{drift} = \frac{1}{2} R_{Elmore} C_{ox} W_{eff} L_{eff} \tag{7.47}$$

$$\frac{1}{\tau} = \frac{1}{\tau_{diffusion}} + \frac{1}{\tau_{drift}} \tag{7.48}$$

In order to derive a simplified NQS model, C_{ox} is used in Equation 7.47 instead of the bias-dependent parameters (C_{sg} + C_{dg}). It is found that the simulated relaxation time τ obtained by Equation 7.48 agrees very well with that obtained by a 2D numerical device simulation under various biasing conditions [35].

In reality, NQS effects are important for long channel devices driven by fast switching inputs. However, NQS effects have also been observed in short channel devices [39,40]. When a MOSFET is operated in the velocity saturation regime, the channel conductivity is reduced, thus increasing the value of τ [35]. The circuit simulation data using NQS model show that the error resulting from the velocity saturation effect in a MOSFET is usually less than 20%. This error can be further reduced by optimizing Elmore constant to achieve an acceptable simulation data for both linear and saturation regions. However, accurate results can be obtained using an empirical model for the relaxation time [35] such as

$$\tau'_{drift} = \begin{cases} \tau_{drift}\left(1 + \frac{3}{8}\left(\frac{V_{ds}}{V_{dsat}}\right)^2\right); & \text{for } 0 < \left(V_{gs} - V_{th}\right) \geq V_{ds} \\ 1.375\tau_{drift}; & \text{for } 0 < \left(V_{gs} - V_{th}\right) \leq V_{ds} \end{cases} \tag{7.49}$$

Again, the simulated relaxation time obtained by empirical expressions in Equation 7.49 and 2D numerical device simulation agree very well [35].

For the simplicity of NQS modeling, it is assumed that the bulk charging current is zero [35]. This assumption is justified since for most applications

the NQS effect from the bulk charge is negligible and does not significantly impact small signal simulation. However, the body current can be included by partitioning Q_{def} between the gate and the body [35,49].

7.4 Modeling Parasitic Elements for RF Applications

With the continuous scaling down of CMOS technologies to the nanoscale regime, RF circuits are realized in a standard CMOS process [50]. Therefore, a compact model for circuit CAD that is valid for a broad range of bias conditions, device sizes, and operating frequencies is of utmost importance. The widely used RF modeling approach is to build subcircuits based on MOSFET models that are suitable for analog/digital applications [40,51–55]. In the subcircuit, parasitic elements around gate, source, drain, and substrate as shown in Figure 7.7 are added to improve the accuracy of the model at high frequencies [40]. An important part of RF modeling is to establish physical and scalable model equations for the parasitic elements at the source, drain, gate, and substrate. The scalability of the intrinsic device is ensured by the core model library developed using the target compact model discussed in Chapters 4 and 5. We will now discuss the techniques to model the gate and substrate resistances for RF and analog applications.

7.4.1 Modeling Gate Resistance

At any low frequency, the gate resistance of a MOSFET can be calculated from the sheet resistance of the gate material and is given by

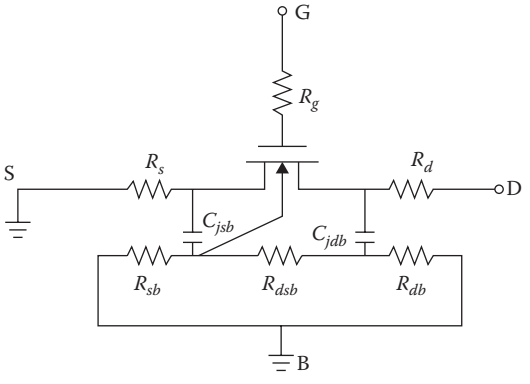

FIGURE 7.7
A subcircuit with parasitic elements added to an intrinsic MOSFET model for RF analysis. (Data from Y. Cheng et al., *Proceedings of the ICSICT,* 416–419, 1998.)

$$R_g = \frac{W_{eff}}{L_{eff}} \rho_{sh,gate} \tag{7.50}$$

where:

W_{eff} and L_{eff} are the effective channel width and channel length of the device, respectively

$\rho_{sh,gate}$ is the gate sheet resistance per square [Chapter 2, Section 2.2.5.3]. The typical sheet resistance for a polysilicon gate ranges between 20 and 40 Ω per square and is significantly lower for silicide as well as for metal stack processes

At high frequencies, the accurate modeling of the gate resistance is very complex due to the distributed transmission-line effect. Therefore, a lumped equivalent gate resistance α times the end-to-end gate resistance given in Equation 7.50 is used such that [56]

$$R_g = \frac{\alpha_g W_{eff}}{L_{eff}} \rho_{sh,gate} \tag{7.51}$$

where:

$\alpha_g = 1/3$ to account for the distributed RC effects when the gate electrode is contacted at one end and $\alpha_g = 1/12$ when the electrode is contacted on both ends [57]

It is found that the distributed RC effect of the gate as well as the NQS effect, that is, the distributed RC effect of the channel, affects the high-frequency characteristics of MOSFET devices. Therefore, an additional component of gate resistance must be considered to account for the distributed RC effect in the channel or NQS effect. Thus, at high-frequency operation of a MOSFET device, the distributed channel resistance *seen* by the signal applied to the gate also contributes to the effective gate resistance in addition to the resistance of the gate electrode. Then the effective gate resistance consists of two parts: the distributed gate electrode resistance ($R_{g,eltd}$) and the distributed channel resistance seen from the gate (R_{gch}), as shown in Figure 7.8 [58].

$$R_{g,eff} = R_{g,eltd} + R_{gch} \tag{7.52}$$

Typically, $R_{g,eltd}$ is insensitive to bias and frequency and, therefore, is obtained from the gate electrode sheet resistance ($\rho_{sh,geltd}$)

$$R_{g,eltd} = \rho_{sh,geltd} \frac{\alpha_g W}{L + \beta} \tag{7.53}$$

where:

α_g is 1/3 when the gate terminal is brought out from one side and 1/12 when connected on both sides

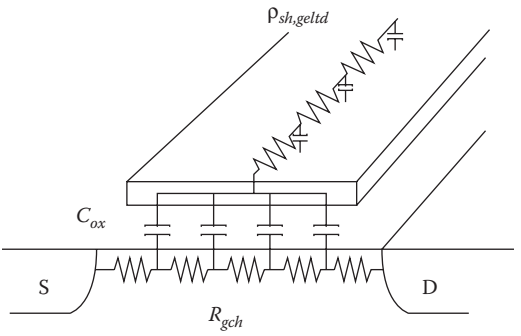

FIGURE 7.8
Distributed nature of gate electrode resistance, channel resistance network, and gate capacitance for modeling the efective gate resistance for RF analysis of MOSFET devices. (Data from X. Jin et al., *IEDM Technical Digest*, 961–964, 1998.)

L is the channel length
the parameter β models the external gate resistance

Now, R_{gch} includes the static channel resistance (R_{st}) that accounts for the DC channel resistance and the excess-diffusion channel resistance (R_{ed}) due to the change of channel charge distribution by the AC excitation of the gate voltage. Both R_{st} and R_{ed} together determine the time constant of the NQS effect. R_{st} is modeled by integrating the resistance along the channel under the QS assumption [58]

$$R_{st} = \int dR = \int \frac{dV}{I_d} = \begin{cases} \dfrac{V_{ds}}{I_{ds}}; & 0 < \left(V_{gs} - V_{th}\right) \geq V_{ds} \\ \dfrac{V_{dsat}}{I_{ds}}; & 0 < \left(V_{gs} - V_{th}\right) \leq V_{ds} \end{cases} \tag{7.54}$$

where:
$V_{dsat} \cong (V_{gs} - V_{th})$ is the saturation drain voltage

Both I_{dsat} and V_{dsat} can be calculated from the target compact model. And, R_{ed} can be derived from the diffusion current (Equation 4.122) as

$$R_{ed} = \frac{L_{eff}}{\eta W_{eff} \mu C_{ox} v_{kT}} \tag{7.55}$$

where:
η is a technology-dependent constant
C_{ox} is the gate capacitance as shown in Figure 7.8

Thus, the overall channel resistance seen from the gate is

$$\frac{1}{R_{gch}} = \gamma\left(\frac{1}{R_{st}} + \frac{1}{R_{ed}}\right) \tag{7.56}$$

where:

γ is a parameter accounting for the distributed nature of the channel resistance

It is shown that if the resistance is uniformly distributed along the channel, the value of γ is 12 [58]. However, this assumption is true only in the saturation region; therefore, γ is used as a fitting parameter. Thus, from Equations 7.53 and 7.56 (along with Equation 7.52), we can model the effective gate resistance network for RF application. The reported simulation results using the above-described model show very good agreement with the data obtained by numerical device simulation.

7.4.2 Modeling Substrate Network

Modeling of the substrate parasitic elements and substrate network is very important for RF applications of MOSFET devices [40,59]. In order to obtain the desired scalable RF model, it is critical to develop scalable model for each component of the substrate network. A three-resistance substrate network equivalent circuit shown in Figure 7.9 is used for modeling the substrate parasitic elements at high frequencies [40].

In Figure 7.9, C_{JSB} and C_{JDB} are the capacitances between the source-body and drain-body *pn*-junctions, respectively, R_{SB} and R_{DB} are the resistances between the source-body and drain-body to account for the resistive losses at the source-drain signal coupling, and R_{BDS} is the substrate resistance. Using a two-port substrate network, the above model has been verified using 2D-numerical device simulation [40,59].

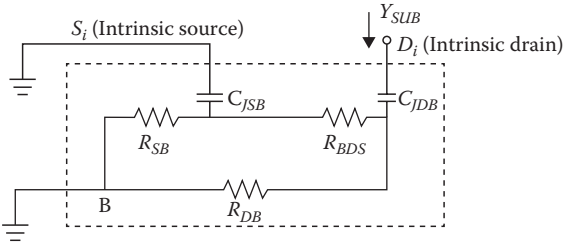

FIGURE 7.9
A three-resistance equivalent circuit for the substrate network: C_{JSB} and C_{JDB} are the capacitances between the source-body and drain-body *pn*-junctions, respectively.

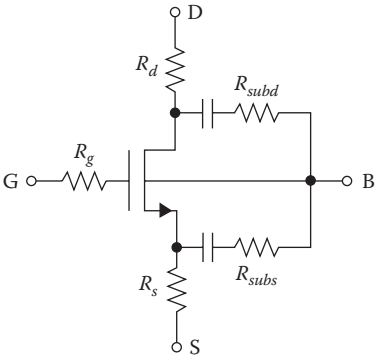

FIGURE 7.10
A MOSFET RF model developed used in conjunction with standard compact model. (Data from J.-J. Ou et al., *IEEE Symposium on VLSI Technology*, 94–95, 1998.)

7.4.3 MOSFET RF Model for GHz Applications

Using the lumped gate resistance and the substrate resistance network models described earlier, a unified compact device model can be realized for simulations of both RF and baseband analog circuits. The equivalent circuit model is shown in Figure 7.10 [54]. The RF MOSFET model can be realized with the addition of three resistors R_g, R_{subd}, and R_{subs} to the existing compact model. In Figure 7.10, R_g models both the physical gate resistance and the NQS effect, whereas R_{subd} and R_{subs} are the lumped substrate resistances between the S/D junctions and the substrate contacts. The values of R_{subd} and R_{subs} may not be equal as they are functions of the transistor layout [54].

7.5 Summary

In this chapter, the basic modeling techniques for analog and RF applications are presented. These approaches include modeling of noise, NQS effect, and the parasitic gate and substrate networks for high-frequency applications. The primary objective of this chapter is to expose readers to these RF models for the sake of completeness of the compact modeling activities and understanding of the advanced VLSI circuits. The readers would benefit from the basic understanding presented in this chapter to use the state-of-the-art RF compact models that are continuously developed and updated.

Exercises

7.1 Use the expression for the inversion charge from the MOS Level 1 model to calculate and plot the thermal noise PSD as a function of the channel length at room temperature for nMOSFET and pMOSFET devices of channel length, $100 < L < 1000$ nm; $W = 1000$ nm. Given that the surface mobilities for electrons and holes are 600 $cm^2\ V^{-1}\ sec^{-1}$ and 400 $cm^2\ V^{-1}\ sec^{-1}$, respectively.

7.2 Complete the mathematical steps to derive the flicker noise PSD given in Equation 7.27. Clearly state any assumptions you make.

7.3 Use the subthreshold region drain current expression (Equation 4.122) to derive the diffusion resistance in the channel given by Equation 7.55. Clearly state any of the assumptions you make.