A Compact Virtual-Source Model for Carbon Nanotube Field-Effect Transistors in the Sub-10-nm Regime—Part I: Intrinsic Elements

Chi-Shuen Lee, Eric Pop, Senior Member, IEEE, Aaron D. Franklin, Senior Member, IEEE, Wilfried Haensch, Fellow, IEEE, and H.-S. Philip Wong, Fellow, IEEE

Abstract—We present a data-calibrated compact model of carbon nanotube (CNT) field-effect transistors (CNFETs) based on the virtual-source (VS) approach, describing the intrinsic current-voltage and charge-voltage characteristics. The features of the model include: (i) carrier VS velocity extracted from experimental devices with gate lengths down to 15 nm; (ii) carrier effective mobility and velocity depending on the CNT diameter; (iii) short channel effect such as inverse subthreshold slope degradation and drain-induced barrier lowering depending on the device dimensions; (iv) small-signal capacitances including the CNT quantum capacitance effect to account for the decreasing gate capacitance at high gate bias. The CNFET model captures dimensional scaling effects and is suitable for technology benchmarking and performance projection at the sub-10-nm technology nodes.

Index Terms—carbon nanotube (CNT), carbon-nanotube field-effect transistor (CNFET or CNTFET), compact model, technology assessment.

I. INTRODUCTION

CARBON nanotube field-effect transistors (CNFETs) based on single-walled semiconducting CNTs have been among the foremost candidates to complement Si and extend CMOS technology scaling in the sub-10-nm technology nodes [1-3]. One of the dominant factors impeding further scaling of Si metal-oxide-semiconductor field-effect transistors (MOSFETs) is the short-channel effect (SCE), which causes FETs at short gate lengths to be difficult to turn off, consequently consuming too much power [4]. To keep Si MOSFETs viable, a tremendous amount of effort has been put into transitioning from two-dimensional (2D) planar device structures to three-dimensional (3D) channel geometry with fin structures [5]. However, further scaling the gate length \( L_g \) of Si-MOSFETs requires an ultra-thin channel body [4], resulting in low drive current due to mobility degradation (caused by the body thickness fluctuation [6]) and low density of states (DOS) [7].

By contrast to bulk 3D materials, a single-walled CNT is essentially a single sheet of graphene rolled into a seamless cylinder with a 1-2 nm diameter. Because of the atomically thin body, the gate control of CNFETs is superior and the SCE can be overcome even for \( L_g \) < 10 nm [1-2,8]. Furthermore, CNTs show promise for energy-efficient computation because of their high carrier velocity and near-ballistic carrier transport property [9-10]. A great deal of effort has been made to demonstrate the potential of CNFETs as the future transistor technology. Recent progress in CNFET technology include a 9-nm \( L_g \) CNFET [2], realization of a gate-all-around (GAA) device [11], complementary n- and p-type CNFETs [12], operation at low (0.4 V) voltages [13], and the demonstration of a simple CNT computer [14]. Despite their great potential, CNFETs suffer from imperfections such as difficulty in obtaining purely semiconducting CNTs [15], hysteresis of the current-voltage (I-V) characteristics [16], mis-positioning and diameter variations [17]. Several techniques have been reported to overcome these imperfections from the fabrication process level up to the system architecture level [18], and more improvement is needed to realize CNFET-based electronics.

For all emerging technologies, early assessment based on both experimental observation and theoretical study is of great value as it facilitates identification of the most promising options and allows resources to be focused on them. Non-equilibrium Green’s function (NEGF) formalism [19], recognized as a physically rigorous approach, has been extensively employed to simulate quantum transport in CNFETs and assess their performance [8,20]. However, NEGF is too computationally expensive for performance assessment at the application level. Compact modeling based on the Landauer formula for ballistic transport in CNTs is another efficient approach for performance assessment of CNFETs [21-23]. While some of these compact models have been validated by numerical simulation or experimental data, the effects of dimensional scaling, series resistance \( (R_s) \), and tunneling leakage current have not been well captured. Attempts were made to address this issue by lumping the scaling and parasitic effects into constant input parameters (e.g. constant \( R_s \), mobility, and subthreshold slope) independent of the device design [21].
As a result, the dimensional scaling effect and variations cannot be studied.

In this paper, we describe a data-calibrated compact CNFET model based on the virtual source (VS) approach [24], which has been implemented in Verilog-A [25] and available online [26]. The main motivation of developing the VS-CNFTET are two-fold: (i) to identify the required improvement in device and materials to achieve performance advantage over similarly scaled FETs, and (ii) to enable performance assessment of CNFET systems at the application level, including device non-idealities and variations. In the model, the VS parameters (e.g. carrier mobility, velocity, and gate capacitance) are connected to the CNFET dimensions and CNT diameter in order to capture the scaling effect. A similar concept has been reported in [27] with preliminary results that did not include several important effects: small-signal capacitances were not properly modeled; CNT quantum capacitance was not considered; the internal VS parameters were independent of CNT diameter; iterations and numerical integral were needed. These deficiencies are addressed in this paper.

Several premises are relied upon in this work: (i) we focus purely on MOSFET-like CNFETs with Ohmic metal-CNT contacts, because they provide better performance and could be realized by heavily doping the source/drain (S/D) extensions [9,28]. Previous efforts on modeling Schottky-barrier CNFETs can be found in [29]; (ii) n-type CNFETs are discussed throughout the paper. Although CNFETs in ambient air are usually p-type based on the preferred injection of holes at the contacts, n-type CNFETs have been achieved by contact or interface engineering [30-31], and from a physical and mathematical point of view the operation of n-type and p-type CNFETs is symmetric due to the symmetry of conduction and valence bands; (iii) only the first sub-band in CNTs is considered because most digital applications call for a low power supply voltage, but higher sub-bands can be easily included with proper modification of the charge model.

This paper is organized as follows: analytical expressions that connect the VS parameters to the CNFET design as well as model calibration are described in Section II; the charge model used to derive the small-signal capacitances is introduced in Section III; in Section IV, the impact of CNT diameter on the intrinsic CNFET performance is presented; finally in Section V, issues pertaining to the VS parameter extraction from CNFETs are discussed. Due to the limited space, the complete derivation of all the equations is detailed in [26]; here we only discuss the physics and key results. Models for the extrinsic elements such as contact resistance and tunneling leakage current will be introduced in Part II of this two-part paper [32].

II. VIRTUAL SOURCE MODEL FOR CNFETS

The VS model is a semi-empirical model with only a few physical parameters, originally developed for short-channel Si MOSFETs that have a gate-controlled source-injection barrier. Recently an enhanced VS emission-diffusion model applicable to both long-channel FETs (in drift-diffusive carrier transport regime) and short-channel FETs (in quasi-ballistic regime) has been reported [33]. Here the VS-CNFTET model is based on the VS model described in [24,34] because we focus on the short-channel FETs (e.g. \( L_g < 30 \) nm) where the carrier transport is assumed to be quasi-ballistic. Based on the VS approach, the drain current \( (I_d) \) of a MOSFET is the product of the mobile charge density and the carrier velocity at the VS, defined as the top of the energy barrier near the source in the on-state, where the lateral electric field is small and the potential is mostly controlled by the gate voltage [24]. There are ten VS parameters: gate length \( (L_g) \); gate capacitance in strong inversion region \( (C_{inv}) \); low-field effective mobility \( (\mu) \); threshold voltage \( (V_t) \); inverse subthreshold slope (SS) factor \( (n_{ss}) \); drain-induced barrier lowering (DIBL) coefficient \( (\delta) \); series resistance \( (R_s) \); VS carrier velocity \( (V_{xo}) \); and fitting parameters \( \alpha \) and \( \beta \) used to smooth the transitions between weak and strong inversion, and between non-saturation and saturation regions, respectively. As conceived originally, the VS model was not meant to be predictive because the VS parameters need to be extracted from current-voltage \( (I-V) \) and capacitance-voltage \( (C-V) \) measurements; the VS-CNFTET model instead associates the VS parameters to the device dimension and CNT diameter so that the CNFET design is connected to the device-level characteristics. Furthermore, by calibrating the VS-CNFTET model to experimental data and rigorous numerical simulations, it becomes possible to make predictive estimates of device behavior as the dimension scales down. A representative GAA-CNFTET device structure used in the VS-CNFTET model is illustrated in Fig. 1 with the critical dimensions labeled. In this section, analytical models to bridge the VS parameters to the device dimension and CNT diameter are introduced.

CNT diameter \( (d) \) is a crucial physical parameter because it determines the CNT band structure and the band gap \( (E_g) \). In this work, \( E_g = 2E_p\pi a_{cc}/d \) is derived from the Hückel tight-binding model [35], where \( E_p = 3 \) eV is the tight-binding parameter, and \( a_{cc} = 0.142 \) nm is the carbon-carbon distance in CNTs, indicating \( E_g \approx 0.85/d \) eV with \( d \) in nm. Corrections to the model of \( E_g \) could be made due to band gap renormalization induced by many-body interaction [36] or substrate-induced polarization effects [37], but they do not alter the essence of the VS model presented here. As will be shown later in this section, \( C_{inv}, \mu, V_t, n_{ss}, \delta, \) and \( v_{xo} \) are all diameter-dependent in the VS-CNFTET model.

A. Inversion Gate Capacitance \( (C_{inv}) \)
In a MOSFET, the mobile charge density in strong inversion at
the VS, where the gradual channel approximation applies
[38], can be approximated as $Q_{so} \approx C_{inv}(V_{gs}-V_{th})$, where
$C_{inv} = C_{ox}C_{s}(C_{ox}+C_{s})$, $C_{ox}$ is the gate oxide,
$C_{s}$ is the semiconductor capacitance defined as $-dQ_{so}/dV_{gs}$,
and $V_{gs}$ is the surface potential [39]. For a GAA structure,
$C_{ox}$ is:

$$C_{ox} = \frac{2\pi \varepsilon_{0} \varepsilon_{r}}{\ln\left(\frac{2t_{ox}+d}{d}\right)}$$

where $\varepsilon_{0}$ is the permittivity in vacuum, $t_{ox}$ and $k_{ox}$ are the
thickness and the relative dielectric constant of the gate oxide,
respectively. In planar bulk semiconductor materials, the DOS
is usually so large that $C_{s} \gg C_{ox}$ and $C_{inv} \approx C_{ox}$; however,
for CNTs, the CNT quantum capacitance ($C_{q}$) needs to be
considered because $C_{q}$ is comparable to $C_{ox}$ due to the relatively
low DOS. In the VS-CNFET model, $C_{inv} = C_{ox}C_{qe}(C_{ox}+C_{qe})$,
where $C_{qe}$ is an empirical parameter representing an effective
CNT quantum capacitance. Strictly speaking, $C_{q}$ is bias dependent [40] as shown in Fig. 2a. However, the numerical
simulation in Fig. 2b shows that the linear relation between $Q_{so}$
and $V_{gs}-V_{th}$ in the inversion region is still retained over a
reasonable range of $V_{gs}$ and $Q_{so}$, implying the viability of having
a constant $C_{q}$ to account for the effect of quantum capacitance.
The numerical simulator used to validate the VS-CNFET model
throughout this paper is provided by [41], which simulates a
GAA-CNFET with heavily doped S/D regions, and the carrier
transport is simulated based on the NEGF formalism. It has
been shown in [40] that the maximum CNT $C_{q}$ is approximately
proportional to $E_{g}^{1/2}$, here $C_{qe}$ is empirically modeled as
$C_{qe} = 0.64E_{g}^{1/2} + 0.1$ (fF/μm), where $E_{g}$ is in the unit of eV,
and the coefficients are extracted by fitting the modeled $Q_{so}$ to the
numerical simulation in Fig. 2b. The modeled $Q_{so}$ is calculated
by substituting $C_{qe}$ into the equation in [26, Eq. (1.3)].

B. Carrier Mobility ($\mu$)

As $L_{g}$ scales down to nanoscale, the carrier transport
approaches the ballistic limit and carrier scattering in the
channel becomes less significant. In this paper, the mobility is
the so-called apparent mobility [38], a concept that connects the
ballistic and diffusive regimes. The apparent mobility could also be understood as another way to express the mean
free path (MFP). As device dimensions become smaller than the
MFP, the carriers travel across the channel nearly without
scattering and scatter only at the source and drain. In this
context, the MFP becomes the channel length.

In the VS-CNFET model, $\mu = GL_{g}/(q_{so})$ is derived from the
one dimensional (1D) quantum transport theory at low fields,
written here for the lowest sub-band [19,42]:

$$G = \frac{4q^{2}}{h} \int_{E_{c}}^{E_{F}} \left[ \lambda_{0}(E) - \frac{\partial}{\partial E} \lambda_{0}(E) \right] dE$$

(2a)

$$n_{i} = \int_{E_{c}}^{E_{F}} (E) f(E,E_{F}) dE$$

(2b)

where $G$ is the CNT conductance, $n_{i}$ is the carrier density, $q_{i}$ is the
elementary charge, $h$ is Planck’s constant, $E_{c}$ is the
conduction band edge, $E_{F}$ is the energy of free electrons, $E_{c}$ is
the Fermi level, $f$ is the Fermi-Dirac distribution function, $g(E)$
is the CNT DOS, and $\lambda_{i}$ is the MFP in CNTs representing the
aggregate effect of optical phonon (OP) and acoustic phonon
(AP) scattering [42]:

$$\frac{1}{\lambda_{i}} = \frac{1}{\lambda_{op}(E,T)} + \frac{1}{\lambda_{op,abs}(E,T)} + \frac{1}{\lambda_{OP,ems}(E,T)}$$

(3)

where $T$ is the temperature, $\lambda_{OP,abs}$ is the OP energy,
$\lambda_{OP,abs}$ and $\lambda_{OP,ems}$ are MFPs for AP scattering, OP
absorption and emission, respectively. The expression for $\lambda_{i}$, its
experimental validation, and treatment across multiple sub-
bands have been detailed in [42] (only the lowest sub-band
is considered here). However, due to the complex expression for $\lambda_{i}$,
Eq. (2a) cannot be integrated analytically; to avoid the use of
a numerical integral in the compact model, an empirical
expression is used to model $\mu$.

$$\mu = \mu_{0} \frac{L_{g}}{\lambda_{i} + L_{g}} (d)^{n}$$

(4)

where $d$ is in the unit of nm, $\mu_{0} = 1350$ cm$^2$/V·s, $\lambda_{i} = 66.2$ nm,
and $c_{i} = 1.5$ are empirical parameters extracted by fitting (4) to
the peak mobility (note that $\mu$ in (2a) depends on $E_{F}$) calculated
by (2) and (3) as shown in Fig. 3. It should be noted that for
device configurations similar to Fig. 1, the source and drain are
in fact separated by $L_{g} + 2L_{ext}$ rather than $L_{g}$. However, since
the extensions are not gated and have higher doping densities
than the region under the gate (thus different MFPs), we treat
the extensions in [32] as extrinsic elements and confine the
scope of intrinsic elements (described by the VS model) to the
region under the gate, leading to a hierarchical model. In
experimental measurements, however, it is not easy to separate
the region under the gate from the extensions and the contacts;

---

Fig. 2. (a) Numerical simulated CNT quantum capacitance $C_{q}$ vs. $V_{gs}$ for different diameters $d$. The peak $C_{q}$ increases as $d$ decreases. (b) Comparison of the VS carrier density $Q_{so}$ between the numerical simulation [41] and the model given (see [26] and [26]). Equivalent oxide thickness (EOT) = 0.7 nm.

Fig. 3. Low-field mobility vs. $L_{g}$ for different diameters. The symbols are the peak mobility calculated numerically by (2) and the lines represent the model given by (4). The mobility decreases towards smaller $L_{g}$ as the conductance becomes constant with quasi-ballistic transport, see (2a).
hence any extraction of mobility for a short-channel CNFET from I-V measurements is actually a reflection of the commingled behaviors of contact injection and carrier transport in the extensions and the channel. Therefore, the use of apparent mobility [38] in the VS model can be viewed as a convenience for describing the experimental I-V curves in a hierarchical model. We note the apparent mobility approaches zero as the channel length (which limits the MFP) approaches zero, consistent with the ballistic limit.

C. SCE Parameters (SS, DIBL, \(V_t\), roll-off)

The SCE is essentially the phenomenon of decreasing \(V_t\) and increasing SS and DIBL as \(L_g\) scales down. In this paper, the SCE parameters are derived from a GAA cylindrical structure based on the scale length theory [43]. The first step is to model the conduction band (\(E_c\)) profile along the channel. In the subthreshold region where the mobile charge in the channel is negligible, the \(E_c\) profile can be obtained by solving the Laplace equation, and the resulting \(E_c\) can be expressed as:

\[
E_c(x) = a_1 e^{-x/a_1} + a_2 e^{-x/a_2} - V_g + E_g / 2 \tag{5}
\]

where \(x\) is the direction along the channel, \(\lambda\) is the electrostatic scale length (also known as the screening length), and \(a_1\) and \(a_2\) are coefficients determined by the boundary conditions: \(E_c(-L_\text{of} - L_d/2) = -E_\text{of} + E_c(L_\text{of} + L_d/2) = -E_\text{of} - V_0\), where \(L_\text{of}\) is an empirical parameter functioning like an extension of the \(L_g\) that captures the finite Debye length at the gate-to-source/drain junctions, and \(E_\text{of}\) is the energy difference from the Fermi level at the gate-to-source/drain junctions, and \(E_\text{of}\) is the Fermi level at the source (i.e. \(E_{\text{is}} = 0\)).

In a GAA cylindrical structure, \(\lambda\) is a solution to the Laplace equation in cylindrical coordinates satisfying the boundary condition at the CNT/oxide interface:

\[
\frac{Y_m(\zeta)}{J_m(\zeta)} = \gamma \frac{Y_0(\zeta)}{J_0(\zeta)} + (1 - \gamma) \frac{Y_0(\zeta + t_\text{ox}/\lambda)}{J_0(\zeta + t_\text{ox}/\lambda)} \tag{6}
\]

where \(J_m\) and \(Y_m\) are Bessel functions of the first kind and second kind of order \(m\), \(\gamma = k_{\text{cnt}}/k_{\text{ox}}\), \(k_{\text{cnt}}\) is the relative dielectric constant of the CNT, and \(\zeta = d/(2\lambda)\). Eq. (6) is a transcendental equation which has no closed form solution for \(\lambda\). Analytical approximations of \(\lambda\) in GAA-MOSFETs have been derived in [44] by assuming that the \(E_c\) profile is parabolic in the transverse direction; however for CNFETs, \(d\) is often smaller than \(t_\text{ox}\), so the approximation made in [44] fails. When \(t_\text{ox} > d/2\), we show that \(\lambda\) can be approximated as:

\[
\lambda = \frac{d + 2t_\text{ox}}{2z_0} \left[ 1 + b \left( \gamma - 1 \right) \right]
\]

\[
b = 0.41 \left( \zeta_0 / 2 - \zeta_0^3 / 16 \right) \left( \pi \zeta_0 / 2 \right)
\]

\[
\zeta_0 = z_0 d / (d + 2t_\text{ox})
\]

where \(z_0 \approx 2.405\) is the first zero of \(J_0\). Derivation of (7) is detailed in [26, Eq. (15)]-[19]). Eq. (7) is compared with the numerical solution to (6) in Fig. 5, showing good agreement when \(t_\text{ox} > d/2\). When \(t_\text{ox} >> d\), Eq. (7) can be simplified to \(\lambda \approx (d + 2t_\text{ox})/z_0\), on the other hand, when \(t_\text{ox} < d\), it has been shown in [45] that \(\lambda \approx (d + 2t_\text{ox})/2z_0\). In both extreme cases, \(\lambda\) increases linearly with \(d\) and \(t_\text{ox}\). In this paper, \(k_{\text{cnt}} = 1\) is used, assuming it is air inside the CNT [46]. However, different values of \(k_{\text{cnt}}\) from 5-10 for semiconducting CNTs have been reported both theoretically [47] and experimentally [48]. Nonetheless, we can show that (7) holds for a wide range of \(k_{\text{cnt}}\) (from 1-20).

By substituting (7) into (5), \(E_c\) profile is calculated and compared to the numerical simulation [41] in Fig. 4, showing good agreement in the gate region. Although the potential “tails” extending into the S/D extensions are not captured by (5), this will not affect the calculation of SCE parameters since only the top of the \(E_c\) \((E_c(\text{max}))\) matters. Modeling of the tails will be discussed in [32] when calculating the tunneling currents. Once the \(E_c\) profile is known, the SCE parameters can be derived as:

\[
n_{\text{ss}} = -\partial E_{\text{c max}} / \partial V_g \big|_{V_g=0} = \left( 1 - e^{-\eta} \right) \tag{8.1}
\]

\[
\delta = -\partial E_{\text{c max}} / \partial V_g \big|_{V_g=0} = e^{-\eta} \tag{8.2}
\]

\[
-\Delta V_t = E_g / 2 - E_{\text{c max}} \big|_{V_g=0} = \left( 2E_{\text{c max}} + E_g \right) e^{-\eta} \tag{8.3}
\]

where \(\eta = (L_\text{of} + 2t_\text{of}) / 2\lambda\), and \(E_{\text{c max}}\) is calculated by substituting \(x = -\lambda / 2 \ln(\alpha / a_1)\) into (5). Eq. (8) is compared to the numerical simulation in Fig. 6. Empirically, \(L_\text{of} \approx t_\text{of}/3\) is found to achieve the best fitting results. A physical interpretation of the relation between \(L_\text{of}\) and \(t_\text{of}\) is that when \(t_\text{of}\) becomes larger, the fringe field from the gate to the S/D extensions will extend, making \(L_\text{of}\) longer. Nevertheless, generally \(L_\text{of}\) should be viewed as a fitting parameter. Note that (8) is a direct result of solving Poisson’s equation without considering non-idealities such as oxide-CNT interface states. Therefore, \(SS \approx 60\) mV/dec and DIBL = 0 for long-channel devices. More discussion on the oxide-CNT interface is included in [32]. Although (8) is derived from a GAA structure, other device structures such as top gate and bottom gate should follow the same trend as long as a proper model for \(\lambda\) is used.
D. Virtual Source Carrier Velocity ($v_{xo}$)

The VS carrier velocity ($v_{xo}$), also known as the injection velocity, is one of the key metrics for the transistor technology [49]. $v_{xo}$ can be associated with $L_g$ through the theory of back scattering of carriers in the channel [50]:

$$v_{xo} = \frac{\lambda_v}{\lambda_v + 2l} v_B$$  \hspace{1cm} (9)

where $v_B$ is the carrier velocity in the ballistic limit, $\lambda_v$ is the carrier MFP, and $l$ is the critical length defined as the distance over which the electric potential drops by $k_B T/q$ from the top of the energy barrier in the channel, where $k_B$ is the Boltzmann’s constant. Strictly speaking, $l$ is proportional to $L_g$ and dependent on $V_{ds}$, as described in [33]. However, since using a bias-independent $v_{xo}$ can fit the experimental $I_{ds}$-$V_{ds}$ data fairly well for different $L_g$’s (as will be seen shortly) and only a small range of $L_g$ is of our interest (e.g. 5 nm < $L_g$ < 30 nm), here $l \approx L_g$ is assumed for the sake of simplicity and $\lambda_v$ is thus empirical. To extract $v_B$ and $\lambda_v$, the VS model [24] is fitted to the $I_{ds}$-$V_{ds}$ data from [51], where three CNFETs on the same substrate with identical structures but different gate lengths were measured.

The extraction flow of $v_{xo}$ involves: (a) $d = 1.2$ nm, $L_g = 15$ nm/300 nm/3 μm, $R_s = 5.5$ kΩ, and SS = 135 mV/dec according to the reported experimental data in [51]; (b) estimating, due to lack of C-V data, $C_{ox} = 0.156$ fF/μm by simulating a metallic cylinder placed on a 10-nm thick HfO$_2$ with a back gate using TCAD Sentaurus [52]; (c) $\mu = 255/10^2/2.1 \times 10^3$ cm$^2$/V·s for $L_g = 15$ nm/300 nm/3 μm respectively, estimated by (2); (d) $\alpha = 3.5$ and $\beta = 1.8$ as suggested in [24]; (e) DIBL and $V_t$ are treated as free parameters because the two parameters are susceptible to the oxide-CNT and air-CNT interface properties and may suffer from different degrees of the hysteresis effect [16]. Fortunately, the extracted $v_{xo}$ is not sensitive to the choice for DIBL and $V_t$. Finally, $v_{xo}$ is treated as a free parameter to achieve the best fitting result as shown in Fig. 7. If uncertainty exists in the exact value of $d$ due to the measurement, the values of $C_{ox}$ and $\mu$ would be adjusted accordingly and the extracted $v_{xo}$ could be slightly different, but the change will be minor and the scaling trend will remain the same. By fitting (9) to the extracted $v_{xo}$’s, $\lambda_v = 440$ nm and $v_B = 4.1 \times 10^5$ cm/s are extracted. $v_{xo}$ for other materials have been extracted from devices at various $L_g$’s, including 1.35$\times$10$^7$ cm/s for 32-nm $L_g$ Si MOSFET [24] and 3.2$\times$10$^7$ cm/s for 30-nm $L_g$ III-V HFET [53].

To model the dependence of $v_{xo}$ on CNT diameter, we refer to the carrier transport theory in MOSFETs [54]: the maximum value of $v_{xo}$ is approximately the equilibrium uni-directional thermal velocity $v_{Th}$. For the non-degenerate case, $v_{Th} = 2k_B T/(\pi m^*)$, where $m^* = \hbar^2/(2\pi e^2 E_g d)$ is the effective mass in CNTs [40]. Therefore we can express $v_B = v_{Th} (d/d_0)^{1/2}$, where $v_{Th} = 4.1 \times 10^7$ cm/s and $d_0 = 1.2$ nm are extracted from [2] set as reference points. To examine the validity of the linear relation between $v_B$ and $d^{1/2}$, the 1D Landauer formula [19] is used to calculate the theoretical ballistic velocity $v_{Th}$:

$$I_{th} = \frac{4q}{h} \int \left[ f_s (E) - f_d (E) \right] dE$$

$$= \frac{4q}{h} k_B T \ln \left[ \frac{1 + \exp \left( \frac{\psi_s - E_f}{k_B T/q} \right)}{1 + \exp \left( \frac{\psi_d - E_f - 2q - V_{Th}}{k_B T/q} \right)} \right]$$  \hspace{1cm} (10)

where $I_{th}$ is the drain current in the ballistic limit calculated by the 1D Landauer formula, and $v_{Th} = I_{th} / h n_s$, where $n_s$ is calculated by (2b). Fig. 8 shows $v_{Th}$ vs. $d^{1/2}$ for different carrier densities, indicating that the linear relation between $v_{Th}$ and $d^{1/2}$ holds for a wide range of $d$ and $n_s$. 

---

**Fig. 6.** Comparison of (a) SS, (b) DIBL, and (c) $V_t$ roll-off between the numerical simulation [41] and the model given by (8) for different gate oxide thickness. Tunneling currents are not excluded.

**Fig. 7.** Extraction of VS carrier velocity. The symbols are experimental data from [51]. (a) $v_{xo} = 3.8 \times 10^7$ cm/s for $L_g = 15$ nm. (b) $v_{xo} = 1.7 \times 10^7$ cm/s for $L_g = 300$ nm. (c) $v_{xo} = 0.47 \times 10^7$ cm/s for $L_g = 3$ μm. Note that the polarity of $V_{gs}$ and $V_{th}$ are flipped compared to the original data to become n-type FETs.
III. TERMINAL CHARGE MODEL

Proper modeling of the terminal charges is required to account for the dynamic operation of a FET. Under quasi-static conditions, the partitioning of charges at the source ($Q_s$) and the drain ($Q_d$) is accomplished through the Ward-Dutton charge-partitioning scheme [55]; the charge at the gate $Q_g = -(Q_s + Q_d)$; and the derivative of terminal charges with respect to the terminal voltage gives the small-signal capacitances [56]. In a short-channel MOSFET, the carrier transport generally falls somewhere in between the drift-diffusion regime and the ballistic transport regime. The charge model employed in this paper is similar to the VS charge model introduced in [26, pp. 21-24]. This section focuses on a correction term in the charge model to account for the effect of CNT quantum capacitance ($C_q$).

As illustrated in Fig. 2a, the CNT $C_q$ increases as $V_g$ increases from zero to $V_t$; then reach a maximum; and finally decreases asymptotically to $C_{q,\text{int}} = 8\pi^2/(3\alpha_x e E_g)$. The decrease in $C_q$ is because of the rapid drop of CNT DOS after the van Hove singularity [35]. The effect of $C_q$ is not considered in the VS charge model originally developed for silicon MOSFETs. While an analytical model for $C_q$ of CNTs has been developed in [22], the equations are relatively complex, making analytical expressions for $Q_g$ and $Q_s$ hard to obtain. Here, the terminal charge is modeled phenomenologically rather than from first-principles to account for the effect of $C_q$:

$$Q_{\text{ch}} = -L_g (Q_{xo} - Q_{\text{tot}})$$

$$Q_{\text{tot}} = (C_{\text{inv}} - C_{\text{orb}}) \cdot n_x \phi_i \cdot \ln \left(1 + \frac{V_g - [V_0 - \frac{\alpha \cdot \phi_f}{n_x} \cdot F (V_{gs})]}{n_x \cdot \phi_f} \right)$$

where $Q_{\text{ch}}$ is the total channel charge, $\phi_i = k_B T q$ is the thermal voltage, $Q_{\text{tot}}$ serves to gradually decrease the absolute value of $Q_{\text{ch}}$ around $V_{tb}$, and $V_{tb}$ is a fitting parameter to be determined. $Q_s$ and $Q_d$ are proportional to $Q_{\text{ch}}$ as described in [26]. Here we discuss a special case of $V_{ds} = 0$ to demonstrate how the model works. At $V_{gs} = 0$, $Q_s = Q_d = V_s Q_{\text{ch}}$, and the small-signal gate capacitance $C_{\text{gg}} = -1/L_g \partial Q_{\text{ch}}/\partial V_{gs}$. When $V_{gs} < V_t$, $Q_{\text{ch}} \approx 0$, and $Q_{\text{tot}} \approx 0$; as $V_{gs}$ increases to $V_t$, $Q_{\text{ch}} < V_{gs} < V_{ds}$, $|Q_{\text{ch}}| < |Q_{\text{tot}}|$, so $Q_{\text{ch}} \approx -L_g Q_{\text{tot}} \approx L_g C_{\text{inv}} (V_{gs} - V_t)$, and $C_{\text{gg}}$ approaches the peak value $C_{\text{inv}}$; when $V_{gs} \gg V_t$, $Q_{\text{ch}}$ becomes appreciable and $Q_{\text{ch}} \approx -L_g C_{\text{inv}} (V_{gs} - V_t) + C_{\text{inv}} (V_{gs} - V_t))$, and $C_{\text{gg}} \approx C_{\text{inv}}$, as expected when $V_{gs}$ approaches infinity. The modelled $C_{\text{gg}}$ is compared with the numerical simulation [41] in Fig. 9, where $V_{tb} = 0.7 E_g q^{0.13}$ is determined empirically to achieve the best fitting result. Compared to the case where quantum capacitance is not considered, the $C_{\text{gg}}$ including the quantum capacitance is lower and gradually decreases at high $V_{gs}$. The resulting charge model is consistent with the current model because they share the same $V_t$ and $Q_{\text{ch}}$.

IV. CNFET INTRINSIC PERFORMANCE AND CNT DIAMETER

In this section, the impact of CNT diameter on the intrinsic CNFET performance is evaluated based on the model described in Section II and III. Inputs to the VS-CNFT model are: $L_g = 8$ nm, supply voltage $V_{dd} = 0.71$ V, and EOT $= 0.51$ nm, selected from the “2023” node of the 2013 International Technology Roadmap for Semiconductors (ITRS) projections [58] which predicts the metal-1 pitch will be scaled down to 25.2 nm in 2023 for high performance logic; a GAA structure is assumed; and $R_s = R_{oh}/2 = h/(2g_p) \approx 3.3$ kΩ per CNT is added to the source and the drain terminals (see Fig. 1) to account for the quantum resistance associated with the interfaces between the 1D CNT channel with the metal S/D contacts (including the lowest band double degeneracy with two spins) [19].

In Fig. 10, the on-state current $I_{on} = I_d(V_{gs}) = V_{ds} = V_{dd}$ per CNT and the intrinsic delay $\tau_{\text{in}} = L_g C_{\text{inv}} V_{dd}/I_{on}$ are plotted against CNT diameter at a fixed off-state current $I_{off} = I_d(V_{gs} = 0, V_{ds} = V_{dd}) = 1$ nA per CNT. As shown in Fig. 10, a 2-nm diameter CNT can deliver 27% higher $I_{on}$ and 21% lower $\tau_{\text{in}}$ than a 1-nm diameter CNT. While $\mu \sim d^2$ has been observed experimentally in CNFETs with relatively long channels ($L_g > 4$ μm) [59], here we predict the ratio of $I_{on}(d = 2$ nm) over $I_{on}(d = 1$ nm) to be 1.27, much smaller than $2^2/1 = 4$, because the channel has become nearly ballistic at $L_g = 8$ nm. The increase in $I_{on}$ for large-diameter CNTs is attributed to higher carrier mobility, velocity, and gate capacitance. The advantage of large-diameter CNTs in $\tau_{\text{in}}$ is not as prominent as in $I_{on}$. On the other hand, the larger $C_{\text{gg}}$ allows a larger swing in the drain voltage before the intrinsic delay becomes significant.
the gate capacitance is also higher. As will be seen in [32], CNT diameter has greater impacts on the parasitic contact resistance and the tunneling leakage currents in a highly scaled CNFET.

V. DISCUSSION

The VS carrier velocity is a crucial metric for the transistor technology because it directly determines the magnitude of the drive current as well as the delay of logic devices. A major advantage of the VS model is its capability of extracting \( v_{xo} \) directly from the measured data. Normally, the inversion gate capacitance \( C_{inv} \) is obtained from the C-V data. Then with \( C_{inv} \) as one of the inputs, fitting the VS model to I-V data determines \( v_{xo} \) [60]. In other words, both I-V and C-V data are needed in order to reliably extract \( v_{xo} \). For emerging devices like CNFETs, however, reliable and reproducible C-V data are often hard to acquire, because of less understanding of CNT-oxide and CNT-metal interfaces and the very small capacitance (aF range) of the 1D channels [61-62]. In this paper, numerical simulation by Sentaurus [52] is used to estimate the \( C_{inv} \) as a compromise for the extraction of \( v_{xo} \) in Fig. 7. In [27], \( v_{xo} = 3 \times 10^7 \text{ cm/s} \) was extracted from a CNFET with \( L_g = 9 \text{ nm} \) [2], smaller than the \( v_{xo} = 3.8 \times 10^7 \text{ cm/s} \) extracted from the \( L_g = 15 \text{ nm} \) CNFET in Fig. 7a. While the contradiction (i.e., \( v_{xo} \) of a 9-nm CNFET is smaller than that of a 15-nm CNFET) might be attributed to the differences in gate oxide, fabrication conditions, CNT quality, or the long-range Coulomb interactions described in [63], the unexpected trend highlights the necessity for a larger number of consistent and systematic characterization of devices to extract \( v_{xo} \) in CNFETs (e.g., CNFETs built on the same CNT with different gate lengths below 100 nm). These high-quality device data are often not readily available because of the difficulties in device fabrication and the hysteresis and instability of experimental devices.

VI. CONCLUSION

The intrinsic elements of a compact CNFET model based on the VS approach has been developed in this paper. A VS carrier velocity of \( 3.8 \times 10^7 \text{ cm/s} \) is extracted from recent experimental CNFET with 15-nm gate length, providing evidence of the superior potential of CNFETs for future transistor technology. The model captures dimensional scaling effects and is used study the impact of CNT diameter on the intrinsic CNFET performance, showing that a 2-nm diameter CNT can deliver 27% higher intrinsic drive current than a 1-nm diameter CNT at \( L_g = 8 \text{ nm} \). The VS-CNFET model has been implemented in Verilog-A and is available online [26]. The model runs smoothly in the SPICE environment (as illustrated in [64]) because all the equations are analytical with no numerical iterations, and the output current is differentiable throughout all regions of operation. A more comprehensive analysis including non-ideal contacts and tunneling leakage is carried out in [32].

ACKNOWLEDGMENT

The authors would like to thank Prof. Lan Wei (Waterloo) and Prof. Shaloo Rakheja (NYU) for the useful discussion on the VS model; Gage Hills and Prof. Subhasish Mitra at Stanford for testing and using the VS-CNFET model while under development. This work was supported in part through the NCN-NEEDS program, which is funded by the National Science Foundation, contract 1227020-EEC, and by the Semiconductor Research Corporation, and through Systems on Nanoscale Information fabric\( \mathcal{S} \) (SONIC), one of the six SRC STARNet Centers, sponsored by MARCO and DARPA, as well as the member companies of the Initiative for Nanoscale Materials and Processes (INMP) at Stanford University.

REFERENCES

[1] M. Luisier, M. Lundstrom, D. Antoniadis, and J. Bokor, “Ultimate device scaling: intrinsic performance comparisons of carbon-based, InGaAs, and Si field-effect transistors for 5 nm gate length,” IEDM Tech. Dig., pp. 11.2.1-11.2.4, 2011.
[2] A. D. Franklin, M. Luisier, S.-J. Han, G. Tulevski, C. M. Breslin, L. Gignac, M. S. Lundstrom and W. Haensch, “Sub-10 nm Carbon Nanotube Transistors,” Nano Lett., vol. 12, no. 2, pp. 758-762, Feb. 2012.
[3] H. B. P. Wong, S. Mitra, D. Akinwande, Z. Zhang, S. Liang, H. Y. Chen, X. Chen, G. Close, J. Deng, A. Hazeghi, J. Liang, A. Lin, L. S. Liyanage, J. Luo, J. Parker, N. Patil, M. Shulakher, H. Wei, L. Wei and J. Zhang, “Carbon nanotube electronics—Materials, devices, circuits, design, modeling, and performance projection,” IEDM Tech. Dig., pp. 23.1.1-23.1.4, 2011.
[4] D. J. Frank, R. H. Denard, E. Norwak, P. M. Solomon, Y. Taur and H.-S. P. Wong, “Device scaling limits of Si MOSFETs and their application dependencies,” Proc. IEEE, vol. 89, no. 3, pp. 259–288, 2001.
[5] M. Bohr and K. Mistry, “Intel’s revolutionary 22-nm Transistor Technology,” [Online]. Available: http://download.intel.com/newsroom/kits/22nm/pdfs/22nm-details_presentation.pdf.
[6] S.-H. Kim, et al., “Experimental Study on Electron Mobility in InGaAs-Planar-on-Insulator Metal-Oxide–Semiconductor Field-Effect Transistors With In Content Modulation and MOS Interface Buffer Engineering,” IEEE Trans. Nanotechnol., vol. 12, no. 4, pp. 621-628, Jul. 2013.
[7] B. Yu, L. Wang, Y. Yuan, P. M. Asbeck and Y. Taur, “Scaling of Nanowire Transistors,” IEEE Trans. Electron Devices, vol. 55, no. 11, pp. 2846-2858, Nov. 2008.
[8] G. Fiori, G. Iannaccone and G. Klimeck, “A Three-Dimensional Simulation Study of the Performance of Carbon Nanotube Field-Effect Transistors With Doped Reservoirs and Realistic Geometry,” IEEE Trans. Electron Devices, vol. 53, no. 8, pp. 1782-1788, Aug. 2006.
[9] A. Javey, J. Guo, Q. Wang, M. Lundstrom and H. Dai, “Ballistic carbon nanotube field-effect transistors,” Nature, vol. 424, pp. 654-657, Aug. 2003.
[10] Z. Yao, C. L. Kane and C. Dekker, “High-Field Electrical Transport in Single-Wall Carbon Nanotubes,” Phys. Rev. Lett., vol. 84, no. 13, pp. 2941-2944, Mar. 2000.
[11] A. D. Franklin, S. O. Koswatta, D. B. Farmer, J. T. Smith, L. Gignac, C. M. Breslin, S.-J. Han, G. S. Tulevski, H. Miyazoe, W. Haensch and J. Tersoff, “Carbon Nanotube Complementary Wrap-Gate Transistors,” Nano Lett., vol. 13, no. 6, pp. 2490-2495, May 2013.
[12] L. Ding, Z. Zhang, S. Liang, T. Pei, S. Wang, Y. Li, W. Zhou, J. Liu and L. M. Peng, “CMOS-Based Carbon Nanotube Pass Transistor Logic Integrated Circuits,” Nature Commun., vol. 3, pp. 677-683, Feb. 2012.
[13] L. Deng, S. Liang, T. Pei, Z. Zhang, S. Wang, W. Zhou, L. J Liu and L.-M. Peng, “Carbon nanotube based ultra-low voltage integrated circuits: Scaling down to 0.4V,” Appl. Phys. Lett., vol. 100, pp. 263116, 2012.
[14] M. M. Shulaker, G. Hills, N. Patil, H. Wei, H.-Y. Chen, H.-S. P. Wong and S. Mitra, “Carbon nanotube computer,” Nature, vol. 501, no. 7468, pp. 746-748, Sep. 2013.
[15] G. S. Tulevski, A. D. Franklin and A. Afzali, “High Purity Isolation and Quantification of Semiconducting Carbon Nanotubes via Column Chromatography,” ACS Nano, vol. 7, no. 4, pp. 2971-2976, 2013.
[16] A. D. Franklin, G. S. Tulevski, S.-J. Han, D. Shahjerdi, Q. Cao, H.-Y. Chen, H.-S. P. Wong and W. Haensch, “Variability in Carbon Nanotube Transistors: Improving Device-to-Device Consistency,” ACS Nano, vol. 6, no. 2, pp. 1109-1115, Jan. 2012.
[17] Q. Cao, S.-J. Han, G. Tulevski, A. Franklin, and W. Haensch, “Evaluation of Field-Effect Mobility and Contact Resistance of Transistors That Use Solution-Processed Single-Walled Carbon Nanotubes,” ACS Nano, vol. 6, no. 7, pp. 6471-6477, 2012.
