Analysis of submicron carbon nanotube field-effect transistors

Toshishige Yamada[a)]

MRJ, T27A-1, NASA Ames Research Center, Moffett Field, California 94035-1000

A theoretical analysis of carbon nanotube based field-effect transistors fabricated by two different groups [Tans *et al.*, Nature (London) **393**, 49 (1998); Martel *et al.*, Appl. Phys. Lett. **73**, 2447 (1998)] is presented. The metal (electrode)-semiconductor (nanotube) contact influences subthreshold channel conductance versus gate voltage V_G, such that the occurrence of a kink depends on the transport mechanism across this contact. Saturation in the turn-on drain current I_D vs V_G seen in experiments reflects the nanotube state density. Saturationless I_D versus drain voltage V_D indicates transport in the weak-localization regime in the absence of carrier–carrier scattering so that pinch-off cannot occur. To compensate for saturationless $I_D(V_D)$ in digital applications, nanotube transistors need to be designed to maximize their transconductance. © *2000 American Institute of Physics.* [S0003-6951(00)01505-9]

Recently, submicron carbon nanotube field-effect transistors (FETs) have been reported by Delft[1] and IBM groups.[2] These studies used Pt and Au for source/drain electrodes, respectively, so that a metal–semiconductor contact was formed. Both groups used a semiconducting nanotube on a SiO$_2$ layer in a back-gate structure. The experiments showed: (1) The drain current I_D versus drain voltage V_D at fixed gate voltage V_G had no definite saturation at large V_D in both Delft and IBM devices, unlike that in conventional metal–oxide–semiconductor (MOS) FETs. Instead, at a few volt V_D, I_D started to increase rapidly in the Delft device. (2) I_D vs V_G at fixed V_D saturated in the IBM device, unlike that in MOSFETs. The threshold voltage V_T is large, around several volts in both. (3) The channel conductance $g_d(V_G)$ did (did not) show a kink around $V_G = 2$ V in the Delft (IBM) device. Some of these features are common in MOSFETs. For example, the rapid increase in $I_D(V_D)$ is due to electron-hole pair creation.[3] The large V_T is attributed to unwanted interface charges[3] unintentionally introduced during fabrication. In this letter, we instead concentrate on the remaining features.

We first pay special attention to $I_D(V_D)$ saturation, since without it a favorable sharp transition in the transfer characteristics of a FET inverter, a key element in digital circuits, may not be satisfactorily achieved[4] unless the transconductance is extremely large. In conventional long-channel FETs where carrier velocity saturation[3] is irrelevant, the $I_D(V_D)$ saturation is caused by carrier–carrier interaction via Coulomb repulsion.[5] Carriers in the channel move in response to the field created by V_D and other carriers, experience a lot of carrier–carrier scattering, and establish a local equilibrium with nearby carriers. The resultant field is self-consistent and determines I_D under given V_D. Since the voltage drops monotonically from source to drain, there always exists a pinch-off point (POP) with theoretically zero carrier density, whenever $V_D > V_G - V_T$. Since the carrier–carrier scattering is effective only between the source and the POP, and the POP does not move very much with V_D in a thin, long chan-

nel (excess V_D beyond $V_G - V_T$ is exhausted between the POP and the drain), the scattering remains almost the same and $I_D(V_D)$ saturation results.[5,6]

Thus, the saturationless $I_D(V_D)$ in the nanotube FETs possessing an extremely thin, long channel means that carriers can reach the drain without experiencing carrier–carrier scattering. This might be counterintuitive for a 0.3 μm long channel at a room temperature, but ballistic transport (the absence of any scattering) has already been observed in the form of conductance quantization for nanotubes as long as 1 μm *floating in the air*.[7] Since carrier–carrier scattering is not sensitive to the presence of the substrate (while elastic scattering is), its absence in the 0.3 μm *nanotube* channel is not surprising. However, the transport in the Delft and IBM devices was *not* ballistic, as indicated by a g_d in the range of a few microsiemens,[1,2] two orders of magnitude smaller than the quantum conductance $\sigma_Q = 2e^2/h$ (e is the unit charge and h the Planck constant). This small g_d indicates that there is a lot of elastic scattering by impurities, defects, bending, or dents in nanotubes *on a substrate*.[2] In mesoscopia, this is called the weak localization regime.[8] Transport in nanotubes on a substrate is in this weak localization regime, while that in nanotubes floating in the air is in the ballistic regime.

Ideal, theoretical transport characteristics for nanotube FETs in either ballistic or weak localization regime are presented next. Delft and IBM experimental data[1,2] are shown in Fig. 1(a). Whether the metal–semiconductor contact in Fig. 1(b) prevents the wave function from penetrating into the electrodes or not (isolating or penetrating contact) plays a critical role. With isolating contacts, a Coulomb island is formed as in Fig. 1(c), and the number of electrons there is quantized. Therefore, a Coulomb staircase[9] is expected for $I_D(V_D)$ at fixed V_G as in Fig. 1(d), and a Coulomb oscillation[10] is expected for $I_D(V_G)$ at fixed V_D as in Fig. 1(e). In either case, the voltage step/period is characterized by the island capacitance C_Σ. An experiment[11] with such an isolating contact has already demonstrated a Coulomb oscillation in $I_D(V_G)$. With penetrating contacts as in Fig. 1(f), a linear response is expected for $I_D(V_D)$ at fixed V_G as in Fig.

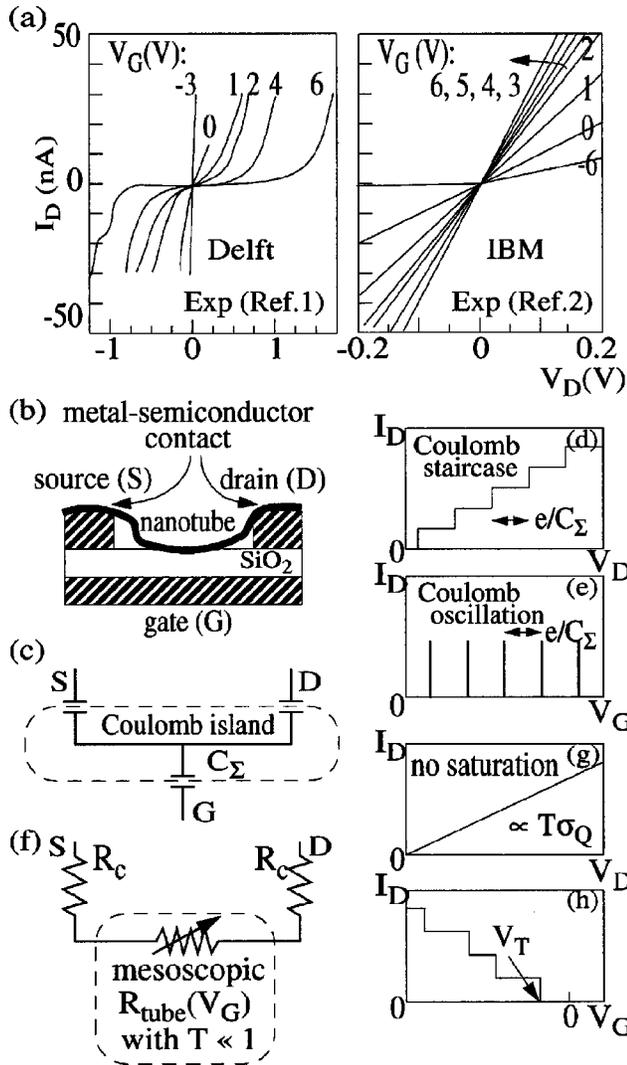[a)]Electronic address: yamada@nas.nasa.gov

FIG. 1. Nanotube FETs: (a) experimental drain current I_D vs drain voltage V_D at fixed gate voltage V_G of Refs. 1 and 2; (b) device structure; (c) equivalent circuit with isolating contact, causing such characteristics as (d) I_D vs V_D and (e) I_D vs V_G; (f) equivalent circuit with penetrating contact, causing such characteristics as (g) I_D vs V_D and (h) I_D vs V_G.
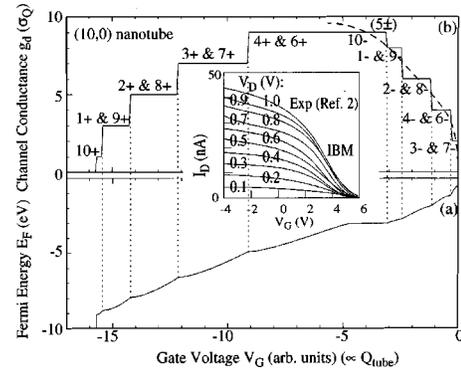

FIG. 2. Turn-on characteristics for (10, 0) nanotube FETs: (a) Fermi energy E_F and (b) channel conductance g_d, respectively, as a function of gate voltage V_G ($\propto$ nanotube charge Q_{tube}), where threshold voltage V_T is chosen at the origin. Nanotube states are indicated by $n\pm$ and the inset is experimental drain current $I_D (\propto g_d)$ vs V_G at fixed drain voltage V_D of Ref. 2.

tributed to this saturation,[2] but we explore here another mechanism inherent to the nanotube channel.

Zigzag nanotubes, denoted by $(N,0)$, have valence bands[13] given by $E(k,k_n) = -|V_{pp\pi}|[1\pm 4\cos(\sqrt{3}ka/2)\times\cos(k_n a/2) + 4\cos^2(k_n a/2)]^{1/2}$, where $|V_{pp\pi}|$ is an overlap integral (3.03 eV) and a is a lattice period. k and $k_n = 2\pi n/aN$ are momenta along and around the nanotube. Each band is denoted by $n\pm$. When N is not divisible by three, the nanotube is semiconducting. In nanotube FETs reported in Refs. 1 and 2, it is not known which nanotubes were used. We will assume a (10, 0) nanotube with a band gap of 1.06 eV, but the qualitative conclusions remain the same for all other semiconducting nanotubes.

V_G, E_F, and charge Q_{tube} on the nanotube channel are all related by

$$\int_0^{E_F} eD(E)dE = Q_{\text{tube}} = C_G(V_G - V_T),\qquad(1)$$

where $D(E)$ is a nanotube state density determined by $E(k,k_n)$ earlier, and the FET geometry defines the gate capacitance C_G. $V_G > V_T$ will demand certain Q_{tube}, which is consistent with C_G. E_F must self-adjust to support this Q_{tube}. This defines $E_F(V_G)$ as shown in Fig. 2(a). Each time a new hole state starts or stops crossing E_F with decreasing V_G, a notch is formed in $E_F(V_G)$. Then, g_d changes abruptly by σ_Q according to the Landauer–Büttiker formula,[14] assuming no scattering, no R_C, and zero temperature. This is shown in Fig. 2(b). There are 20 modes in the several-electron-volt neighborhood of E_F in the (10, 0) nanotube, but some modes do not overlap at all. $5\pm$ states are dispersionless and do not contribute to the transport. For these reasons, g_d is at most $9\sigma_Q$. In practice, $I_D(V_G)$ will be rounded due to a finite temperature and may look somewhat like a broken line in the plot. This already catches the main feature of $I_D(V_G)$ saturation in the IBM data (inset).[2] Furthermore, $I_D(V_G)$ is shrunk vertically because it is multiplied by $T<1$ due to elastic scattering, and is stretched horizontally in some nonlinear manner due to R_C that generally depends on V_G (metal–semiconductor barrier width modulation). Saturation in $I_D(V_G)$ is thus explained. Taken Fig. 2(b) literally, there might be decrease in g_d for extremely large V_G, but this is unlikely to be seen experimentally. Before

1(g) and a staircase for $I_D(V_G)$ at fixed V_D as in Fig. 1(h) as discussed later.

We claim that both Delft and IBM devices[1,2] had penetrating contact. Then, $I_D(V_D)$ will be linear with a slope given by $1/(R_{\text{tube}} + 2R_c)$. R_{tube} is a mesoscopic resistance[8] for the nanotube channel and reflects mesoscopic mode selection at the source as well as elastic scattering characterized by a total transmission coefficient T. R_C is a contact resistance[12] across the metal–semiconductor source/drain contact. V_G can change R_{tube} through the Fermi energy E_F modulation, and thus changes the slope. The linear response prevails until nonlinear effects such as electron-hole pair creation occurs.

$I_D(V_G)$ will be a staircase function. Every time a new state joins the transport with increasing V_G, there is a sudden jump in I_D. The corresponding g_d, however, will be smaller than integer multiples of σ_Q, due to $T<1$ as well as $R_c \sim R_{\text{tube}}$ in practice. These effects, together with finite-temperature rounding, will smooth out the steps in $I_D(V_G)$. In fact, the observed $I_D(V_G)$ in the IBM device was smooth and tended to saturate.[2] $R_c > R_{\text{tube}}$ at large V_G certainly con-
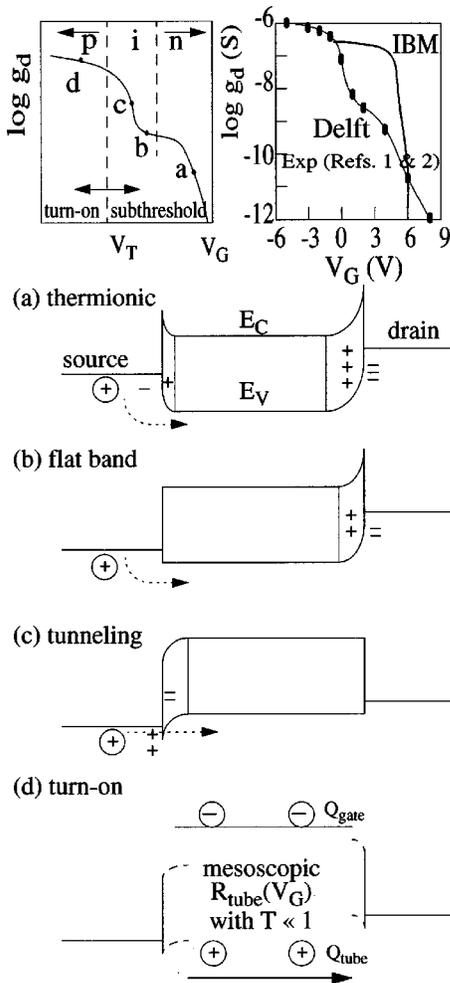
FIG. 3. Channel conductance g_d vs gate voltage V_G in subthreshold and turn-on regions, with experimental data of Refs. 1 and 2. Band diagrams are drawn for selected operating points: (a) thermionic; (b) flat band; (c) tunneling; and (d) turn-on.

pends on V_G, somewhat more slowly than linearly, as discussed in Fig. 2. The charges Q_{tube} are neutralized by the gate charges Q_{gate}. In a ballistic channel, a distinction has to be made as usual for the electrostatic potential, and the left- and right-going chemical potentials. In our weak-localization channel, the electrostatic potential may still not be classical in the absence of carrier–carrier scattering, and thus the band diagram is not explicitly drawn, although the transport is certainly in the linear $I_D - V_D$ regime. Now we claim that Delft group observed an entire transition from (a) to (d), while IBM group observed a latter transition from (c) to (d), and this would explain the presence and absence of the kink in the respective data. The initial doping in each nanotube is very likely different since even one impurity atom out of a thousand carbon atoms would be enough to cause this effect. Another probability is that different metal electrodes of Pt and Au can create different valence band offsets to cause the effect.

In summary, submicron nanotube FETs[1,2] have been analyzed theoretically. The structures in $g_d(V_G)$ reflect the transport at the metal–semiconductor contact, and saturation-less $I_D(V_D)$ indicates the absence of carrier–carrier scattering in the nanotube channel, which may not be compatible with digital applications. In order to compensate for it, we need to design a nanotube FET to maximize the transconductance, so that we can inherit the same circuit scheme as that for conventional FETs.

this occurs, there would be large tunneling current between the gate and the channel, and eventually the transistor would break down.

$g_d(V_G)$ characteristics in the subthreshold and turn-on regions are shown, along with Delft and IBM experimental data[1,2] in Fig. 3. This can be explained based on the transport across the metal–semiconductor contact. In the $g_d(V_G)$ characteristics for a p-channel nanotube FET in Fig. 3, four operating points are shown with a band diagram, where circled charges are mobile. I_D is negligible for (a)–(c) in the subthreshold region, and thus, the nanotube bands are flat except near the contacts. Finite I_D starts to flow for (d) in the turn-on region. In Fig. 3(a), the channel is slightly, unintentionally n-type doped, and the hole conduction is thermionic, where I_D depends exponentially on V_G. In Fig. 3(b), a flat band condition at the source contact is realized, where I_D stays constant with V_G. Figure 3(c) is an onset of tunneling, where I_D again depends exponentially on V_G. In Fig. 3(d), inversion charges are formed and I_D now monotonically de-

[1] S. J. Tans, A. R. M. Verschueren, and C. Dekker, Nature (London) **393**, 49 (1998).
[2] R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and Ph. Avouris, Appl. Phys. Lett. **73**, 2447 (1998).
[3] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley, New York, 1981).
[4] R. W. Keyes, Science **230**, 138 (1985).
[5] W. Shockley, Proc. IRE **40**, 1365 (1952).
[6] V. G. K. Reddi and C. T. Sah, IEEE Trans. Electron Devices **12**, 139 (1965).
[7] S. Frank, P. Poncharal, Z. L. Wang, and W. A. de Heer, Science **280**, 1744 (1998).
[8] C. W. J. Beenakker and H. van Houten, in *Solid State Physics*, edited by H. Ehrenreich and D. Turnbull (Academic, San Diego, 1991); H. R. Sea, R. Martel, and Ph. Avouris, the 46th American Vacuum Society Meeting, Seattle, WA, 25–29 October, 1999.
[9] D. A. Averin and K. K. Likharev, *Mesoscopic Phenomena in Solids*, edited by B. L. Altshuler, P. A. Lee, and R. A. Webb (Elsevier, Amsterdam, 1991).
[10] C. W. J. Beenakker, Phys. Rev. B **44**, 1646 (1991).
[11] S. J. Tans, M. H. Devoret, H. Dai, A. Thess, R. E. Smalley, L. J. Geerligs, and C. Dekker, Nature (London) **386**, 474 (1997).
[12] R_C is discussed for metallic nanotubes in J. Tersoff, Appl. Phys. Lett. **74**, 2122 (1999).
[13] R. Saito, M. Fujita, G. Dresselhouse, and M. S. Dresselhouse, Phys. Rev. B **46**, 1804 (1992).
[14] Y. Imry, *Introduction to Mesoscopic Physics* (Oxford, New York, 1997).