Available online at www.sciencedirect.com
ScienceDirect
Solid-State Electronics 52 (2008) 899–908
SOLID-STATE ELECTRONICS
www.elsevier.com/locate/sse

Small-signal performance and modeling of sub-50 nm *n*MOSFETs with f_T above 460-GHz

V. Dimitrov [a], J.B. Heng [a], K. Timp [a], O. Dimauro [a], R. Chan [a], M. Hafez [a], J. Feng [a], T. Sorsch [b], W. Mansfield [b], J. Miner [b], A. Kornblit [b], F. Klemens [b], J. Bower [b], R. Cirelli [b], E.J. Ferry [b], A. Taylor [b], M. Feng [a], G. Timp [a,*]

[a] *University of Illinois at Urbana-Champaign, Electrical and Computer Engineering, 3041 Beckman Institute, 405 N Mathews, Urbana, IL 61801, United States*
[b] *New Jersey Nanotechnology Consortium, Murray Hill, NJ 07974, United States*

Received 18 May 2007; received in revised form 13 January 2008; accepted 29 January 2008
Available online 17 March 2008

The review of this paper was arranged by Prof. A. Zaslavsky

Abstract

We have fabricated and tested the performance of sub-50 nm gate *n*MOSFETs to assess their suitability for mixed signal applications in the super high frequency (SHF) band, i.e. 3–30 GHz. For a 30 nm × 40 μm × 2 device, we found $f_T = 465$ GHz at $V_{ds} = 2$ V, $V_g = 0.67$ V, which is the highest cut-off frequency reported for a MOSFET produced on bulk silicon substrate so far. However, our measurements of f_{max} and noise figure indicate that parasitics impose limitations on SHF operation. We also present a high frequency ac model appropriate to sub-50 nm gate length nanotransistors, which incorporates the effects of the parasitics. The model accurately accounts for measurements of the *S*- and *Y*-parameters in the frequency range from 1 to 50 GHz.

Keywords: Radio frequency (RF) MOSFET; RF MOSFET modeling; High frequency; Mixed signal

1. Introduction

Mobile, wireless communications are changing everything. Portable communication devices like the cell phone, along with 3 G, WLAN, and Bluetooth® are spurring the demand for high frequency, mixed signal integrated circuits that are inexpensive, reliable and have a long battery life. CMOS technology can satisfy these demands. The relentless scaling of CMOS toward nanometer-scale gate lengths has produced MOSFETs with digital and RF performance that is suitable for mixed signal applications.

The merit of a transistor depends on the circuit design. While large signal digital integrated circuits often use gate delay as a metric, the same loading conditions do not generally apply to RF circuits [1]. Three figures-of-merit appropriate to small-signal RF performance are the cut-off frequency associated with the short-circuit current gain,

$$f_T \approx g_m/2\pi(C_{gs} + C_{gd}) \sim 1/L_g^n \tag{1}$$

the maximum frequency of oscillation (where the unilateral power gain vanishes),

$$f_{max} \approx f_T/2\sqrt{R_g(g_{ds} + 2\pi f_T C_{gd})} \sim 1/(L_g^n W) \tag{2}$$

and the noise figure F_{min},

$$F_{min} \approx 1 + Kf\sqrt{g_m(R_g + R_s)}/f_T \tag{3}$$

where g_m is the transconductance, C_{gs} and C_{gd} are the gate-to-source and gate-to-drain capacitances, R_s is the

* Corresponding author. Tel.: +1 2172449629; fax: +1 2172446622.
E-mail address: gtimp@uiuc.edu (G. Timp).

0038-1101/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.sse.2008.01.025

parasitic source resistance, R_g is the (lumped) gate resistance, g_{ds} is the output conductance, W is the total gate width, n is an index that ranges from $1 < n < 2$ depending on the transistor model (short or long channel), and K is a constant that depends on the technology. Ostensibly, improvements in f_T follow from scaling of the gate length, L_g. There has been a progressive increase in f_T to 330 GHz for a 60 nm gate length *n*MOSFET [2–5], which is comparable to observations in sub-100 nm InP HEMTs [6], but still inferior to reports on SiGe [7] and InP/InGaAs HBTs [8–10]. On the other hand, in MOS technology, f_{max} generally lags behind f_T – the disparity can be accounted for by parasitics, [2,11–13] such as R_g and g_{ds} that are not optimized in a core CMOS manufacturing process. With acceptable gain, the noise figure, F_{min}, also has to be minimized to make effective use of CMOS technology for RF. Like f_T, F_{min} has been found to improve (diminish) with each technology generation [2], but just like f_{max}, F_{min} also depends on parasitic elements that are sensitive to the gate bias and geometry.

Along with the RF performance, another prerequisite for the implementation of mixed circuits in CMOS is accurate, high frequency models for the MOSFET [13–24] and passives. Specifically, the MOSFET model must accurately represent the power gain, input and output impedance and phase delay between the gate voltage and the drain current. A microwave table-based approach to modeling can be very accurate, but requires a large database obtained from numerous measurements and computationally intensive simulations – it becomes intractable for designing highly integrated CMOS communications circuits. Instead, a compact physics-based model is preferred, but a physics-based model has to be valid over a range of bias conditions, temperatures, and frequencies. Consequently, it has to account, not only for parasitic resistances and capacitances, but also for non-quasi-static (NQS) or distributed effects in the gate, substrate and channel resistances due to the channel propagation delay, and non-reciprocal capacitances that account for the different effect of the gate and drain on each other in terms of charging currents [19,25].

Here, we report on the fabrication, testing and modeling of the RF performance of sub-50 nm gate length *n*MOSFETs to assess their suitability for mixed signal applications in the super high frequency (SHF) band, i.e. 3–30 GHz. Using a conventional process flow suitable for a digital technology, we fabricated *n*MOSFETs with gate lengths as short as 30 nm, and then we measured the DC and RF performance. Following Rashkin [26–28] we extended the usual de-embedding methodology [29] to account for additional access capacitances associated with metal interconnections running between the (SOLT) reference plane and the contacts to the drain, source and gate. With this refinement, we extracted an $f_T = 465$ GHz from measurements of a 30 nm × 40 μm × 2(finger) *n*MOSFET taken in the frequency range $1\text{ GHz} < f < 50\text{ GHz}$ at $V_{ds} = 2$ V, $V_g = 0.67$. This is the highest cut-off frequency reported for a MOSFET so far. It represents a substantial improvement over previous extrapolations using the same transistors ($f_T = 290$ GHz) that do not account for the access capacitance [30]. However, our measurements of $f_{max} \leqslant 135$ GHz and noise figure $F_{min} = 0.9$ dB at 8 GHz indicate that parasitics still impose limitations on SHF operation. The equivalent circuit that we developed to model the RF performance is based on Tsividis's work [19] incorporating the effect of extrinsic and access parasitics and NQS. This model gives an accurate accounting of the measurements of the Y-parameters in the frequency range from 1 to 50 GHz and scales appropriately with the transistor layout.

2. Fabrication

We have produced sub-50 nm *n*MOSFETs using a conventional process flow suitable for digital applications. The layout for the unit cell for each RF-MOSFET uses a co-planar waveguide structure to facilitate the RF measurements with a gate that is folded into a multi-legged structure and contacted from one side of the channel. An example of a two-legged MOSFET is shown in Fig. 1a. The aspect ratio of the folded gate guarantees that the total gate resistance is determined by the length of the legs while the shared drain contact reduces the overall junction capacitance. Using this geometry, we expect f_{max} to scale closely with the length of the leg. An especially dramatic illustration of the performance that can be recovered through gate engineering like this is the recent demonstration by Tiemeijer et al. [12] of $f_{max} = 150$ GHz obtained using a 180 nm gate length CMOS foundry process that exhibits a cut-off frequency of only $f_T = 70$ GHz on a resistive 10 Ω-cm substrate.

A transmission electron micrograph of a cross-section through a representative MOSFET structure is shown in Fig. 1b. The transistors were fabricated using 90 nm thick LOCOS isolation on float-zone (>800 Ω-cm) wafers. A 30 keV boron implant with a fluence of 5.0E13 cm^{-2} was used to form the p-tub, and a 7 keV boron implant at a fluence of 1.5×10^{13} cm^{-2} was used to form the channel. Subsequently, a 1.3 nm thick gate oxide was grown using rapid thermal oxidation, which was followed by the deposition of a 95 nm thick in situ phosphorus-doped polysilicon layer. The 30–50 nm gate electrodes, defined by electron beam lithography in Sumitomo resist NEB31, were transferred into the poly using reactive ion etching without using a hard mask. The gate-etch stopped on the thin gate oxide without trenching. Following the definition of polysilicon gates, an arsenic implant at 2.5 keV with fluence of 3.0×10^{14} cm^{-2} was used to form the shallow source/drain (S/D) extensions. Subsequently, a 30 nm silicon nitride layer was deposited on top of a thin 10 nm silicon dioxide layer and subsequently etched to create the sidewall spacers. A 47 keV arsenic implant at 4×10^{15} cm^{-2} was then used to define the S/D. The doping was activated using a rapid thermal anneal for 1 s at 1050 °C in nitrogen with a 75 °C/s ramp. Finally,

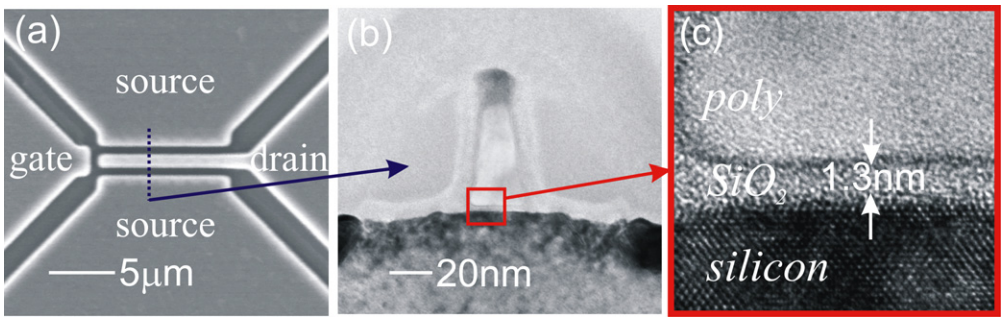

Fig. 1. (a) A scanning electron micrograph of a MOSFET tester (metal contacts to the source, gate, and drain shown here). The cross-section of the MOSFET (blue dashed line) is shown in (b). (b) A transmission electron micrograph of one finger of a nominally 30 nm gate length *n*MOSFET with a 1.3 nm thick gate oxide. The gate is comprised of heavily doped polysilicon 95 nm thick with a $CoSi_2$ strap. The 40 nm sidewalls consist of a 10 nm thick oxide beneath 30 nm of silicon nitride. A magnified view of the area outlined in red is shown in (c). The gate oxide appears to be about 1.3 nm thick. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

we used a cobalt-salicide process to lower the gate and S/D resistance to about 8 Ω/sq. The salicide consisted of the deposition of 5.0 nm of Cobalt followed by anneals for 30 s at 550 °C and for 30 s at 750 °C for the formation of a highly conductive $CoSi_2$ layer.

The back end of the process used 400 nm of dielectric (TEOS) to isolate the 0.75 μm aluminum co-planar waveguide structures (used for RF measurements) from the high-resistivity float-zone substrate. A 50 nm silicon nitride layer prior to the TEOS deposition was used to facilitate the etching of 0.5 μm vias by protecting the silicided gate and S/D extensions. After the vias to the S/D and the gate were etched, a 30 nm Ti/80 nm TiN layer was deposited and annealed at 700 °C for 20 s prior to aluminum deposition to serve as a diffusion barrier and aid the adhesion of the aluminum. Finally, before the metal contact lithography, 25 nm of TiN was deposited to cap the aluminum.

3. DC characteristics

We measured the dc current–voltage characteristics of MOSFETs fabricated this way. Typical results obtained for a gate length 45 nm × 20 μm × 2 (gate length × gate width × number of fingers) and $L_g = 30$ nm × 40 μm × 2 MOSFETs are shown respectively in Fig. 2a–c and d–f. For the 30 nm gate, we found a drive current $I_D = 1.9$ mA/μm at $V_{ds} = V_G = 2$ V and an off-state current of 310 μA/μm, measured at $V_G = 0$ V and $V_{ds} = 2$ V, giving an $I_{on}/I_{off} = 6$. The maximum transconductance in the saturation region is $g_m = 0.94$ mS/μm, which is found at $V_{ds} = 2$ V for $V_G = 0.67$ V. For the 45 nm device, we observe a drive current $I_D = 1.8$ mA/μm at $V_{ds} = V_G = 2$ V with an off-state current of 55 μA/μm, giving an $I_{on}/I_{off} = 33$. The maximum transconductance in the saturation region is $g_m = 1.1$ mS/μm at $V_{ds} = 2$ V and $V_G = 0.67$ V.

From the subthreshold characteristic shown in Fig. 2c and f, we infer that short channel effects like drain induced barrier lowering and channel length modulation affect the device performance. In particular, the drain characteristic indicates a finite output resistance that can be represented

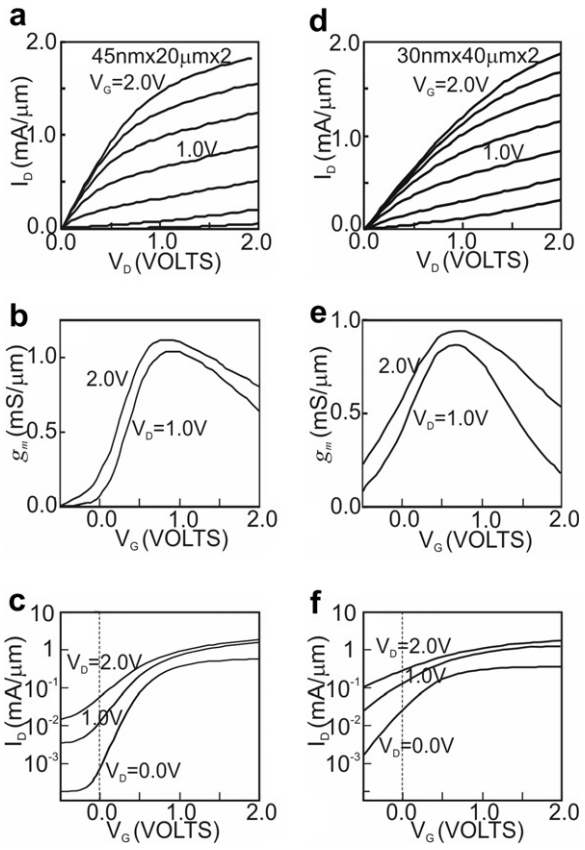

Fig. 2. The drain, transconductance, and subthreshold characteristics measured in 45 nm × 20 μm × 2 (a–c) and 30 nm × 40 μm × 2 (d–f) *n*MOSFETs, respectively. The drive current $I_D = 1.8$ mA/μm and 1.9 mA/μm at $V_{ds} = V_g = 2$ V for (a) and (d), respectively. The maximum transconductance in the saturation region is $g_m = 0.94$ mS/μm and 1.1 mS/μm at $V_{ds} = 2$ V and $V_g = 0.67$ V for (b) and (e), respectively.

by a lumped element of 30 Ω in the intrinsic model for the MOSFET see Table 1). The short channel effects are especially evident in the 30 nm device, which result in a reduced transconductance even though the gate length is smaller. Another detrimental factor may be the series resistance associated with the relatively thin (0.75 μm) aluminum layer used for the coplanar waveguide structure.

Table 1
Extracted MOSFET parameters from model (30 nm × Wμm × 2)

Intrinsic parameters	Quantity*				Extrinsic parameters	Quantity*			
	5 μm	10 μm	20 μm	40 μm		5 μm	10 μm	20 μm	40 μm
g_m (intrinsic) (mS)	36.4	75.7	89.5	305	R_{ge} (kΩ)	2.5	0.7	3.1	11.0
g_{mb} (intrinsic) (mS)	6.4	1.1	26.7	89.3	R_{se} (Ω)	40.5	21.4	7.0	3.1
g_{ds} (mS)	9.8	23.3	57.1	250	R_{de} (Ω)	0.02	0.01	0.62	1.0
C_{gdi}	0.1 fF	1.0 fF	0.1 fF	0.1 aF	R_{dse} (Ω)	613	129	33.8	918
C_{gsi} (fF)	9.4	10.0	20.5	47.3	L_{ge} (pH)	853	896	0.48	7.3
R_{gdi} (Ω)	6.2	4.7	0.1	1.0	C_{gbee} (fF)	5.4	9.5	7.1	10.9
R_{gdt} (MΩ)	99	137	180	40.0	C_{dsee} (fF)	16	20.4	56.8	190
R_{gsi}	1.9 kΩ	768 Ω	2.6 kΩ	4.8 kΩ	C_{da} (fF)	6.1	7.8	8.9	1.9
R_{gst} (kΩ)	25.7	29	50.1	41.1	$C_{gsee} + C_{ga}$ (fF)	3.9	5.8	7.1	6.0
R_1 (Ω)	287	223	7.3	5.0	$C_{gdee} + C_{gda}$ (fF)	0.8	0.71	2.2	2.2
R_2 (Ω)	218	160	0.03	6.8					
C_1 (aF)	48	39	0.025	47	C_{ga}**	6.3 fF			
C_2 (aF)	75	100	93	75	C_{da}**	2.4 fF			
R_{dbi}	2.6 kΩ	1.1 kΩ	801 Ω	997 Ω	C_{gda}**	3.9 fF			
R_{sbi} (Ω)	578	258	691	604					
C_{dbi}	20.5 fF	187 fF	5.1 nF	2.1 nF					
C_{sbi}	9.9 nF	7.6 nF	50.7 fF	18.5 fF					

* These values are derived from measurements of the S-parameters at the following bias conditions: $V_{ds} = 2$ V, V_{gs} at g_{mmax} ($V_{gs} = 0.83$ V for $W = 5$ μm, $V_{gs} = 0.75$ V for $W = 10$ μm, $V_{gs} = 0.75$ V for $W = 20$ μm, $V_{gs} = 0.67$ V for $W = 40$ μm).

** Access capacitances extracted from data using linear regression with respect to gate width W following Eq. (4).

This parasitic resistance may have an adverse effect on the transconductance, and consequently adversely affect both f_T and f_{max}. The gate current measured in the 30 nm × 40 μm × 2 MOSFET is $I_G = 4.9$ nA/μm when it is on, biased at $V_G = 2$ V and $V_D = 2$ V, and < 28 nA/μm when it is off, biased at $V_G = 2$ V and $V_D = 0$ V. Because of the small active area, the gate leakage current associated with quantum mechanical tunneling through the 1.3 nm gate oxide does not affect the dc characteristics appreciably although it may contribute shot noise [24].

4. Small-signal SHF model

According to Tsividis [19], a physics-based model for the high frequency operation of a MOSFET can be separated into intrinsic and extrinsic contributions as illustrated in Fig. 3. The intrinsic elements, which are inside the red[1] broken line in Fig. 3, are supposed to depend on the bias conditions and the geometry of the active area of the device. The intrinsic parameter, g_m, represents the small-signal gate transconductance; g_{mb} represents the substrate transconductance; and g_{ds} the output conductance. For nanometer-scale transistors the small-signal model also has to account for tunneling between the gate electrode and the source and drain contacts, which is exponentially dependent on the voltage and oxide thickness. Two resistances, R_{gst} and R_{gdt} are used to represent the different currents associated with tunneling through the same 1.3 nm thick gate oxide. To account for charge storage, we assume quasi-static operation and use the intrinsic capacitances C_{gsi}, C_{sbi}, C_{gdi} and C_{dbi} required to represent the change in the gate and depletion charges. The capacitances are bias dependent and may not be reciprocal. Non-reciprocal effects, which are modeled by adding transcapacitances and corresponding dependent current-sources to accurately account for the charging current, are expected to affect the prediction of the transadmittances Y_{21} and Y_{12}, although they are usually ignored for frequencies <50 GHz [25].

The small-signal model for the drain current consists, not only of the contribution from the channel (drain-to-source), but also has a parasitic component due to impact ionization at high electric field that flows between the substrate and the drain. The role of the source in the above analysis will be played by the substrate instead. The parameters R_{dbi} and C_{dbi}, and R_{sbi} and C_{sbi} reflect the substrate contribution. And finally, the nonquasi-static effects (NQS) in the intrinsic model (i.e. the finite charging time in the channel inversion layer due to the history of the voltage biases) are accounted for by using the voltage-control-current-sources (VCCS) connected in parallel with the intrinsic capacitances and admittances [19]. For example, the lag in the gate current behind a quickly varying drain signal is accounted for by parameters such as R_{gdi} and C_{gdi}. The corresponding parameters associated with a source excitation are R_{gsi} and C_{gsi}. The coefficients of the VCCS's (g_m and g_{mb}) in Fig. 3 would usually be complex, but following Tsividis we circumvented this complication by using the series combinations of R_1 and C_1, and R_2 and C_2, respectively (which are highlighted in red in Fig. 3) [19]. These simple circuit elements are chosen to draw a negligible current in comparison to the R_{gsi} and C_{gsi}, and R_{bsi} and C_{bsi} combinations.

In contrast with the intrinsic components, the extrinsic elements are supposed to be independent of bias, but scale

[1] For interpretation of the reference in colour in text the reader is referred to the web version of this article.

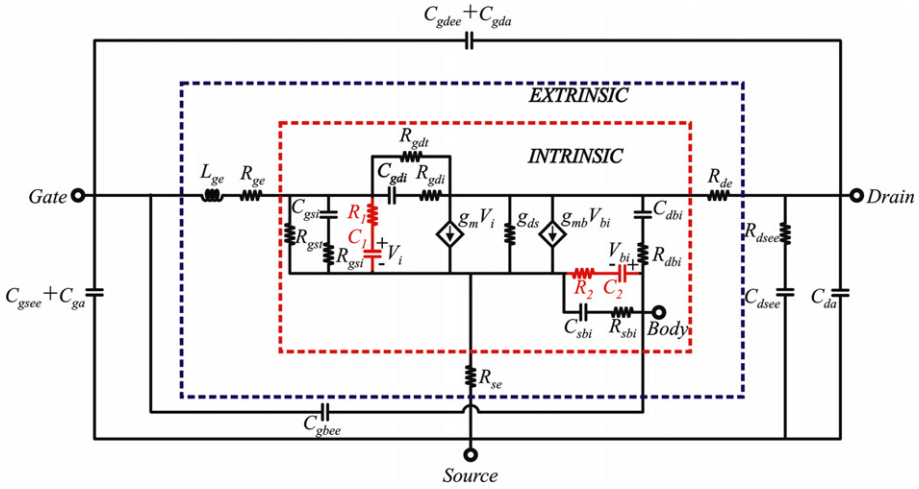

Fig. 3. The compact equivalent circuit of a sub-50 nm MOSFET. This model accounts for the NQS effect in the SHF region. The equivalent circuit of the intrinsic part of the transistor which was used to fit the measurements includes the pairs C_1–R_1 and C_2–R_2 to represent the delay associated with the long gate fingers. In this case g_m and g_{mb} are not complex.

with the active area of the device. At high frequency, the admittance of these extrinsic capacitances can be large compared to that of the intrinsic elements in parallel with them. Thus, a comprehensive small-signal model of the extrinsic effects must also include parasitic resistances and inductances. This is accomplished economically using only a few lumped elements. For example, as shown in Fig. 3, the resistive material associated with the source, drain, and gate are represented respectively by the resistances, R_{se}, R_{de}, and R_{ge}. R_{se} and R_{de}, which include the losses associated with the metal contacts and the contact resistance associated with the source and drain implants, are inversely proportional to the gate width. The extrinsic gate resistance, R_{ge}, includes both the resistance of the gate finger, which is proportional to the gate width, and the metallic losses associated with the connection of the gate to the co-planar waveguide (CPW) up to the reference plane. The contribution to R_{ge} due to the series resistance of the silicide on top of the polysilicon layer is usually assumed to predominate according to $R_{ge} = r_s W/3L_g$, where r_s is the sheet resistivity of the silicide [11]. (For aggressively scaled technologies there is an additional contact resistance that may have to be taken into account that is associated with the interface due to the Schottky contact between the silicide and polysilicon given by: $R_{gec} = r_C/WL_g$ where r_C is the contact resistivity [31]. For a typical transistor used in these experiments $W \geqslant 5\ \mu m$ and $R_{ge} > R_{gec} \sim 100\ \Omega$.) The parasitic inductances are defined similarly. However, because of the nanometer-scale channel length and wide source-drain regions, the intrinsic and extrinsic inductances have values <1 pH, and so they are negligible in the SHF band. Due to the size of the gate finger and gate resistance, we found it necessary for a good fit to include only the extrinsic gate inductance L_{ge}.

We experimentally determined that there are additional parasitic couplings to account for between the metal wires outside the intrinsic region of the device, which can be adequately represented by capacitances that span the drain-source, gate-source and gate-drain terminals [26–28]. In Fig. 3, we denote these capacitances as C_{dsee}, C_{gsee}, and C_{gdee}, i.e. extrinsic–extrinsic elements which are independent of bias and proportional to the gate width. These elements do not account for the access admittances like C_{da} associated with the metal leads as illustrated in Fig. 4,

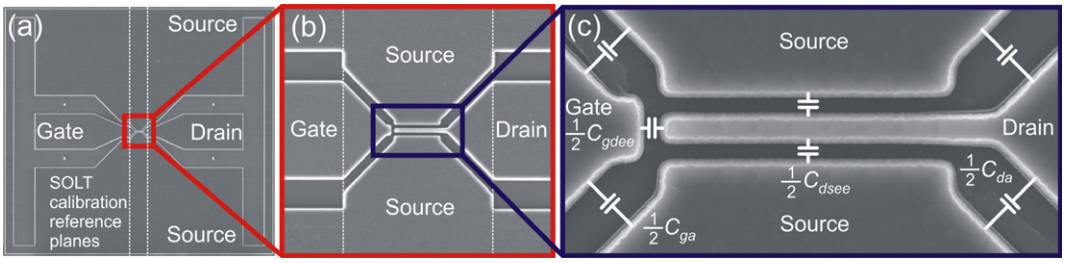

Fig. 4. (a) Is a top–down view (SEM) of the coplanar waveguides used in conjunction with GND-signal-GND (GSG) RF probes to measure the Y-parameters of our device. The signal of port1 of the PNA is gate, and the signal of port2 of the PNA is drain. The SOLT calibration brings the reference planes for the RF measurement to the edges of the coplanar waveguides, marked by the dotted lines. (b) A magnified image of the device area to show the location of the measurement reference planes. (c) Shows the metallization above the active device area and the parasitic capacitances associated with it. The access capacitances C_{ga} and C_{da}, the gate-to-drain and drain-to-source extrinsic–extrinsic capacitance C_{gdee} and C_{dsee} are not accounted for with the SOLT calibration. C_{ga}, C_{da}, and C_{gdee} are relatively constant as a function of W. C_{dsee}, on the other hand, scales with W.

however. Although the access admittances can be adequately represented by capacitances in parallel with these extrinsic–extrinsic capacitances, they scale differently. So, Fig. 3 indicates that the access capacitances C_{ga} and C_{gda} are summed with the corresponding extrinsic–extrinsic capacitances.

5. Measurements of the small-signal SHF performance

The RF performance from 50 MHz to 50 GHz of nanotransistors embedded in a CPW was measured using an Agilent E8364B PNA. To launch the signal from the RF probe tips to the device under test (DUT), a CPW is used in a ground-signal-ground configuration as shown in the micrograph of Fig. 4. Because of the relative size of the device, parasitic elements associated with the contact pads and interconnects cannot be neglected. So, the intrinsic performance of the transistor had to be de-embedded from the data.

We used Y-parameters to characterize the RF measurements. To obtain an accurate assessment of the Y-parameters, we used a two step procedure. First, using a full 2-port SOLT calibration [26,29], we automatically subtracted the Y-parameters associated with on-wafer standards consisting of a short, an open, a load (nominally 50 Ω, and a thru (SOLT) from the DUT measurements using the E8364B. (This extends the reference plane to the dotted line in Fig. 4b.) While this calibration technique is renowned for its accuracy [29], it does not account for the access admittances like C_{da} indicated in Fig. 4c associated with the metal leads extending from the SOLT reference plane and to the source, drain and gate contacts. Since the MOSFET gate width does not extend into the ground plane of the CPW, (see Fig. 4b), the access admittances can be adequately represented solely by capacitors.

The second step involves subtracting the admittance associated with the access capacitances from the imaginary part of the SOLT calibrated measurements of the Y-parameters. The access capacitances were inferred using a linear regression to determine the intercepts of the effective capacitance (Im[Y/ω]) versus gate width from measurements of the Y-parameters using extraction techniques based on the ColdFET and Inversion approximations [25,28], with the MOSFET biased OFF, (i.e. $V_{ds} = 0$ V and $V_g < V_t$), the active area of the device can be neglected, and the imaginary part of the measured Y-parameters is the sum of the extrinsic capacitances and the access capacitances, i.e.

$$\mathrm{Im}[Y_{11}] \cong \omega(C_{ga} + C_{gda} + C_{gdee} + C_{gsee}) \quad (4a)$$

$$\mathrm{Im}[Y_{12}] \cong -\omega(C_{gda} + C_{gdee}) \quad (4b)$$

$$\mathrm{Im}[Y_{22}] \cong \omega(C_{da} + C_{gda} + C_{gdee} + C_{dsee}) \quad (4c)$$

Eqs. 4(a–c) are insufficient for discriminating between the access capacitances and the extrinsic (or intrinsic) capacitances. However, the extrinsic and extrinsic–extrinsic capacitances like C_{dsee} are directly proportional to the gate width, while the access capacitances are independent of W. Thus, by using the intercepts of a regression line through the effective capacitance Im(Y/ω) versus gate width, we can determine the access capacitances. Two deficiencies of this method are: (1) the capacitances obtained include other capacitances such as the gate-to-source capacitance due to the enlarged region of the polysilicon used for contact access that is located on top of the LOCOS isolation oxide (usually negligible); and (2) there is additional error introduced by the regression equation.

Using Agilent's Advanced Design System software to fit the model to the data, we arrived at the final estimates for the lumped elements represented in Fig. 3. For example, we extracted the intrinsic and extrinsic parameters associated with $30\ \mathrm{nm} \times W\mu\mathrm{m} \times 2$ and $45\ \mathrm{nm} \times W\mu\mathrm{m} \times 2$ MOSFETs biased at $V_{ds} = 2.0$ V and V_g at the maximum of the transconductance. Table 1 lists the values for all parameters in the compact model of the $30\ \mathrm{nm} \times W\mu\mathrm{m} \times 2$ nMOSFET. These parameters are consistent with analytical estimates given by Tsvidis [19]. For example, from [19] we expect $C_{gsi} \sim (2/3)C_{ox}$ and $C_{gdi} \sim 0$ and $R_{gsi}C_{gsi} > R_{gdi}C_{gdi}$, despite the observation that the fit is insensitive to the value of R_{gdi}. We find that without varying R_{gsi}, there is <1–2% change in the parameters and <50% change in the residual for $1\ \Omega < R_{gdi} < 2R_{gsi}$.

Comparisons between the predictions of the model for the 30 nm × 40 μm × 2 and 45 nm × 10 μm × 2 MOSFETs using the parameters obtained from the extraction technique and the measurements are illustrated in Figs. 5 and 6, respectively. Generally, we find very close correspondence between the Y-parameters calculated from the model and the measured data up to 50 GHz. Measurements of both the real and imaginary parts of the Y-parameters fol-

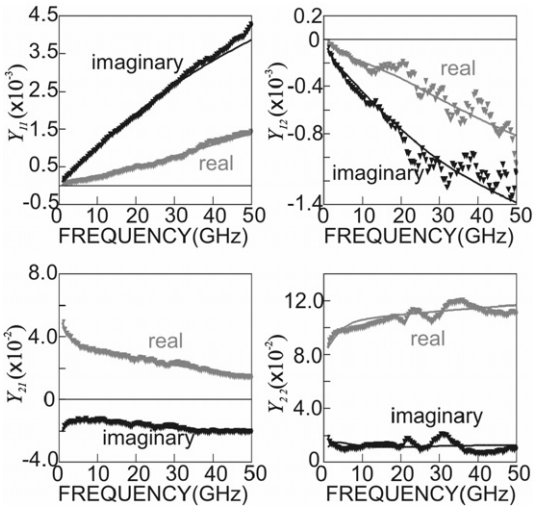

Fig. 5. Comparison of the Y-parameters of the model and measured data for a 30 nm × 40 μm × 2 device from 1 GHz to 50 GHz. Both the real (gray lines) and imaginary (black lines) parts of the model are compared with the real (grey triangles) and imaginary (black triangles) parts of the measured Y-parameters respectively. We find excellent agreement between the model and measurements in the 3–30 GHz band.

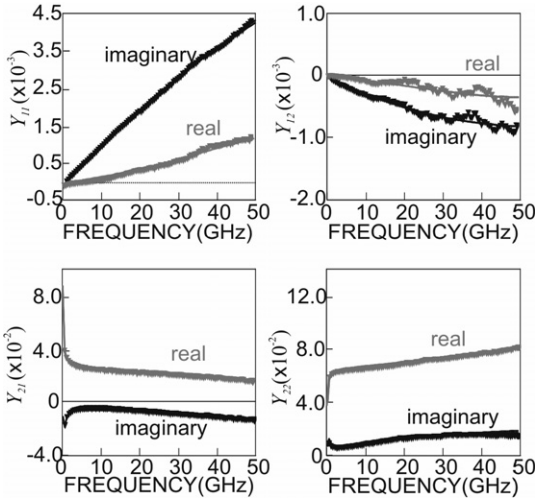

Fig. 6. Comparison of the Y-parameters of the model and measured data for a 45 nm × 10 μm × 2 device from 1 GHz to 50 GHz. Both the real (gray lines) and imaginary (black lines) parts of the model are compared with the real (grey triangles) and imaginary (black triangles) parts of the measured Y-parameters, respectively. We find excellent agreement between the model and measurements in the 3–30 GHz band.

low the predictions of the model shown in Fig. 3, even though nonreciprocal transcapacitances (such as C_{gdi} and C_{dgi}) are not included in the model, which was expected to make it difficult to accurately predict Im[Y_{21}] and Im[Y_{12}] at the same time [25]. Unlike the measurements obtained on 5, 10, and 20 μm channel width MOSFETs, neither Im[Y_{21}] nor Im[Y_{22}] approach zero in the frequency range shown in Figs. 5 and 6, which the model of Fig. 3 attributes to the high gate resistance associated with the 40 μm channel widths.

One advantage of this model of Fig. 3 is economy. It contains fewer parameters than the distributed version [30], and yet it still accurately accounts for the measurements in the frequency range from 1 GHz to 50 GHz, and scales appropriately with geometry. While this model is compact, which facilitates IC design, the fit to the data deteriorates slightly compared to the distributed model [30]. The error function χ^2, which is based on the sum of the squares of the residuals, is doubled from approximately 0.03 to 0.05. This slight degradation of the fit, however, is imperceptible in the plots of the Y- parameters, the power gain and the current gain.

The extracted parameters match the results obtained from low frequency or dc measurements and closely correspond to parameters estimated from the physical structure of the MOSFET. Fig. 7a and b compare low frequency (dc) measurements of the effective g_m, and g_{ds}, with the intrinsic parameters extracted from the model of Fig. 3 in the low frequency limit (5 MHz). We find that the low frequency measurements generally correspond with the extracted parameters, except for a gate width of $W = 40$ μm where the model predicts larger values, beyond the error attributed to the fit. The discrepancies found for a gate width of $W = 40$ μm may be associated with the inability of the lumped element model to capture the distributed nature of the gate, source, drain and channel resistances. It is not observed for a distributed model for the gate [30]. A comparison of measurements of the intrinsic capacitance C_{gsi} and the extrinsic parameter C_{dsee} versus gate width are shown in Fig. 7c and d. To compare with the values extracted from the model (Fig. 3), we have estimated C_{gsi} from PADRE simulations of the transistor geometry. C_{gsee} was inferred from the measured Y-parameters using extraction techniques based on the ColdFET approximation [25].

6. Figures-of-merit for RF-MOSFET amplifier

Fig. 8a illustrates the RF performance of 30 nm × 40 μm × 2 nMOSFETs biased at the transconductance maximum $g_m = 0.94$ S/mm found at $V_g = 0.67$ V for $V_{ds} = 2$ V. The performance is indicated by the short-circuit current gain, h_{21}, and Mason's unilateral gain, U. Both h_{21}, and U exhibit a nearly ideal single-pole response with frequency, and so we infer f_T and f_{max} by extrapolating to unity gain at −20 dB/decade from data in the range 40–50 GHz. For the 30 nm gate length MOSFET we find f_T =465 GHz as indicated in Fig. 8a. However, a similar extrapolation using a single-pole response for data in the range 1–30 GHz indicates that $f_T \cong 500$ GHz for the same MOSFET as shown in Fig. 8b. This discrepancy from the single-pole response observed above 30 GHz is an indication of the difficulty realizing adequate calibration at the high frequency end of the PNA.

An nMOSFET $f_T = 465$ GHz is comparable to the best SiGe-BJT reported in literature so far [6]. On the other hand, the power gain performance measured by $f_{max} = 45$ GHz as shown in Fig. 8c is reduced by a factor of 10 from the cut-off frequency. The degradation of f_{max} is an indication of the detrimental effect of parasitics indicated by Eq. (2). Relying on the agreement between the model of Fig. 3 and measured current and power gain shown in Fig. 8c, we infer that parasitic elements such as the gate resistance, R_{ge}, and the series source and drain resistances, R_{se} and R_{de}, adversely affect the power gain in the SHF band. Moreover, previous extrapolations based on the same measurements showed that $f_T = 290$ GHz and $f_{max} = 45$ GHz for the 30 nm gate length [30]. The prior extrapolations reflect the detrimental effects of the access capacitances on the estimate of f_T.

Figs. 9a and b illustrate measurements of the RF performance in the same 30 nm × 40 μm × 2 and 45 nm × 20 μm × 2 nMOSFETs of Fig. 2, showing the same trend in current and power gain. There is an improvement gleaned in the current gain from decreasing the channel length: i.e. the 45 nm device shows $f_T = 350$ GHz while the 30 nm MOSFET has an f_T above 460 GHz. Fig. 9c summarizes the measured dependence of f_T on the gate length inferred from top–down scanning electron microscopy (SEM) for devices with the same gate width. Within the error, we find that f_T data is consistent with a $1/L_g$ dependence associated with short channel MOSFETs over

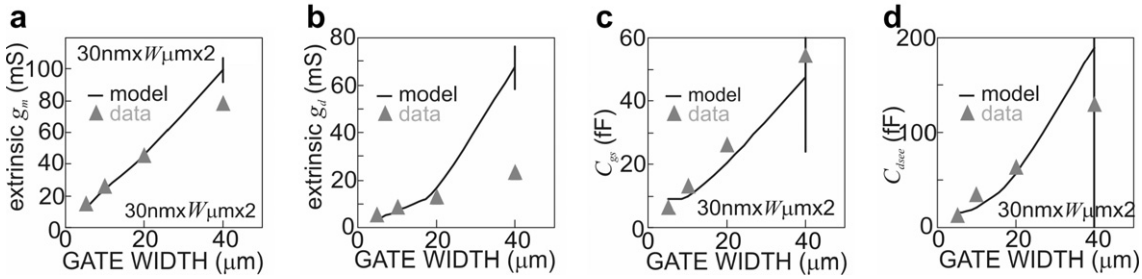

Fig. 7. A comparison of parameters obtained from the model to parameters obtained from extraction based on the Y-parameter matrix and DC measurements. The triangles represent values obtained from measurements or estimates, while the solid line mark values obtained from the model. For the intrinsic model, for example, the transconductance (left) tracks the transconductance values obtained from DC measurements. For the extrinsic parameters, the drain-to-source capacitance C_{dsee} also tracks the values extracted from measurements.

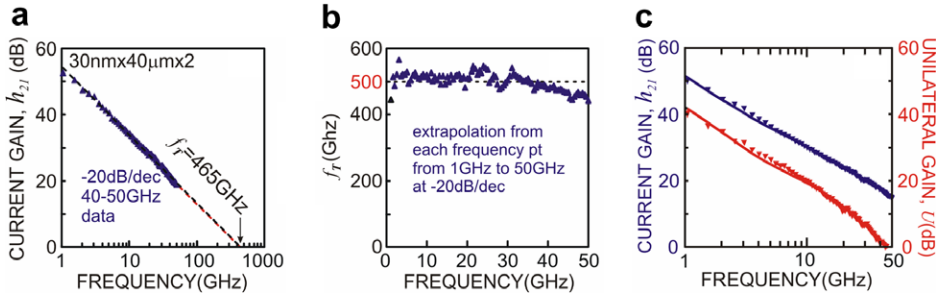

Fig. 8. (a) Cut-off frequency, f_T, of 30 nm × 40 µm × 2 inferred from a −20 dB/dec extrapolation from the points in the 40–50 GHz range and (b) the distribution of the cut-off frequency inferred at −20 dB/dec from each point in the 1–50 GHz range. The deviation above 30 GHz hints at the difficulties associated with realizing good calibration at very high frequencies. The data below 30 GHz extrapolates to a cut-off frequency of 500 GHz. (c) A comparison between the measured (triangles) and model (solid line) short-circuit current gain $|h_{21}|^2$ and Mason's Power Gain U for the same transistor. The model accurately represents the performance over the measured frequency range.

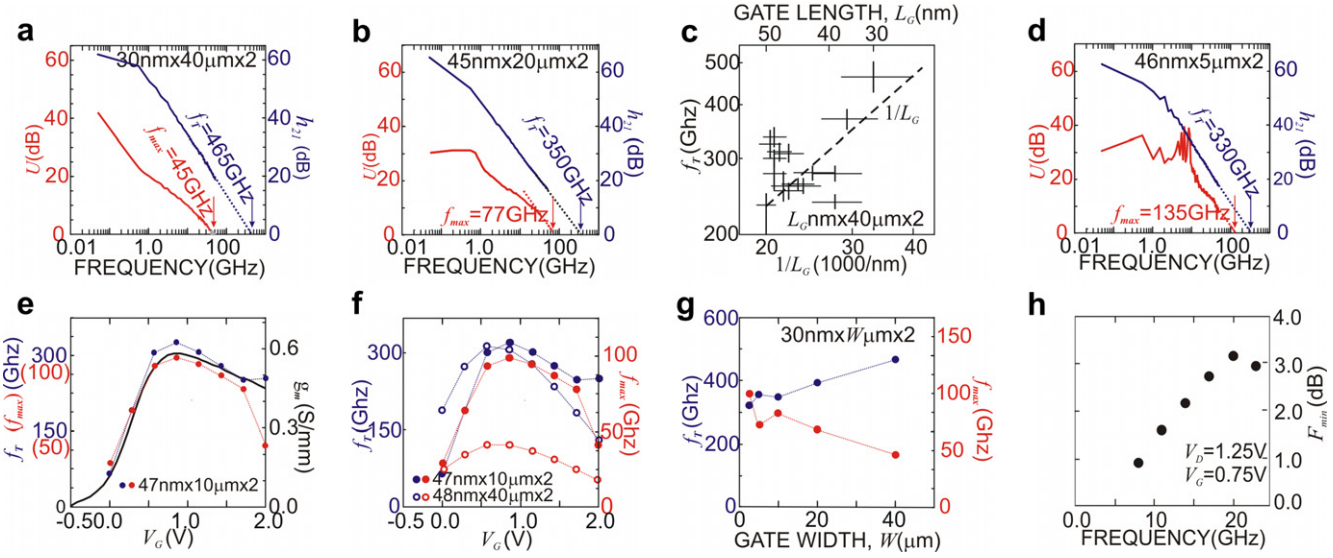

Fig. 9. (a) and (b) Measurements of the RF performance of the 30 nm × 40 µm × 2 and 45 nm × 20 µm × 2 nMOSFETs shown in Fig. 2 f_T and f_{max} were obtained at $V_{ds} = 2$ V by extrapolating measurements in the 40–50 GHz range using a slope of −20 dB/decade. Despite the degradation in f_T relative to the device in (a), f_{max} improves substantially with an increase in the channel length and a decrease in channel width. In (c) f_T is plotted as a function of $1/L_g$ for L_g(nm) × 40 µm × 2 nMOSFETs. A linear fit of the log[f_T] vs. log[$1/L_g$] data reveals a $1/L_g{}^{0.5}$ dependence of f_T. (d) Measurements of the RF performance of the 46 nm × 5 µm × 2 nMOSFET for comparison with (b). The dramatic change in f_{max} is mainly associated with a change in gate width. (e) f_T and f_{max} measured on a 47 nm × 10 µm × 2 nMOSFET at $V_{ds} = 2$ V are compared to the transconductance g_m as a function of gate voltage, V_g. Both figures-of-merit are observed to track with g_m. (f) f_T and f_{max} measured in a 47 nm × 10 µm × 2 and 48 nm × 40 µm × 2 nMOSFET as a function of gate voltage, V_g. Notice that f_{max} decreases substantially with increase in gate width. (g) The dependence of f_T and f_{max} on the gate width, W measured at the peak in g_m for $V_{ds} = 2$ V. f_{max} changes by about 300% for MOSFETs with nominally the same $L_g = 30$ nm, while f_T changes by less than a factor of 2 for the same transistors. (h) f_{min} as a function of frequency for a 45 nm × 40 µm × 2 nMOSFET measured at $V_g = 0.75$ V and $V_{ds} = 1.25$ V.

the narrow range examined (30–50 nm) as anticipated from Eq. (1). The SEM measurements were found to be consistent with cross-sectional TEM assessments done on the same devices using focused ion beam milling to prepare the specimens. The error associated with the length measurement develops from variations in the gate length along the width of the device associated with the lithography and etching. The error associated with the estimate of f_T develops mainly from errors in the measurement of the gain in the 40–50 GHz band.

Both g_{ds} and R_{ge} increase as the gate length decreases, which has a detrimental effect on f_{max}. So, we anticipate that $f_{max} \sim 1/W$ as indicated in Eq. (2), provided that g_m, R_{ge}, g_{ds}, C_{gsi} and C_{gdi} scale linearly with the gate width [11]. Figs. 9b and d show two *n*MOSFETs of comparable gate length, but different gate widths: i.e. 45 nm × 20 μm × 2 and 46 nm × 5 μm × 2, respectively. As illustrated by a comparison of the two figures, devices with a shorter gate width deliver higher f_{max} without compromising f_T. The 46 nm × 5 μm × 2 device of Fig. 9d exhibits an $f_T = 330$ GHz and $f_{max} = 135$ GHz, while the 45 nm × 20 μm × 2 device has $f_T = 350$ GHz and $f_{max} = 77$ GHz. Devices with a larger gate length can potentially deliver higher f_{max} despite a reduction in f_T since improvements in R_{ge} and g_{ds} could compensate for the degradation. The slight degradation of f_T measured in the device of Fig. 9d is offset by a decrease in W, R_g, and R_{ds}.

We methodically explored the dependence of f_T and f_{max} on gate voltage, width and length. The measured gate voltage-bias dependence of f_T and f_{max} is illustrated in Fig. 9e for a 47 nm × 10 μm × 2 MOSFET. The observed dependence follows from the gate voltage dependence of the transconductance as indicated in Eqs. (1) and (2). Generally, for the devices we measured, both f_T and f_{max} scale with g_m in the range $0 < V_g < 1.7$ V. The measured dependence of f_T and f_{max} on gate width is illustrated in Fig. 9f for similar gate length *n*MOSFETs (47 nm versus 48 nm) with different gate widths (10 μm versus 40 μm, respectively). Notice that even though the peak values of $f_T \approx 300$ GHz are nearly identical, the peak value of f_{max} for the 47 nm × 10 μm × 2 MOSFET is about 100 GHz, while the peak value for the 48 nm × 40 μm × 2 is less than half that value, $f_{max} = 41$ GHz. Fig. 9g summarizes the trends observed in f_T and f_{max} as a function of the gate width for nominally 30 nm gate length MOSFETs. f_T increases as the gate width increases while f_{max} increases with decreasing W. The increase in f_T with gate width W is due to both proximity corrections in the electron beam lithography and gate etch that generate a narrower gate length for larger gate widths.

Finally, to make effective use of CMOS technology for RF applications, the noise figure, F_{min}, also has to be minimized with an acceptable associated gain. So, we examined the broadband noise performance in the SHF band. Fig. 9h shows a linear frequency dependence of F_{min} that we anticipated in Eq. (3). F_{min} is about 0.9 dB at 8 GHz. Generally, we find a larger noise figure with increasing frequency – approaching 3 dB for frequencies above 20 GHz, which compares favorably to data reported on 80 nm *n*MOSFET at the same operating point [2]. According to Eq. (3), we expect an improvement in the noise figure if the gate resistance improves. For this 47 nm × 10 μm × 2 *n*MOSFET at 8 GHz we find a maximum available gain of 22 dB with $F_{min} < 1$ dB, while at 20 GHz, we find a maximum available gain of 11 dB with $F_{min} < 3.5$ dB.

7. Conclusion

In summary, measurements of sub-50 nm gate *n*MOSFETs reveal the highest RF performance of an *n*MOSFET so far. For a 30 nm × 40 μm × 2 device, we found $f_T >$ 460 GHz at $V_{ds} = 2$ V, $V_g = 0.67$ V after accounting for access capacitances. However, measurements of $f_{max} \leqslant$ 135 GHz and the noise figure indicate that parasitics adversely affect the circuit performance. Very accurate models of the measurements in the SHF band that account for non-quasi-static effects and incorporate parasitic elements indicate that the gate resistance and the source-drain limit the power gain.

Acknowledgments

We gratefully acknowledge support from the Air Force (FA9550-04-1-0214), NSF (CCR 01-21616) and ONR (N00014-03-0268). The Center for Microanalysis of Materials, supported by the DOE (DEFG02-91-ER45439), was instrumental in this work.

References

[1] Taylor GW, Simmons JG. Figure of merit for integrated bipolar transistors. Sol State Electron 1986;29:941–6.
[2] Kuhn K et al. A comparison of state-of-the-art NMOS and HBT devices for analog/mixed-circuit/RF circuit application. VLSI Tech Symp Tech Dig 2004;224.
[3] Momose HS, Ohguro T, Morifuji E, Sugaya H, Nakamura S, Iwai H. Cutoff frequency and propagation delay time of 1.5-nm gate oxide CMOS. IEEE Trans Electron Dev 2001;48:1165–74.
[4] Zamdmer N et al. A 243-GHz f_t and 208-GHz f_{max}, 90-nm SOI CMOS SoC technology with low-power millimeter-wave digital and RF circuit capability. VLSI Tech Symp 2004:98–9.
[5] Lee S, Wagner L, Janannathan B, Csutak S, Pekarik J, Zamdmer N, et al. Record RF performance of sub-46 nm Lgate NFETs in microprocessor SOI CMOS technologies. IEDM Tech Dig 2005:241–4.
[6] Suemitsu T, Enoki T, Yokoyama H, Ishii Y. Impact of two-step-recessed gate structure on RF performance of InP-based HEMTs. Electron Lett 1998;34:220–2.
[7] Khater M et al. SiGe HBT technology with $f_{max}/f_T = 350/300$ GHz and gate delay below 3.3 ps. IEDM Tech Dig 2004:247–50.
[8] Hafez W, Feng M. 0.25 μm emitter InP SHBTs with $f_T = 550$ GHz and $BV_{CEO} > 2$ V. IEDM Tech Dig 2004:549–52.
[9] Hussian T et al. First demonstration of sub-0.25 μm-width emitter InP-DHBTs with >400 GHz f_t and >400 GHz f_{max}. IEDM Tech Dig 2004:553–6.
[10] Yu D et al. Ultra high-speed 0.25-μm emitter InP-InGaAs SHBTs with f_{max} of 687 GHz. IEDM Tech Dig 2004:557–60.
[11] Razavi B, Yan R, Lee KF. Impact of distributed gate resistance on the performance of MOS devices. IEEE Trans Circuit Syst-I 1994;41:750–4.

[12] Tiemeijer LF et al. Record RF performance of 90 nm CMOS technology. IEDM Tech Dig 2004:441–4.
[13] Dambrine G et al. What are the limiting parameters of deep-submicron MOSFETS for high frequency applications? IEEE Elec Dev Lett 2003;24:189–91.
[14] Liu W, Gharpurey R, Chang MC, Erdogan U, Aggarwal R, Mattia JP. RF MOSFET modeling ac-counting for distributed substrate and channel resistances with emphasis on the BSIM3v3 SPICE model. IEDM Tech Dig 1997:309–12.
[15] Pehlke DR, Schroter M, Burstein A, Matloubian M, Chang MF. High-frequency application of MOS compact models and their development for scalable RF MOS libraries. Proc IEEE Custom IC Conf 1998:219–22.
[16] Ou J-J, Jin X, Ma I, Hu C, Gray P. CMOS RF modeling for GHz communication IC's. VLSI Technol Symp 1998:94–5.
[17] Jen SH, Enz C, Pehlke DR, Schroter M, Sheu BJ, Accurate MOS transistor modeling and parameter extraction valid up to 10-GHz. In: Proceedings of European solid-state device research conference Bordeaux, France; 1998. p. 484–7.
[18] Cheng Y et al. RF modeling issues of deep-submicron MOS-FET's for circuit design. Proc IEEE Int Conf Solid-State IC Tech 1998:416–9.
[19] Tsividis YP. Operation and modeling of the MOS transistor. McGraw-Hill; 1987.
[20] Enz C. An MOS transistor model for RF IC design valid in all regions of operation. IEEE Trans Microw Theory Techn 2002;50:342–58.
[21] Cho D, Kang S, Kim K, Lee S. An accurate intrinsic capacitance modeling for deep submicrometer MOSFETs. IEEE Trans Electron Dev 1995;42:540–8.
[22] Cheng Y, Chen C, Matloubian M, Deen MJ. High-frequency small signal ac and noise modeling of MOSFETs for IC design. IEEE Trans Electron Dev 2002;49:400–8.
[23] Enz CC, Cheng Y. MOS transistor modeling for RF IC design. IEEE J Solid-State Circ 2001;35:186–201.
[24] Eminente S, Alessandrini M, Fiegna C. Comparative analysis of the RF and noise performance of bulk and single-gate ultra-thin SOI MOSFETs by numerical simulation. Solid-State Electron 2004;48(4):543–9.
[25] Kwon I, Je M, Lee K, Shin H. A simple and analytical parameter-extraction method of a microwave MOSFET. IEEE Trans Microw Theory Tech 2002;50:1503–9.
[26] Raskin JP, Gilon R, Dambrine G, Chen J, Vanhoenacker D, Colinge JP. Accurate characterization of SOI MOSFETs for the design of low-voltage, low-power RF ICs. Analog IC Signal Proc 2000;25:133–55.
[27] Goffioul M, Vanhoenacker D, Raskin J-P. Direct extraction techniques of microwave small-signal model and technology parameters for subquarter micron SOI MOSFETs. 5th Symp Diagn Yield 2000:8.
[28] Dehan M, Raskin J-P, Vanhoenacker D. Comparison of different extraction methods of small-signal parameters for SOI MOSFETs. Proc 32th Eur Microw Conf 2002.
[29] Heuermann H, Schiek B. Robust algorithms for TAN network analyzer self-calibration procedures. IEEE Trans Instrum Meas 1994;43(1):18–22.
[30] Dimitrov V et al. High performance, sub-50 nm MOSFETS for mixed signal applications. IEDM Tech Dig 2005:213–6.
[31] Litwin A. Overlooked interfacial silicide–polysilicon gate resistance in MOS transistors. IEEE Trans Electron Dev 2001;48(9):2179–81.