Compact Model for Carbon Nanotube Field-Effect Transistors Including Nonidealities and Calibrated With Experimental Data Down to 9-nm Gate Length

Jieying Luo, Student Member, IEEE, Lan Wei, Member, IEEE, Chi-Shuen Lee, Aaron D. Franklin, Member, IEEE, Ximeng Guan, Member, IEEE, Eric Pop, Senior Member, IEEE, Dimitri A. Antoniadis, Fellow, IEEE, and H.-S. Philip Wong, Fellow, IEEE

Abstract—A semianalytical carbon nanotube field-effect transistor (CNFET) model based on the virtual-source model is presented, which includes series resistance, parasitic capacitance, and direct source-to-drain tunneling leakage. The model is calibrated with recent experimental data down to 9-nm gate length. Device performance of 22- to 7-nm technology nodes is analyzed. The results suggest that contact resistance is the key performance limiter for CNFETs; direct source-to-drain tunneling results in significant leakage due to low effective mass in carbon nanotubes and prevents further downscaling of the gate length. The design space that minimizes the gate delay in CNFETs subject to off-state leakage current (I_{OFF}) constraints is explored. Through the optimization of the length of the gate, contact, and extension regions to balance the parasitic effects, the gate delay can be improved by more than 10% at 11- and 7-nm technology nodes compared with the conventional 0.7× scaling rule, while the off-state leakage current remains below 0.5 μA/μm.

Index Terms—Carbon nanotube (CNT), carbon nanotube field effect transistor (CNFET), contact resistance, direct source-to-drain tunneling.

Manuscript received January 21, 2013; revised April 2, 2013; accepted April 2, 2013. Date of publication May 6, 2013; date of current version May 16, 2013. This work was supported in part by the MSD and the C2S2 Focus Centers, two of six research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation subsidiary, and the member companies of the Stanford Initiative for Nanoscale Materials and Processing. It is additionally supported by IBM through its membership at the Center for Integrated Systems at Stanford. J. Luo and L. Wei are additionally supported by the Stanford Graduate Fellowship. E. Pop has been supported by the NSF CAREER under award ECCS 0954423. The review of this paper was arranged by Editor A. C. Seabaugh.

J. Luo, C.-S. Lee, and H.-S. P. Wong are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: ivy.luo@stanford.edu; chishuang@stanford.edu; hs.p.wong@stanford.edu).

L. Wei was with the Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. She is now with the Altera Corporation, San Jose, CA 95134 USA (e-mail: lwei@altera.com).

A. D. Franklin is with the IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: aaronf@us.ibm.com).

X. Guan was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with the IBM Semiconductor Research and Development Center, Hopewell Junction, NY 12533 USA (e-mail: xgguan@us.ibm.com).

E. Pop is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: epop@illinois.edu).

D. A. Antoniadis is with the Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: daa@mtl.mit.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2013.2258023
it is difficult to obtain large amounts of reliable data, yet there is a strong need for a realistic model to assess their potential for future use. In our proposed model, we relate the empirical parameters, as much as possible, with the device structures such as gate length and contact length to enable projections that reflect changes in the device design, while the other parameters are extracted from a few sets of experimental data. This model captures CNFET’s physical properties such as diameter-dependent tunneling and CNT–metal contact resistance, which are important for the purpose of performance benchmarking, projection, and circuit design and optimization.

The VS model is one of the components of this hierarchical model. While each of the constituent components have been published by authors before, the constituent components do not lend themselves to gaining physical insights into how to optimize the CNFET and how to use them for design optimization and circuit simulation. The way the constituent models are put together is the main contribution of this paper.

This paper is organized as follows. In Section II, each level of the model is described. Calibration with experimental data from 300-nm down to 9-nm gate length is presented in Section III. Based on the parameters extracted from Section III, the performances of 22- to 7-nm technology nodes are projected and the challenges related to parasitic effects and direct tunneling are highlighted in Section IV. Finally, Section V illustrates the use of the model for device optimization of CNFETs at the 11- and 7-nm technology nodes.

II. CNFET MODEL

Fig. 1 illustrates the modeled device structure. It is a planar top-gated CNFET with an undoped CNT channel and ungated, highly-doped source/drain (S/D) extensions. A single CNFET contains multiple CNTs in parallel within the channel to boost the drive current. The ungated S/D extensions can offset the gate electrode from the S/D contact plugs and reduce the parasitic capacitance [27]. Present experimental CNFETs often employ a back gate with gate-to-S/D overlap [28], [29] or undoped (not intentionally doped), underlapped, and ungated extension regions [30], because a stable and well-controlled CNT doping technology is not yet available. However, it is well known that the structure in Fig. 1 can achieve better device performance with CMOS-compatible fabrication [31], [32] due to its effectiveness in reducing parasitic capacitances.

The model takes the CNFET structure design and physical parameters as inputs. The design parameters include the device pitch \( L_{\text{pitch}} \), length of the gate \( L_g \) and the contact \( L_c \), the gate width \( W \), the dielectric constant \( \epsilon \) and thickness \( t_\text{ox} \) of the gate oxide, the height of the gate \( H_g \) and the S/D contact plugs \( H_c \), the CNT diameter \( d_{\text{CNT}} \), the CNT density \( N_{\text{CNT}} \), and the doping density at the extensions \( n_{\text{SD}} \). The physical parameters are the velocity at the VS \( \upsilon \), the carrier mean free paths in the CNTs (denoted by \( \lambda_c \) and \( \lambda_{\text{ext}} \) for the parts under the S/D metal contacts and extensions, respectively), and the specific contact resistance of the metal–CNT contact \( \rho_c \), all of which can be extracted from the experiments.

The hierarchical structure of the model is shown in Fig. 2. The first level is composed of models for the intrinsic and extrinsic components including the mobility, gate-to-channel capacitance, series resistances, parasitic capacitances, SS and drain-induced-barrier-lowering (DIBL) coefficient. The second level has two models: 1) the VS model utilizing the outputs of the first level to generate the thermionic emission current \( I_{\text{VS}} \) and 2) a semi-analytical model for the tunneling current \( I_{\text{TUNNEL}} \). The final output drain current is given by \( I_D = I_{\text{VS}} + I_{\text{TUNNEL}} \). The details of each part are described below.


A. Models for Intrinsic and Extrinsic Components

1) Mobility ($\mu$): We employ an experimentally corroborated physics-based mobility model for CNTs that applies both to the diffusive and quasi-ballistic transport regimes [34], taking into account the acoustic phonon (AP) and optical phonon (OP) scattering.

$$\mu = \frac{4qL_g}{\hbar n} \sum_i \int_0^\infty \frac{\lambda(E)}{L_g} \left( -\frac{\partial f}{\partial E} \right) dE$$  \hspace{1cm} (1a)

$$\lambda = \frac{1}{\lambda_{AP}(E, T) + \left(1 - f(E + \hbar\omega_{OP})\right)\lambda_{OP, abs}(E, T) + \left(1 - f(E - \hbar\omega_{OP})\right)\lambda_{OP, ems}(E, T)}$$  \hspace{1cm} (1b)

where $n$ is the charge density, $f$ is the Fermi–Dirac distribution, $i$ is the summation index over the first and second subbands, $\hbar\omega_{OP} \approx 0.18$ eV is the OP energy, $\lambda_{AP}$, $\lambda_{OP, abs}$, and $\lambda_{OP, ems}$ are free path for AP scattering, OP absorption, and emission, respectively, which depend on the energy and the temperature. Low-field mobility at a charge density of $n = 0.01$ nm$^{-1}$ is used in the model. Because $\lambda_{AP}$, $\lambda_{OP, abs}$, and $\lambda_{OP, ems}$ are proportional to $d_{CNT}$ [34], [35], $\mu$ increases when $d_{CNT}$ increases.

2) Gate-to-Channel Capacitance ($C_{GC}$): Analytical expression of $C_{GC}$, including the screening effects between multiple CNTs under a single planar gate, has been derived in [36], and the details are presented in the Appendix. $C_{GC}$ is one of the most important factors determining the drive current because it is proportional to the amount of carriers induced by the gate electric field.

3) Series Resistances ($R_S$): The series resistance $R_S$ has two components, one being the resistance of the ungated S/D extensions ($R_{ext}$) [37].

$$R_{ext} = L_{ext}/(\lambda_{ext}G_{1D}) \quad G_{1D} = \frac{4q^2}{\hbar^2} \frac{e^{\Delta/kT}}{1 + e^{\Delta/kT}}$$  \hspace{1cm} (2)

where $\Delta \equiv E_{FS} - E_C$ ($E_V$) for n-type (p-type) CNFETs, and $E_{FS}$ is the Fermi level at the S/D extensions related to the doping density $n_{SD}$. When the CNTs are highly doped, $1/G_{1D} \approx 4q^2/\hbar \approx R_Q = 6.45$ k$\Omega$ is the quantum resistance for CNTs with two-band degeneracy. The other component is the contact resistance ($R_c$) between the CNTs and metal contacts [38].

$$R_c = \frac{1}{4} + \frac{\rho_C}{\lambda_c R_Q} \coth \left( \frac{R_Q}{\rho_C L_c} + \frac{R_Q^2}{4\rho_C^2 L_c} \right)$$  \hspace{1cm} (3)

where $\rho_C$ is equivalent to the reciprocal of the contact conductance $g_c$ in [38]. Note that $\rho_c$ here is different from the conventional contact resistivity for Si MOSFETs. In the transmission line model, $1/\rho_c$ is the conductance per unit length along the CNT. When $L_c$ is smaller than the transfer length, $R_c \approx R_Q/2 + \rho_c/L_c$. The total parasitic resistance for a single CNT is the sum of these two, that is, $R_s = R_{ext} + R_c$.

4) Parasitic Capacitances ($C_P$): The parasitic capacitance of a CNFET consists of two components: the outer-fringe capacitance ($C_{OF}$) between the gate and the CNTs in the S/D extensions, and the gate-to-plug capacitance ($C_{GTP}$) between the gate and the S/D contact plug [36]. The details are described in the Appendix.

5) SS and DIBL Coefficient $\delta$: SS and DIBL coefficient $\delta$ are calculated by the ratio of the capacitances coupling to the VS in the channel [39].

$$SS = \ln(10) \frac{kT}{q} \left( \frac{C_{GC} + C_S + C_D}{C_{GC}} \right)$$  \hspace{1cm} (4a)

$$\delta = C_D/(C_{GC} + C_S + C_D)$$  \hspace{1cm} (4b)

where $C_{GC}$, $C_S$, and $C_D$ are capacitances coupling from the gate, source, and drain to the channel, respectively. The models of $C_S$ and $C_D$ are verified by Maxwell 3-D [40], and the details are in the Appendix. This model can be easily applied to other structures as long as appropriate models for $C_{GC}$, $C_S$, and $C_D$ are used.

B. VS Model

The VS model is a semiempirical model applicable to short-channel MOSFETs in all regions of operation, from ballistic transport to diffusive velocity saturation [26]. The inputs of the VS model include: $\phi$, $\mu$, SS, $\delta$, $R_s$, and $C_p$. $\phi$ can be extracted from experimental data and will be discussed in Section III. $\mu$, SS, $\delta$, $R_s$, and $C_p$ are calculated from the models in Section II-A.

There are two fitting parameters in the VS model, $\alpha$ and $\beta$, which model the channel charge from weak to strong inversion and the transition from linear to saturation region, respectively. As a start, the values $\alpha = 3.5$ and $\beta = 1.4$, borrowed from Si MOSFETs, are used. More experiments are required to determine the specific $\alpha$ and $\beta$ for CNFETs.

With all the inputs of the VS model calculated appropriately in Section II-A and the two fitting parameters empirically set, the VS current can be generated.

C. Tunneling Leakage Current

Two tunneling mechanisms are modeled as illustrated in Fig. 3(a): direct source-to-drain tunneling current ($I_{DSDT}$) from the source conduction band (CB) into the drain CB, and the band-to-band tunneling (BTBT) from the source valence band into the drain CB.

To evaluate the tunneling probability, knowing the band profile along the channel is indispensable. In this paper, we employ a gate-all-around (GAA) cylindrical structure as illustrated in Fig. 3(b) to derive a semianalytical model for the band profile as an estimation of the impact of $I_{DSDT}$, because a planar CNFET structure like that in Fig. 1 has no
analytical solution. The CNT is divided into the channel region and extension regions denoted by I, II, and III in Fig. 3(b). In the subthreshold region, mobile charge is negligible and the surface potential in region I is approximated as the solution to the Laplace’s equation in cylindrical coordinates [41].

\[
\phi_I(z) \approx J_0(\zeta_0) \left[ C_1 \exp(z/\Lambda) + C_2 \exp(-z/\Lambda) \right] + V_{\text{CNT}}
\]

\[
C_2 = \frac{(V_{bi} - V_{\text{CNT}})(e^{z/L_g} - 1) - V_{DS}}{2J_0(\zeta_0) \sinh(L_g/\Lambda)}
\]

\[
C_1 = \frac{V_{bi} - V_{\text{CNT}}}{J_0(\zeta_0)} - C_2
\]

\[
Y_0'(\zeta_0) \frac{J_0(\zeta_0)}{J_1(\zeta_0)} = \kappa_e \frac{Y_0(\zeta_0)}{J_0(\zeta_0)} + (1 - \kappa_e) \frac{Y_0(\zeta_0 + L_o/\Lambda)}{J_0(\zeta_0 + L_o/\Lambda)}
\]

where \( J_0 \) and \( Y_0 \) are Bessel functions of the first and second kinds, respectively, \( \zeta_0 = d_{\text{CNT}}/\Lambda \), \( \Lambda \) is the scale length given in (5b) to satisfy the continuity of perpendicular component of electric field at the CNT/dielectric interface. \( \kappa_e = \epsilon_{\text{CNT}}/\epsilon_{\text{OX}} \) is the ratio of the dielectric constant of CNT to the gate oxide. In this paper, \( \epsilon_{\text{CNT}} = 1 \) and \( \epsilon_{\text{OX}} = 16 \) are chosen. \( C_1 \) and \( C_2 \) are coefficients determined by the boundary conditions \( \phi(0) = V_{bi} \) and \( \phi(L_g) = V_{bi} + V_{DS} \) and \( V_{bi} \) is the built-in potential proportional to \( E_{FI} - E_{FS} \). The reference point is chosen at \( E_{FS} = 0 \) so that the intrinsic Fermi level is equal to \(-q\phi(\zeta)\). To account for the mobile charge induced by the gate, \( V_{\text{CNT}} \) is introduced as the actual voltage dropper on CNTs satisfying

\[
V_{GS} - V_{FB} = qn/C_{\text{OX}} + V_{\text{CNT}}
\]

\[
n = n_1 \exp\left(-\frac{\phi_I(z_{\text{max}})}{kT/q}\right), \quad z_{\text{max}} = \frac{\Lambda}{2} \ln \frac{C_2}{C_1}
\]

\[
n_1 = 4 \sqrt{eE_g/(3\sqrt{\pi}d_{\text{CNT}})} \cdot \exp\left(-E_{G}/(2kT)\right)
\]

\[
C_{\text{OX}} = 2\pi \epsilon_{\text{OX}}/\ln\left[(d_{\text{CNT}} + 2L_0)/d_{\text{CNT}}\right]
\]

where \( V_{GS} \) and \( V_{FB} \) are the gate and flat-band voltages, respectively, and \( z_{\text{max}} \) corresponds to the position of the band maximum in the channel. Derivation of the CNT intrinsic carrier density \( n_1 \) was elucidated in [42].

At the junctions of the channel and the S/D extensions, the electric field does not terminate abruptly but penetrates into the extensions, leading to a tail in the band profile [43] and affects \( I_{DS}\text{DT} \). In this paper, we use an exponential function to phenomenologically model the descending potential in the S/D extensions.

\[
\phi_{II}(z) = \left[ \phi_I(0) - \frac{\phi_{SD}}{q} \right] \exp\left(\frac{\phi_I(0) - \phi_{S}}{q} - z\right) + \frac{\phi_{SD}}{q}
\]

\[
\phi_{III}(z) = \left[ \phi_I(L_g) - \frac{\phi_{SD}}{q} - V_{DS} \right] \exp\left[\frac{\phi_I(L_g) - \phi_{S}}{q} - z\right] + \frac{\phi_{SD}}{q} + V_{DS}
\]

where \( \phi_{SD} = E_{FS} - E_{FI} \), and \( E_{FI} \) is the intrinsic Fermi level. Equation (8) links the band profile in regions I, II, and III smoothly.

Given the potential profile, \( I_{DS}\text{DT} \) in the ballistic transport regime can be evaluated as [44]

\[
I_{DS}\text{DT} = \frac{4q}{h} \int_{E_{FS}/2}^{E_{FS}/2+\phi_{max}} \left[ f(E, E_{FS}) \right. \left. - f(E, E_{FS} - qV_{DS}) \right] dE
\]

where the prefactor of four arises from the double degeneracy of the first sub-band and electron spin. The tunneling probability \( T(E) \) is calculated by the transfer matrix method [45] that takes the band profiles calculated in (6) and (8) as inputs. More details of the calculation of tunneling probability and analytical expressions of band profile will be discussed in a later publication.

BTBT current \( (I_{BTBT}) \) is calculated by the Wentzel–Kramers–Brillouin (WKB) method using the triangular barrier approximation [46].

\[
I_{BTBT} \approx \frac{4q}{\hbar} kT \cdot T_{\text{WKB}} \left[ \ln \left( \frac{e^{(E+qV_{DS})/kT} + 1}{e^{E_{FS}/kT} + 1} \right) - qV_{DS} - E_{FS} + E_{FS}/2 \right]
\]

\[
T_{\text{WKB}} \approx \exp \left( -\frac{\pi}{4} \frac{E_{g}^{2}}{\hbar^{2}q^{2}F} \right)
\]

where \( v_F \sim 10^{6} \text{ m/s} \) is the Fermi velocity, \( F \) is the electric field in the junction at the drain calculated from \( \partial \phi/\partial z \) in (7b). \( T_{\text{WKB}} \) in (9) is a result of the hyperbolic band structure of CNTs and is different from the bulk semiconductors such as silicon or germanium. It is worth noting that phonon-assisted tunneling [47] is not yet included in the model and remains a subject for future works. Therefore, only when \( V_{DS} > E_{g} \), can \( I_{BTBT} \) be appreciable.

\( I_{\text{TUNNEL}} \) is simply the sum of \( I_{DS}\text{DT} \) and \( I_{BTBT} \) and is superimposed on \( I_{GS} \) obtained in Section II-B by matching the threshold voltage \( V_T \), which is defined as the \( V_{GS} \) for which the derivative of transconductance \( \partial g_m/\partial V_{GS} \) is a maximum [48]. Although the device configuration used to derive the tunneling leakage is different from the planar structure in Fig. 1, it provides an efficient and physically logical means to capture the impact of tunneling for the scaled CNFETs.

The modeled \( I_{DS}\text{DT} \) is compared to an open-source simulator which solves the Poisson and Schrödinger equations self-consistently using the NEGF formalism and calculates the current in ballistic transport regime in CNFETs with a GAA cylindrical geometry and doped S/D extensions [25]. Fig. 4 shows the comparison of the band profile as well as \( I_{DS}\text{DT} \) obtained from the model and the NEGF simulation. In Fig. 4(a) and (b), the potential tails in the S/D extensions broaden the tunneling barrier width. Good agreement between the modeled \( I_{DS}\text{DT} \) and the numerical simulation is observed when high-\( k \) dielectric is used \( (\epsilon_{\text{OX}} > 8) \) as shown in Fig. 4(c) and (d).

III. MODEL CALIBRATION WITH EXPERIMENTS

The CNFET model is fitted to the latest experimental \( I-V \) characteristics for long-channel \( (L_e = 300 \text{ nm}, L_c = 100 \text{ nm}) \) and short-channel \( (L_e = 20 \text{ nm}, L_c = 20 \text{ nm} \) and \( L_e = 9 \text{ nm}, \) and \( L_c = 200 \text{ nm} \) CNFETs [28], [29].
Tunneling model is not included in the calibration, as the tunneling current is not significant in the experimental data. \( d_{\text{CNT}} = 1.2–1.4 \, \text{nm} \) was observed in all the devices. Thus, an average of \( d_{\text{CNT}} = 1.3 \, \text{nm} \) is used in the model when fitting to the data, which corresponds to \( E_g \approx 0.66 \, \text{eV} \). The calibrated \( I-V \) curves along with the experimental data are shown in Fig. 5. The signs of \( V_{GS} \) and \( V_{DS} \) are flipped over to make the device plots \( n \)-type-like for convenience. A local bottom gate was utilized to modulate the carriers in the CNTs, which is different from the modeled structure shown in Fig. 1. Thus, the intrinsic capacitance \( C_{\text{GC}} \) of the back-gated structure is obtained from TCAD Sentaurus [49] and serves as the input. Because of the device structure difference between the experimental device and the model, SS and DIBL are not calculated from (4) and (5) but remain as fitting parameters. To accommodate the notable hysteresis observed in the experiments, the \( V_T \) in the \( I_D-V_D \) characteristics is shifted by a constant value (\( \leq 0.2 \, \text{V} \)) compared with the \( V_T \) extracted from \( I_D-V_{GS} \) characteristics. Finally, the parasitic resistance is purely the contact resistance, since there are no S/D extensions in the experimental devices.

The VS velocity \( \nu \), an empirical parameter to the VS model, and \( \rho_c \) and \( \lambda_c \) of the contact resistance model as in (3) are extracted by tuning these parameters to match the experimental data. \( \nu = 3 \times 10^5 \, \text{m/s} \), \( \rho_c = 420 \, \text{k}\Omega \cdot \text{nm} \), and \( \lambda_c = 250 \, \text{nm} \) are obtained from the calibration. It is worth noting that the \( \nu \) is proportional to the saturation current at high \( V_{GS} \), and \( \rho_c \) determines how fast the \( R_c \) increases with the shrinking \( L_c \). A large \( \rho_c \) leads to large \( R_c \) and less steep slope in the \( I_D-V_D \) characteristics. The diameter-normalized contact resistivity is \( \rho_c d_{\text{CNT}} = 546 \, \text{k}\Omega \cdot \text{nm}^2 \), a little higher than the one reported in [28]. Because the devices used for calibration have no extension regions, \( \lambda_{\text{ext}} \) cannot be extracted and is assumed to be equal to \( \lambda_c \). When the technology for CNT doping is better developed, \( \lambda_{\text{ext}} \) can be characterized more accurately.
projected number in Table I is in a certain range, for example, $L_{\text{pitch}} = 0.5 \times (30 + 40) = 35$ nm for the 7-nm technology node. Following the definition in Fig. 1(a), $L_{\text{ext}}$ is equal to $0.5 \times (L_{\text{pitch}} - L_c - L_g)$. Section IV-A to IV-C examines the impact of series resistance, parasitic capacitance, and tunneling leakage current on the CNFET’s performance, as the feature size is scaled down.

### A. Series Resistance

Series resistance increases rapidly with the downscaling of device dimensions and becomes dominant starting from the 11-nm technology node as shown in Fig. 6. The channel resistance is calculated by $R_{\text{ch}} = R_{\text{tot}} - 2R_{\text{ext}} - 2R_c$, where $R_{\text{tot}}$ is the total resistance in the on-state, that is, $R_{\text{tot}} = V_{\text{DD}}/I_{\text{ON}}$. $R_{\text{ch}}$ increases with the technology nodes, because higher $V_T$ is required to keep $I_{\text{OFF}}$ low due to increasing SS. Unlike Si MOSFETs in which the parasitic capacitance is the limiting factor for advanced technology nodes [19], today’s CNFETs are more limited by the CNT-to-metal contact resistance [52].

As described in Section II, the series resistance is composed of the extension resistance $R_{\text{ext}}$ and the contact resistance $R_c$. $R_{\text{ext}}$ is proportional to the ratio of $L_{\text{ext}}$ to the mean free path $\lambda_{\text{ext}}$, while the contact resistance has a more complex dependence on the physical parameters and the contact length. The physics that determine the resistance of the CNT/metal contact are: 1) carrier injection from CNT to metal, whose rate is proportional to $1/\rho_c$, and (2) carrier transport in the CNTs that are covered by the contact metal, which is characterized by $\lambda_c$. In the quasi-ballistic regime, $1/\lambda_c$ is the scattering probability in CNTs per unit length and is highly dependent on the CNT growth and doping conditions. A wide range of $\lambda_c$ from 20 to 380 nm has been reported in [38] and [53]. A pessimistic estimation of $\lambda_c = 15$ nm is used in the following analysis. Because of the distributed nature of the CNT–metal interface, $R_c$ decreases with a hyperbolic cotangent dependence on $L_c$ [54], as shown in Fig. 7, with the lower bound equal to the quantum resistance. The first derivative of $R_c$ with respect to $L_c$ reaches 1 kΩ/nm when $L_c < 20$ nm, demarcating the region where the $R_c$ starts to shoot up. This is a manifestation of the “short contact regime” where $R_c$ starts to increase drastically with decreasing $L_c$. At 7-nm node with $L_c = 10$ nm, $\rho_c$ has to be reduced to 120 kΩ·nm in order to achieve $2R_c < 0.3R_{\text{tot}}$, around 30% of its original value.

### B. Parasitic Capacitance

The ratio of intrinsic capacitance $C_{\text{GC}}$ to the parasitic capacitances ($C_{\text{OF}} + C_{\text{GTP}}$) decreases from 69% to 60% when scaling from 22-nm node to 7-nm node. Parasitic capacitances impose extra burdens on the device, which not only slow down the switching but also raise the dynamic power consumption. Both capacitances $C_{\text{OF}}$ and $C_{\text{GTP}}$ are functions of $L_{\text{ext}}$ as shown in Fig. 8. As $L_{\text{ext}}$ is shortened, $C_{\text{GTP}}$ increases reciprocally due to stronger gate-to-plug coupling, while $C_{\text{OF}}$ decreases proportionally with $L_{\text{ext}}$. $C_{\text{GTP}}$ and $C_{\text{OF}}$ intersect around $L_{\text{ext}} = 20$ nm at $H_g = 30$ nm. Beyond
Lg gate length is scaled down to 9 nm and beyond. SS is thus drastically affected as the distances, capacitances, and tunneling leakage current increase with the reduction in length. Reduction in Hg can effectively reduce CGTP and push the intersection of CGTP and COFF toward the left allowing further scaling of Lext. Another benefit of shortening Lext is to reduce one of the series resistances Rext. Therefore, there is a tradeoff between parasitic capacitance and series resistance.

C. Tunneling Leakage Current

Direct source-to-drain tunneling becomes significant when the gate length is very short. As shown in the inset of Fig. 9, IDSDT dominates over thermionic emission current when the gate length is scaled down to 9 nm and beyond. SS is thus deteriorated with decreasing Lg. Note that the calculation of IDSDT in Fig. 9 is based on the GAA cylindrical structure. Whether direct source-to-drain tunneling was appreciable in [29] needs to be further investigated. It has been reported that the metal/CNT contacts also play an important role in the I–V characteristics and affect the SS because the contacts are modulated by the bottom gate [55]. Temperature-dependent measurements might be helpful to elucidate the impact of tunneling leakage current [23].

On the other hand, BTBT occurs when the drain bias (VDS) is larger than the bandgap or the barrier is raised so high by the gate voltage that the valence band in the channel is lifted above the CB at the source. To avoid BTBT, CNTs with smaller diameters are preferred due to larger bandgap as well as lower tunneling probability.

V. Design Space and Structure Optimization

The primary driver for technology scaling is to reduce cost by shrinking the device pitch Lpitch, which is equal to (Lc + 2Lext + Lg). The effects of scaling the individual components have been discussed in Section IV: parasitic resistances, capacitances, and tunneling leakage current increase drastically as Lc, Lext, and Lg are scaled down, respectively. This paper aims to optimize the ratio of Lc, Lext, and Lg for fixed Lpitch for the 11- and the 7-nm technology node to minimize the gate delay as a demonstration of the model’s capability. As a metric for the device performance, the gate delay is defined as \( \tau = \frac{CN_VDD}{I_{ON}} \), where \( CN = CGC + COG + 2CGD = C_{GC} + 3(C_{OF} + CGTP) \) including the Miller effect. The threshold voltage \( V_T \) is determined by maximizing \( I_{ON} \) under the constraint of \( I_{OFF} \leq 0.5 \mu A/\mu m \) at \( NCNT = 250/\mu m \). Using this model, one can, for example, explore the interaction between the nanotube diameter and the leakage current; the interplay between the contact resistivity, contact length, and device pitch; the selection of the proper work function for the gate electrode; and the choice of the power supply voltage. These explorations will be part of a future study to examine the energy–delay tradeoffs for CNFET device/circuit co-optimization.

The CNFET dimensions can be found in Table I. Lpitch is restrained at 47.5 nm for the 11-nm node and 35 nm for the 7-nm node. The optimization result and the explored design space for the 11-nm node is shown in Fig. 10. The dark blue regions represent the forbidden designs due to intolerable Ioff (>0.5 \( \mu A/\mu m \)) or Lext < 1 nm. The delay can be greatly reduced by increasing Lc (or reducing Rext).

It is essential to investigate the tradeoffs between them. In this section, we optimize the ratio of Lc, Lext, and Lg with fixed Lpitch for the 11- and the 7-nm technology node to minimize the gate delay as a demonstration of the model’s capability. As a metric for the device performance, the gate delay is defined as \( \tau = \frac{CN_VDD}{I_{ON}} \), where \( CN = CGC + COG + 2CGD = C_{GC} + 3(C_{OF} + CGTP) \) including the Miller effect. The threshold voltage \( V_T \) is determined by maximizing \( I_{ON} \) under the constraint of \( I_{OFF} \leq 0.5 \mu A/\mu m \) at \( NCNT = 250/\mu m \). Using this model, one can, for example, explore the interaction between the nanotube diameter and the leakage current; the interplay between the contact resistivity, contact length, and device pitch; the selection of the proper work function for the gate electrode; and the choice of the power supply voltage. These explorations will be part of a future study to examine the energy–delay tradeoffs for CNFET device/circuit co-optimization.

The CNFET dimensions can be found in Table I. Lpitch is restrained at 47.5 nm for the 11-nm node and 35 nm for the 7-nm node. The optimization result and the explored design space for the 11-nm node is shown in Fig. 10. The dark blue regions represent the forbidden designs due to intolerable Ioff (>0.5 \( \mu A/\mu m \)) or impractical device structure (Lext < 1 nm). Compared with conventional 0.7x scaling rules, \( \tau \) is improved from 0.62 to 0.48 ps for the 11-nm node after the optimization of the ratio of the gate, contact, and extension lengths. For 7-nm node, \( \tau \) is improved from 0.63 to 0.55 ps (explored design space is not shown here). Note that the optimized Lc is much larger than its original value (Lc = 12–23 nm for the 11-nm node and Lc = 10–15 nm for the 7-nm node), indicating the significance of reducing contact resistances. Assuming that \( p_c \) could be reduced by half (420–210 KΩ-nm), the optimized \( \tau \) for the 7-nm node would be further reduced from 0.55 to 0.42 ps.

The major obstacle for Lg scaling is direct source-to-drain tunneling leakage that prevents Lg from scaling below 10 nm. Employing CNTs with a smaller diameter in a conventional CNFET can reduce the tunneling probability due to a larger bandgap at the cost of increasing effective mass. Constraint of
Lg scaling would be more rigorous when the process variation is included, and can also be evaluated by the model and alleviated by a careful design. Another information from the optimization of the 11- and 7-nm nodes is that the minimum gate delay is increased from the 11-nm node to the 7-nm node, contradicting the traditional Dennard-scaling expectation [56]. This occurs mainly because the decrease in the contact length greatly raises the contact resistance and reduces the drive current.

VI. CONCLUSION

In this paper, we introduced a compact, physical, and intuitive CNFET model capturing both intrinsic and extrinsic device properties, and the model could be implemented in SPICE or VerilogA. The model was calibrated with the latest experimental results from 300-nm down to 9-nm gate lengths. Based on the GAA cylindrical configuration, the irregular potential profile along the channel was modeled semianalytically, providing an efficient path to study the impact of the tunneling leakage current in the ultrascaled devices. Through careful optimization of the device structure, made possible by the use of this compact model, the impact of the extrinsic components could be alleviated, and the projected gate delay could be improved by more than 20%. From the exploration of the design space, we observed that: 1) contact resistance is the key limiter of the CNFET performance. Substantial improvement in delay can be achieved if $r_c$ is reduced, showing the importance of improving the CNT–metal interface and 2) direct source-to-drain tunneling limits the downscaling of $L_g$. This result is believed to be universal in all kinds of FETs suggesting further study in material and device structure to minimize the tunneling current.

APPENDIX

The intrinsic and extrinsic capacitances are derived in [36]. The gate-to-channel capacitance $C_{GC}$ including screening effects between multiple CNTs under a single planar gate is as shown in (A.1), where $C_{GC,	ext{INF}}$, $C_{GC,	ext{E}}$, and $C_{GC,	ext{M}}$ are $C_{GC}$ for single CNT and for CNTs at the edge and in the middle of the CNT array, respectively; $C_{GC,	ext{SR}}$ is the equivalent series capacitance due to channel screening; $s$ is CNT pitch; $k_1$ and $k_2$ are the relative permittivity of the oxide and substrate, respectively; $\varepsilon_0$ is the vacuum permittivity; $r$ is the radius of CNT; and $h$ is the distance between the center of CNT and the gate.

$$C_{GC,	ext{E}} = C_{GC,	ext{INF}} \cdot C_{GC,	ext{SR}}/(C_{GC,	ext{INF}} + C_{GC,	ext{SR}})$$

$$C_{GC,	ext{M}} = 2C_{GC,	ext{E}} - C_{GC,	ext{INF}}$$

$$C_{GC,	ext{SR}} = \frac{4\pi k_1 \varepsilon_0 L_g}{\ln \left( \frac{\sqrt{s^2 + 2(b-h)(b+\sqrt{h^2-r^2})}}{\sqrt{s^2 + 2(b-h)(b-\sqrt{h^2-r^2})}} \right) + \lambda_1 \ln \left( \frac{(b+2r)^2 + r^2}{9r^2 + s^2} \right) \tanh \left( \frac{h+r}{s-\rho} \right)}$$

$$C_{GC,	ext{INF}} = \frac{\cosh^{-1}\left( \frac{(t_{\text{ox}} + r)}{r} \right) + \lambda_1 \ln \left( \frac{(b + 2r)}{3r} \right)}{k_1 - k_2}$$

$$\lambda_1 = \frac{k_1 - k_2}{k_1 + k_2}$$

(A.1)


Jieying Luo (S’11) received the B.S. and M.S. degrees from the Department of Physics, Peking University, Beijing, China, and Department of Electrical Engineering, Stanford University, Stanford, CA, USA, where she is currently pursuing the Ph.D. degree.

Lan Wei (S’06–M’11) received the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 2010. She is a Technical Staff Member with the Altera Corporation, San Jose, CA, USA.

Chi-Shuen Lee is currently pursuing the Ph.D. degree in electrical engineering with Stanford University, Stanford, CA, USA. His research focuses on modeling and simulation of nanoelectronic and carbon nanotube devices and analysis of circuit-level performance.

Aaron D. Franklin (M’09) received the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 2008. His research focuses on integration of carbon-based nanomaterials into electronic devices, including high-performance transistors, thin-film transistors, supercapacitors, and photovoltaic cells.

Eric Pop (M’99–SM’11) received the Ph.D. degree from Stanford, University, Stanford, CA, USA, in 2005. He is an Associate Professor of electrical and computer engineering with the University of Illinois Urbana-Champaign, Urbana, IL, USA.

Ximeng Guan (S’06–M’11) was a Post-Doctoral Scholar with the Department of Electrical Engineering, Stanford University, Stanford, CA, USA. He is currently with the IBM Semiconductor Research and Development Center, Hopewell Junction, NY, USA, working on logic performance benchmarking.

Jieying Luo

Dimitri A. Antoniadis (M’79–SM’83–F’90) is the Ray and Maria Stata Professor of electrical engineering at the Massachusetts Institute of Technology, Cambridge, MA, USA.

His current research interests include nanoscale electronic devices in Si, Ge, and III–V materials.

H.-S. Philip Wong (S’81–M’82–SM’95–F’01) received the Ph.D. degree from Lehigh University, Bethlehem, PA, USA.

He is the Willard R. and Inez Kerr Bell Professor with the School of Engineering, Stanford University, Stanford, CA, USA.