Nano Research
2021, 14(9): 3051–3069

ISSN 1998-0124 CN 11-5974/O4
https://doi.org/10.1007/s12274-021-3459-z

Review Article

Carbon nanotube transistor technology for More-Moore scaling

Qing Cao (✉)

Department of Materials Science and Engineering, Department of Electrical and Computer Engineering, Department of Chemistry
Seitz Materials Research Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61820, USA

Received: 14 January 2021 / **Revised:** 2 March 2021 / **Accepted:** 21 March 2021

ABSTRACT

Scaling of silicon field-effect transistors has fueled the exponential development of microelectronics in the past 60 years, but is now close to its physical limits with the critical dimensions of state-of-the-art silicon devices approaching the sub-10 nm regime. Carbon nanotubes have been suggested to hold great promise of replacing the central role of silicon in the next-generation logic switches with their unique geometrical and electrical properties. In this article, I firstly examine the scaling advantages of carbon nanotubes compared to silicon from technology-development perspective, and then review the latest progress on addressing the manufacturability issues for scaled carbon-nanotube transistors, from materials to device-integration levels. Finally, the possible pathways for nanotube transistors to transition into commercial applications are discussed.

KEYWORDS

carbon nanotube, transistor, scaling

1 Introduction

Since 1960s, Moore's law, which predicts that the number of transistors on general-purpose microprocessors doubles approximately every two years, has powered the development of computer and communication technologies that have transformed all aspects of our modern life. The exponential reduction of transistor size, as dictated by Moore's law, constantly makes our computer chips and information-processing machines cheaper, faster, and more energy efficient. The progress so far has been astonishing: the central-processing units (CPUs) of smartphones we are using today have over 100,000 times the processing power of the Apollo computer that landed man on the moon 50 years ago, together with more than 10 times reduction in power consumption. The most recent progress under Moore's law laid the foundation of today's breakthroughs in artificial intelligence and blockchains, by giving computers the capability to practically process massive amount of data with reasonable throughput and cost.

Despite its tremendous success, Moore's law based on the scaling of silicon metal-oxide-semiconductor field-effect transistors (MOSFETs) is approaching its physical limits as evident from the slowing-down progress by the semiconductor industry in the past decade. First, although the number of transistors in our microprocessors still managed to stay on the exponential growth curve, the clock frequency and the chip single-thread performance have both stagnated since early 2000s (Fig. 1(a)). Second, the individual-transistor dimensions have been shrinking at an increasingly slower rate, with only about 10% reduction in the contacted-gate pitch (the smallest possible distance between gates of adjacent transistors on chip) going from 10 to 5 nm technology node (Fig. 1(b)). However, in the meantime, the demand on the continuous progress in computational power is surging with emerging applications.

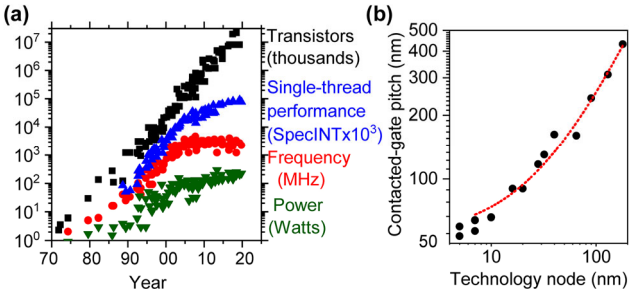


Figure 1 (a) Trend of different aspects of microprocessors including the transistor count (black), single-thread performance (blue), clock frequency (red), and total power consumption (green) over the last 48 years (source: Karl Rupp, 48 years of microprocessor trend data). (b) Scaling trend of the contacted-gate pitch of logic transistors from 180 to 5 nm technology node (source: Wikichip).

For example, further development of artificial intelligence requires more powerful and energy-efficient computers to train and execute more and more complex deep-neural-network models, which are composed of millions of nodes and billions of parameters, for more accurate and cognitive learning and inference tasks both on cloud and edge [1]. The business applications of blockchains require exponentially higher hash rate of the hardware to accommodate the skyrocketing difficulty associated with the expansion of the network size [2]. The virtual-reality and augmented-reality applications require higher computing power under much lower power consumption on mobile devices to improve the portability and user experience. Although improvements on the software and architecture levels can meet some of these challenges [3], transistor technologies enabling smaller footprint and better performance on device level (a strategy called more-Moore) is still more desirable.

When scaling becomes more and more difficult, materials

Address correspondence to qingcao2@illinois.edu

innovations are playing an increasingly important role to sustain Moore's law. We have already witnessed the replacement of aluminum with copper interconnect at 180 nm technology node, the replacement of SiO_2 with organosilicate glass as interlayer dielectrics at 90 nm node, and most recently the replacement of SiO_2 gate dielectric/polysilicon gate with HfSiON-La_2O_3 high-k gate dielectrics/metal gate at 45 nm node. The silicon channel becomes the next target. Among many potential material choices, including III-V semiconductors and various two-dimensional (2D) nanomaterials [4, 5], single-walled carbon nanotubes (SWNTs) are an especially attractive candidate with their unique combination of intrinsic nanometer size and exceptional electrical properties. In this review, I outline the case for replacing silicon with carbon nanotubes in the next-generation extremely scaled logic transistors from technology-development perspective, summarize recent progress made by researchers from both academia and industry to transform nanotube transistors into a practical technology, and discuss the most critical problems that remain to be overcome, followed by my perspective on the potential pathway for nanotube transistors to become a mainstream device technology.

2 Scaling advantages of carbon nanotubes: What really matters?

The structure of a SWNT can be visualized as a rolled-up single-layer graphene with about one-nanometer diameter, as first observed by high-resolution transmission-electron microscopy (TEM) in 1993 [6, 7]. The exceptional electrical properties of SWNTs were soon predicted in theory and verified in experiments [8–10]. The carrier mobility of carbon nanotubes is more than 30 times higher than that of silicon and allows ballistic carrier transport at room temperature over several microns [11], which led to the initial enthusiasm to apply nanotube transistors as the next switch beyond silicon MOSFETs in 2000s [12, 13]. However, impressive as these attributes are, they do not address what really matters for a scaled logic transistor. In this section, I will first discuss what the limiting factors for silicon transistors are and why the unique properties of nanotubes are suitable to overcome these hurdles, based on recent simulation and experimental results.

2.1 Chip power and power density

Under Moore's law, with more and more transistors squeezed into the same chip area, the power consumption of each individual device has to be dramatically reduced to manage the chip power density. It used to be accomplished by following the Dennard scaling law. According to Dennard's law, the device physical dimensions (both width W and length L) and its operating voltages, including both the drive voltage (V_{DD}) and the threshold voltage (V_T), should be reduced in tandem by the same factor α. As a result, the device on-state current (I_{on}) density remains the same, while the device power consumption ($W{\cdot}I_{on}{\cdot}V_{DD}$) and the overall area ($W{\cdot}L$) are both reduced by the factor of α^2, keeping the chip on-state power density (power consumption per area) as a constant [14]. However, this strategy has already reached its limit because MOSFETs, unlike a perfect switch, have a fundamental limitation on the abruptness of their turn-on characteristics, where the applied gate voltage (V_{GS}) needs to be changed by at least 60 mV to cause a corresponding change of the source-drain current (I_{DS}) by one decade at room temperature, as limited by thermionic emission. Therefore, the reduction of V_T causes exponentially higher device off-state leakage current (I_{off}), as schematically illustrated in Fig. 2(a). As a result, the chip passive power,

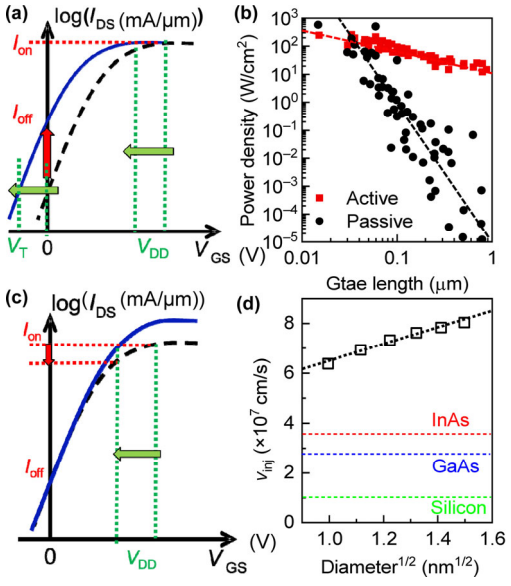


Figure 2 (a) Schematic illustrating the impact of voltage scaling on I_{off} and I_{on}. Transfer curves of MOSFETs with high (black dashed line) and low (solid blue line) V_T are plotted in semi-logarithmic scale. Green dotted lines serve as visual guide to mark the commensurate shift of V_{DD} and V_T. Red dotted line marks the invariance of I_{on} density. (b) Scaling trend of the passive (black dots) and active (red squares) power densities of silicon logic transistors as a function of the device gate length [15]. (c) Schematic illustrating the maintenance of the device I_{on} and I_{off} densities under the same V_T, lower V_{DD}, and higher v_{inj} of the channel. Transfer curves of MOSFETs with high (solid blue line) and low (black dashed line) v_{inj} are plotted in semi-logarithmic scale. Green dotted lines serve as visual guide to mark the shift of V_{DD}. Red dotted lines mark the reduction of I_{on} density of a device with identical V_T and v_{inj} but lower V_{DD}. (d) Theoretical v_{inj} of carbon nanotubes as a function of the square-root of nanotube diameter (black), as benchmarked with that in silicon (green), GaAs (blue), and InAs (red) transistors [4, 19].

the energy consumed by microprocessors even while doing nothing with transistors biased steadily into the off state with the applied V_{GS} = 0 V, has been growing rapidly with the reduction of device size, and has already approached the active dynamic power in state-of-the-art silicon-transistor technologies (Fig. 2(b)). Continuing with this trend will soon increases the overall chip power consumption beyond the cooling capacity we can practically provide at chip or package level [15].

If we choose to keep the V_T constant but only reduce V_{DD} in scaling, we can limit both the standby and dynamic power consumptions of silicon MOSFETs, but will sacrifice the I_{on} density which is proportional to the device switching speed. For nanoscale logic transistors, I_{on} density can be calculated as the product of the carrier concentration (Q_{inv}) and the carrier velocity in the channel at the top of the barrier near the source (so-called injection velocity v_{inj}). The carrier concentration Q_{inv} is further proportional to the gate-oxide capacitance per unit area (C_{ox}) and the difference between V_{DD} and V_T as $Q_{inv} = C_{ox}{\cdot}(V_{DD} - V_T)$ [16, 17]. Therefore, maintaining the V_T while decreasing V_{DD} reduces their difference ($V_{DD} - V_T$), and thus the concentration of free carrier. If the v_{inj} of the semiconductor channel remains the same, the reduced carrier concentration will then inevitably lead to lower I_{on} density. So in this case, our computers will run slower than the previous model, which is also unacceptable [18].

Therefore, to overcome such power-dissipation bottleneck, which has become one of the most critical show stoppers for more-Moore scaling of conventional silicon MOSFETs, an effective approaches is to increase the v_{inj} of the semiconductor channel, which will counterbalance the reduction of Q_{inv} caused by the decrease of (V_{DD} - V_T) so that we can not only constrain

the power consumption but also maintain the I_{on} density for better performance, as illustrated in Fig. 2(c). It can be successfully accomplished by replacing the silicon channel with carbon nanotubes, since the unique band structure of nanotubes and their low carrier effective mass lead to several times higher theoretical v_{inj}, as benchmarked against in devices based on either silicon or III-V semiconductors (Fig. 2(d)) [19]. In addition to theoretical predictions, v_{inj} was recently extracted in experiment from nanotube transistors with 10–15 nm gate length (L_g) to be in the range of 3×10^7 – 4×10^7 cm/s [19, 20], which is more than triple that in silicon MOSFETs and compares favorably against InGaAs and InAs high-electron-mobility transistors (HEMTs). These results are critical because they suggest that, with carbon nanotubes as the semiconductor channel of MOSFETs, we can maintain the current V_T at 0.3 V while reducing the V_{DD} drastically from 0.9 to 0.5 V to significantly decrease the device power consumption but still delivering the same or even better performance compared to their silicon-based competitors; and this unique capability of carbon nanotube transistor resulting from SWNT's intrinsically higher v_{inj} is essential to support further scaling of logic transistors toward higher integration density.

2.2 Short-channel effects

Another critical limiting factor for MOSFET scaling is the so-called short-channel effects. For transistors, the abruptness of the subthreshold transition, which is quantitatively characterized by a parameter called the subthreshold swing (SS) as defined by the relationship: SS = $dV_{GS}/d\log_{10}I_{DS}$, is limited by not only the thermionic emission over the barrier in the off-state but also the effectiveness of the electrostatic coupling between the gate and the channel. More specifically, because the applied source–drain bias creates depletion regions penetrating the device channel underneath the gate stack, the gate electrode partially loses the electrostatic control over the channel. As a result, a larger gate bias will be required to shift the channel potential and the resultant carrier concentration by the same amount compared to an ideal MOSFET. Such undermined electrostatic coupling therefore leads to the deviation of the device SS from the 60 mV/dec thermodynamic limit. In addition, the device V_T becomes dependent on drain bias applied, exhibiting so-called drain-induced-barrier-lowering (DIBL) effect. Since the relative portion of these depletion regions extending from the source and drain into the channel increases with the reduction of the device L_g, their impact is more severe in scaled transistors. In nanometer-size devices, these short-channel effects can lead to substantially larger SS, and thus drastically increased I_{off} or device passive power consumption, compared to long-channel transistors in older technology nodes (Fig. 3(a)), and severe DIBL contributing to the device V_T variation.

One way to mitigate the short-channel effects is to reduce the silicon-channel thickness (t_{Si}) together with the L_g. If the t_{Si} is considerably smaller than the source/drain junction depth, the extension of the formed depletion regions around source and drain electrodes into the device channel will be greatly suppressed [21]. Moreover, the shape of the channel can be modified so that the gate control can be exerted from multiple directions in a three-dimensional (3D) multigate configuration to better compete with the variation in the electric field along the channel direction arising from the source-drain bias [22]. Quantitatively, a parameter called the natural length (λ) is used to represent the extension of the electric field lines from the source and drain into the channel as $\lambda = \sqrt{\frac{\varepsilon_{Si}}{N\varepsilon_{ox}} t_{ox} t_{Si}}$,

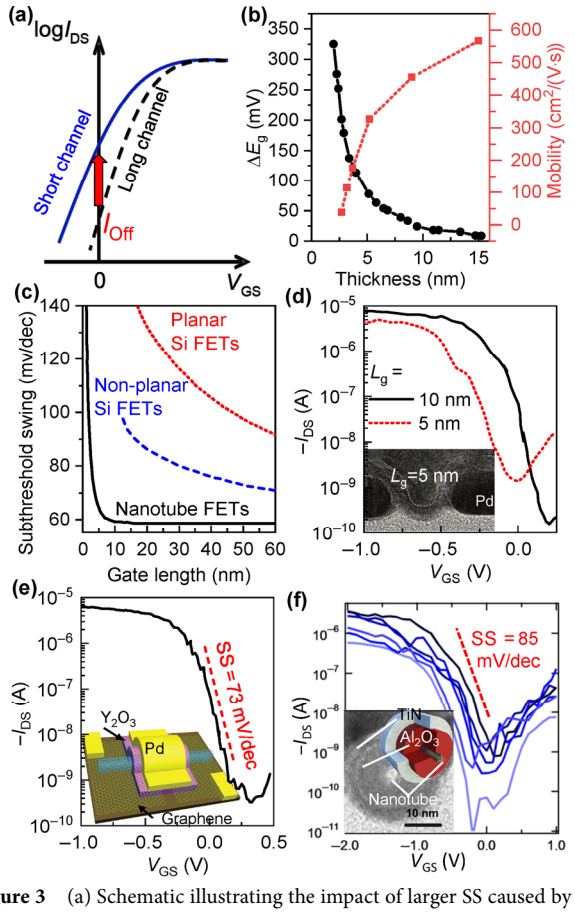


Figure 3 (a) Schematic illustrating the impact of larger SS caused by the short-channel effects on device I_{off}. (b) The change of the bandgap (ΔE_g, black, left axis) and mobility (red, square, right axis) as a function of the silicon-body thickness [23, 26]. (c) Scaling characteristics of the SS for nanotube FETs (black solid line), silicon planar-gate FETs (red dotted line), and silicon multigate FETs (blue dashed line) [19, 30]. (d) Logarithmic plots of experimentally measured transfer curves for nanotube transistors with L_g of 10 nm (black solid line) and 5 nm (red dotted line), respectively. Applied V_{DS} is –0.1 V. Inset: cross-sectional TEM micrograph showing the 5 nm-L_g nanotube FET with Pd contacts. (e) Transfer characteristics and the schematic diagram (inset) of the 5 nm-L_g nanotube FET with graphene contacts. Applied V_{DS} is –0.1 V. Reproduced with permission from Ref. [32], © American Association for the Advancement of Science 2017. (f) Transfer curves of gate-all-around nanotube transistors. Applied V_{DS} is –0.5 V. Inset: cross-sectional schematic and TEM image showing the device gate stack. Dashed red line serves as a visual guide to extract the SS. Reproduced with permission from Ref. [36], © American Chemical Society 2013.

where ε_{ox} and ε_{Si} are the dielectric constants of the gate oxide and the silicon channel, respectively, t_{ox} is the gate-oxide thickness, and N is the effective gate number correlating with the gate configuration (1 for planar gate, 2 for double gate, 3 for trigate, and 4 for gate-all-around). Generally, the L_g needs to be at least six times longer than λ to ensure that the short-channel effects will not significantly degrade the device operations. Therefore, in order to suppress the short-channel effects, the t_{Si}, i.e. the fin width in state-of-the-art silicon fin field-effect transistors (finFETs), at 7/5 nm technology nodes has been limited to merely 5–7 nm for L_g down to 16 nm. Further reducing the L_g below 10 nm requires the t_{Si} to become less than 5 nm, which creates two fundamental problems. The first is the quantum confinement, which splits both the conduction and valence bands of silicon into sub-bands and thus widens the effective bandgap E_g. And this effect becomes very pronounced when t_{Si} is below 5 nm (Fig. 3(b)) [23, 24]. Since V_T is closely correlated with E_g, a small fluctuation of the device fin width in this regime caused by the process variability, for example

the line-edge roughness in the lithography process [25], will create substantial V_T variations, which is a big problem for integrated circuits where we want billions of transistors to turn on and off at the same voltage. The second problem is that the greater confinement of electrons within thinner silicon films leads to enhanced phonon scattering and surface roughness scattering. As a result, the effective carrier mobility drops substantially in such ultrathin silicon films with $t_{Si} < 5$ nm (Fig. 3(b)) [26–28], which will degrade the device performance by reducing the v_{inj}. Therefore, although 10 nm-L_g silicon transistors with $t_{Si} \leq 5$ nm have been demonstrated in experiment showing adequate short-channel control [29], it is not clear that they can deliver the required performance and uniformity for commercial applications.

Carbon nanotubes, as a quasi-one-dimensional (1D) nanomaterial with diameters, i.e. their intrinsic semiconductor-body thickness, down to around one-nanometer range, offer the possibility to achieve superior electrostatic control by the gate electric field even for devices with extremely scaled L_g, without sacrificing their electrical properties. Simulation indicates that the SS of nanotube transistors remains at the 60 mV/dec limit and the DIBL is kept below 50 mV/V, with gate-all-around device structure and 2 nm t_{ox}, down to L_g below 5 nm, which are significantly better compared to either planar or nonplanar (trigate finFET) silicon devices (Fig. 3(c)) [19, 30]. In addition to simulation, nanotube transistors' immunity to the short-channel effects has been extensively verified in recent experiments [31–33]. At 10 nm L_g, the nanotube transistors with conventional metal contacts and planar gate exhibited SS down to 70 mV/dec [31, 32], which is much lower than the 90–120 mV/dec SS of highest-performing 10 nm-L_g silicon counterparts (Fig. 3(d)) [29, 34, 35]. Although the SS for nanotube transistors degraded to ~ 110 mV/dec in experiment when the L_g was further reduced to 5 nm [32], it can be potentially improved by three strategies. The first is to reduce the thickness of the source/drain metal contacts. Thinner contacts will minimize their screening against the gate electric field, and thus allow the segments of nanotubes beneath the contact being depleted by the gate electric field to increase the effective L_g in the device off-state. For example, by adopting single-layer graphene as the contact, the SS down to 73 mV/dec has been demonstrated for 5 nm-L_g nanotube transistors (Fig. 3(e)) [32]. However, to make this strategy most effective, there must be substantial overlaps between the gate and the source-drain contacts, which will cause large parasitic capacitance to degrade the device switching speed. Another option is to adopt the gate-all-around structure, where the gate completely surrounds the nanotube channel from all directions for maximum electrostatic control, as in the theoretical simulation studies [19]. Such gate-all-around nanotube transistors have been demonstrated in experiment with the high-k gate dielectric deposited by atomic-layer deposition (ALD) on a suspended nanotube channel (Fig. 3(f)) [36]. Since the nanotube surface is inert to ALD precursors, a monolayer of NO_2 was physically adsorbed on the nanotube surface first, which was subsequently converted to an AlO_xN_y adhesion layer upon reaction with the trimethylaluminum precursors [37]. ALD TaN was then deposited as the gate metal. However, it has not been proven in experiment whether the improved electrostatics of the gate-all-around device architecture would indeed lead to better SS scaling behavior for nanotube devices, since the smallest L_g experimentally demonstrated for these gate-all-around nanotube transistors was above 20 nm, as limited by the added device structural complexity. And even for these relatively long-channel transistors, the measured SS was over 85 mV/dec, far from the 60 mV/dec limit as largely hindered by the presence of traps at the nanotube-AlO_xN_y interface. Further engineering optimizations of the direct ALD growth of high-quality gate dielectrics on nanotubes might help to address this problem in the future, and reveal the intrinsic L_g-scaling limit of nanotube transistors with wrapped gate [38, 39]. The last option is to adopt nanotubes with smaller diameters, which can suppress both the direct source-drain tunneling and the band-to-band tunneling across the nanotube channel with their larger E_g and lower tunneling probability according to simulation [40]. However, smaller nanotube diameters will also lead to smaller I_{on} with the lower v_{inj} (Fig. 2(d)) and the larger parasitic resistance at the nanotube-metal contacts. It is likely that an optimized diameter needs to be selected in future nanotube transistor technologies based on whether the devices are tuned toward high-performance, where high I_{on} is more important, or low-power applications, where small SS to reduce I_{off} and V_{DD} is more critical. There is no experimental data on the DIBL reduction effect of short-channel nanotube transistors, as it is hard to be extract accurately due to the presence of hysteresis in nanotube transfer characteristics [41].

2.3 Scaling of contact length

Transistor scaling requires the proportional miniaturization of both the semiconductor channel and the metallic source-drain contacts. Reduction of the device channel length decreases the channel resistance up to its ballistic limit. However, the contact resistance, on the contrary, increases rapidly with the reduction of the device contact length (L_c, the length of the metal contact typically in overlap with the underneath semiconductor). As shown in Fig. 4(a), for silicon, with the standard nickel-silicide contacts (Pt is incorporated to improve the thermal stability), the extracted overall contact resistance ($2R_c$) increases rapidly from below 100 Ω·μm to above 500 Ω·μm when the L_c is reduced to 25 nm [42]. And the situation is even worse for III-V semiconductors [43]. Such scaling effect of the L_c on $2R_c$ can be adequately described by the classical transmission-line model, where the resistance of the semiconductor corresponds to the series resistance of the transmission line, the interface resistance represents the parallel shunt-line resistance, and the contact metal, whose resistance is neglected here, forms the return lead (Fig. 4(b)). By solving the transmission-line equations as a superposition of forward and reflected-back waves, the contact resistance, as the input resistance across the two ports of the transmission line, can be calculated as $R_c = \frac{\sqrt{R_s \rho_c}}{W} \coth\left(\sqrt{\frac{R_s}{\rho_c}} L_c\right)$, where R_s is the sheet resistance of the semiconductor, ρ_c is the contact interface resistivity, and W is the contact width [44, 45]. This universal correlation between L_c, which occupies a substantial portion of the overall device footprint, and $2R_c$ has become a significant hurdle for scaling, as it indicates that the reduction of L_c into the 10-nanometer scale inevitably leads to drastically higher parasitic resistance and therefore a commensurate drop in device performance.

Carbon nanotubes, again with their intrinsic quasi-one-dimensional shape, offer a potential solution to this problem. Instead of depositing metals on top of the nanotubes to form the similar overlapping contact structure as in silicon or III-V transistors where the metal–nanotube interface are bonded through relatively weak van der Waals interactions (Fig. 4(c)), the quasi-zero-dimensional open ends of the nanotubes can be directly welded to the metal contacts where the carbon atoms of the nanotubes bond directly to the metal atoms of the contacts through strong covalent carbide bonds (Fig. 4(d)) [46]. In this so-called "end-bonded" contact scheme, most carriers from the nanotube channel are collected into the metal contacts with high transmission probability through this strongly coupled

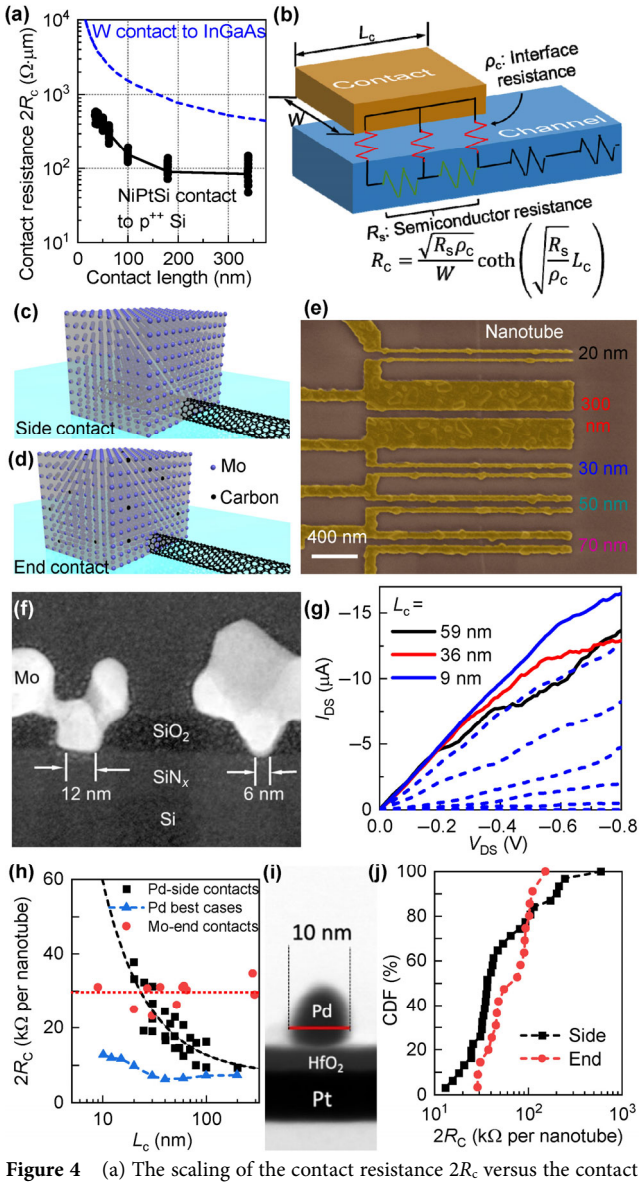


Figure 4 (a) The scaling of the contact resistance $2R_c$ versus the contact length for 3 nm-thick NiPtSi contact to heavily boron-doped silicon (black solid) and tungsten contact to InGaAs (blue dashed) [42, 43]. (b) Schematic of the transmission-line model for distributed planar metal–semiconductor contacts. (c) and (d) Schematics of the side-bonded (c) and end-bonded (d) contacts to carbon nanotubes. (e) False-colored SEM image showing the test structure composed of a set of quasi-ballistic transistors with varying L_c fabricated on the same nanotube to extract the L_c-scaling characteristics. (f) Cross-sectional TEM showing the sub-10 nm length end-bonded contacts to nanotubes. (g) Current–voltage characteristics of end-contacted nanotube transistors with their L_c varied from 59 nm (black), 36 nm (red), to sub-9 nm (blue). Reproduced with permission from Ref. [46], © American Association for the Advancement of Science 2015. (h) Plots of $2R_c$ as a function of L_c extracted from sets of nanotube transistors adopting either Pd side-bonded contacts (black) or Mo end-boned contacts (blue). Each set of devices with varied L_c was fabricated on the same nanotube. The lowest $2R_c$ for each L_c extracted from a large group of nanotube transistors is plotted together to illustrate the best-case L_c-scaling characteristics for Pd side contacts (blue triangles). (i) Cross-sectional TEM image showing the 10 nm-length Pd side-bonded contact to nanotubes. Reproduced with permission from Ref. [50], © American Chemical Society 2019. (j) Comparison of the cumulative distribution function (CDF) plots of $2R_c$ for 10 nm-length Pd side-bonded contacts (black squares) and Mo end-bonded contacts (red circles) to semiconducting nanotubes.

quasi-zero-dimensional interface without changing their transport directions, and therefore the contact resistance is both low, because of high transmission probability resulting from strong carbon-metal coupling, and independent of the contact length, because of the quasi-zero-dimensional nanotube-metal interface. In experiment, the end-bonded contacts can be formed by reacting carbide-forming metals deposited on top of the nanotube, such as molybdenum and nickel, with the underneath carbon nanotube through solid-state chemical reactions upon annealing [46, 47]. Their L_c-scaling behavior was evaluated in experiment by constructing a series of devices with identical ~ 20 nm L_g but various L_c on the same nanotube to avoid the dependence of $2R_c$ on the semiconducting nanotube diameter and bandgap (Fig. 4(e)). Since the ~ 20 nm device L_g is much smaller than the carrier mean free path of nanotubes, the channel is quasi-ballistic. Therefore, the measured overall nanotube-transistor resistance can be approximated as the nanotube contact resistance $2R_c$ plus the ballistic resistance of the nanotube channel (~ 6.5 kΩ per tube). Based on these test structures, we found that, with end-bonded contacts, even under smallest L_c below 10 nm (Fig. 4(f)), the output resistances of these quasi-ballistic nanotube transistors were indeed independent of L_c with uncompromised device performance (Fig. 4(g)), confirming a constant $2R_c$ regardless the contact length. Such immunity to the L_c-scaling effects, enabled by the unique end-bonded contact scheme, represents another critical scaling advantage of carbon nanotubes as a low-dimensional nanomaterial compared to bulk silicon.

Despite their better scalability, for technology development, it is necessary to quantitatively compare the performance of such end-bonded contacts with the conventional sided-bonded contacts to semiconducting nanotubes. For side-bonded Pd contacts to nanotubes, the L_c scaling experiments performed using the same test structures, where multiple quasi-ballistic transistors with varied L_c were constructed on the same nanotube, suggested a scaling trend similar to that of silicon devices. The extracted $2R_c$ could be pretty low down to ~ 6 kΩ per tube at long contacts, but it rises rapidly with the reduction of the device L_c (Fig. 4(h)) [32, 48, 49]. In contrary, the end-bonded contacts can achieve the L_c-independent $2R_c$ around 36 kΩ per tube [46]. Although quantitatively they are inferior to side-bonded contacts at long L_c, end-bonded contacts can provide at least two to three times lower $2R_c$ when the L_c gets into the sub-10 nm regime. However, it is still premature to decide that end-bonded contacts would be preferred for scaled nanotube transistors. In a recent survey, the parasitic contact resistances were extracted from a large number of transistors fabricated on nanotubes with varied diameters for side-bonded Pd contacts [50]. It found that the $2R_c$ for Pd side contacts could exhibit a much weaker dependence on L_c in the best-case scenario, with merely 13 kΩ per tube for 10 nm L_c on some "hero" devices (Fig. 4(i)). These results give hope to rely on the low-temperature-deposited side contacts to achieve even smaller parasitic resistance in scaled nanotube transistors. However, we cannot make any definitive conclusion as the results are obscured by the large variability observed for both contact schemes. Figure 4(j) compares the cumulative $2R_c$ distributions for both 10 nm-L_c side-bonded Pd contacts and end-bonded Mo contacts to semiconducting nanotubes. They all exhibit a large variation with greater than 10× spread. The underlying mechanisms causing this large variability are not clear yet, but the random defects associated with either the nanofabrication process or the nanotube material itself are likely the dominant contributors [50]. Reducing such variability, in my opinion, is currently the most critical challenge to improve contact properties in nanotube transistors.

2.4 Extremely scaled nanotube transistors and their benchmark with silicon-based competitors

To experimentally verify the above-mentioned scaling advantages

of carbon nanotubes, i.e. the high intrinsic v_{inj} as well as the immunities to both short-channel and L_c-scaling effects, it requires the fabrication of nanotube transistors with both sub-10 nm L_g and L_c, as well as the benchmark of their performance against state-of-the-art silicon-based competitors. This target has been accomplished by our team in 2017 [33]. As shown in the schematic (Fig. 5(a)) and the cross-sectional TEM micrograph (Fig. 5(b)), the device incorporated a semiconducting nanotube as the channel with 10-nm wide end-bonded source-drain contacts composed of Co-Mo alloy. Cobalt acted as a catalyst to drastically reduce the carbide-formation temperature from 800 °C as required for pure molybdenum contacts down to 650 °C, which is critical to minimize the distortion of the device dimensions caused by the high-temperature annealing process for these nanoscale transistors. 5 nm Al_2O_3 deposited by ALD was employed as the high-k gate dielectric and spacers separating the 10 nm-wide Pd metal gate with the Co-Mo source-drain contacts. The overall device footprint, including the source-drain contacts, the gate, and the spacers, spanned only 40 nm. It corresponds to a 30 nm contacted-gate pitch in integrated circuits, which is two times smaller, or in another word at least two technology nodes ahead (based on the conventional shrink factor of 0.7 in Moore's law), than state-of-the-art 5 nm-node silicon transistors [51].

Their electrical performance was then firstly benchmarked with some of the most advanced silicon MOSFETs, including high-performance finFETs and gate-all-around nanowire transistors at 10 nm node as well as fully depleted silicon-on-insulator (FDSOI) transistors at 14-nm node (Fig. 5(c)). Even with 2 times smaller L_g, the fabricated extremely scaled nanotube transistors still exhibited similarly abrupt SS, which verifies the suppression of the short-channel effects by the intrinsic ultrathin body of carbon nanotubes. The device on-state current density was compared after normalizing I_{on} of the nanotube transistor to nanotube pitch of 2 and 5 nm, respectively, both of which are achievable with the current nanotube assembly processes to be discussed in Section 3.2. Under a lower V_{DS} of −0.5 V, the nanotube transistors achieved similar (5 nm pitch) or twice (2 nm pitch) as high I_{on} density at −0.5 V V_{GS} with the same I_{off} of 2 μA/μm, which verifies the capability of scaled nanotube transistors to deliver the same or even faster switching speed under lower power consumption as enabled by nanotube's higher v_{inj} and lower parasitic contact resistance. The I_{on} density of such 40 nm-footprint nanotube transistors (assuming 2 nm pitch) was further compared to those of the highest-performing silicon devices with 10 nm L_g under identical I_{off} and 0.5 V V_{DD} (Fig. 5(d)). We found that the nanotube transistors could deliver a pitch-normalized I_{on} density up to 900 μA/μm, while that of the best silicon-based competitors is more than four times lower at only about 200 μA/μm [29, 34, 35]. Finally, to benchmark our extremely scaled nanotube transistors with incumbent silicon transistor technologies on chip level, a device model was established based on these experimental results. For nanotube transistors with both their L_g and L_c scaled to around 10 nm to afford an overall contacted-gate pitch of 30 nm (Fig. 6(a)), which exhibit I_{on} ~ 740 μA/μm, I_{off} ~ 2 μA/μm, and SS ~ 83 mV/dec as comparable to the values determined in experiment (Fig. 5(c)), they would enable more than two-fold faster logic transactions under less than half energy consumption in a microprocessor based on the International Business Machines (IBM) Power 7 architecture, compared to what is possible with 7/5 nm technology-node silicon transistors (Fig. 6(b)) [52, 53]. All these benchmark results have confirmed that carbon nanotubes, when incorporated into technology-relevant, extremely scaled transistor form factor in experiment, will enable the more-Moore scaling of MOSFETs beyond the physical limits of silicon by addressing their most critical scaling challenges with nanotube's intrinsic material properties, which therefore put carbon nanotube transistors at the forefront in the competition for the next switch in high-performance integrated logic circuits.

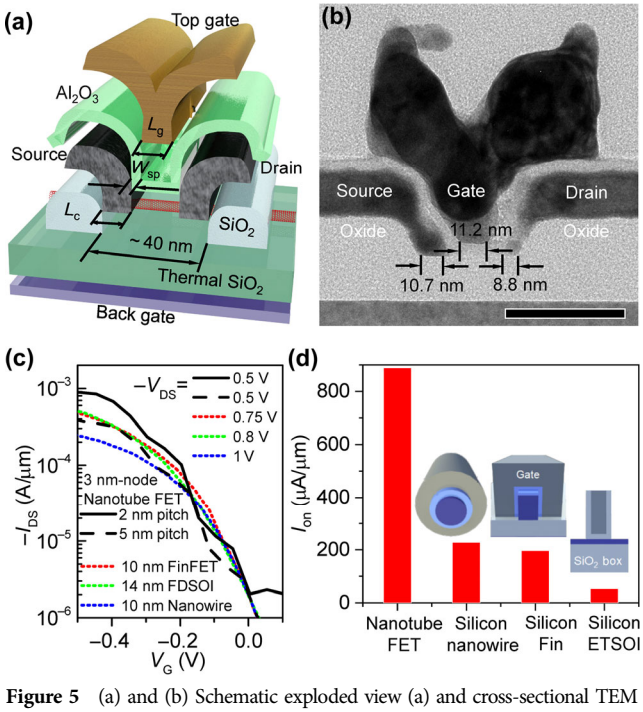


Figure 5 (a) and (b) Schematic exploded view (a) and cross-sectional TEM micrograph (b) showing the structure of the extremely scaled nanotube transistor with 10 nm-L_c end-bonded contacts, 10 nm L_g, and 5 nm spacer width (W_{sp}) to afford overall 40 nm device footprint. Scale bar in (b): 50 nm. (c) Comparing the transfer characteristics of pitch-normalized 40-nm footprint nanotube transistors biased under −0.5 V V_{DS} (black solid line for 2-nm and dashed line for 5-nm nanotube pitch, respectively) against the 10-nm-node Si finFET biased under −0.75 V V_{DS} (red dotted line), the 14-nm-node FDSOI FET biased under −0.8 V V_{DS} (green dotted line), and the 10-nm-node Si nanowire FET with gate-all-around configuration biased under −1 V V_{DS} (blue dotted line). Reproduced with permission from Ref. [33], © Cao, Q. et al. 2017. (d) Benchmarking the pitch-normalized I_{on} of 40 nm-footprint nanotube transistor with that of best-reported sub-10 nm L_g silicon MOSFETs under identical V_{DD} of 0.5 V and Ioff of 2 μA/μm.

3 Manufacturability challenges for carbon nanotube transistors

The next step is to convert what we see in the chip simulation (Fig. 6(b)) into a nanotube CPU as a commercial product. However, despite the great promise of carbon nanotube transistor

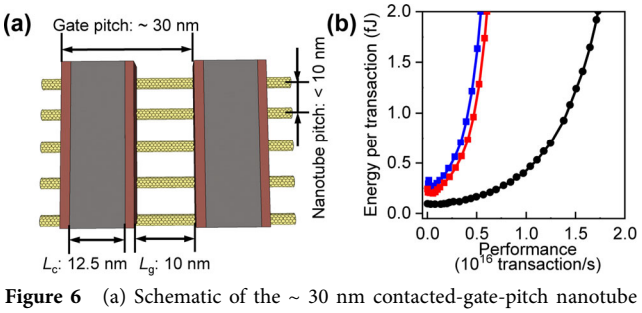


Figure 6 (a) Schematic of the ~ 30 nm contacted-gate-pitch nanotube transistor with 10 nm L_g, 12.5 nm L_c, and 2.5 nm W_{sp} built on closely packed parallel nanotube arrays as the channel. (b) Simulated energy *versus* performance for microprocessors built with the scaled nanotube transistors displayed in frame (a) (black), as well as 5-nm-node (blue) and 7-nm-node (red) silicon MOSFETs.

technology as discussed in the previous section, it is a daunting challenge to manipulate these nanomaterials for the fabrication of nanotube transistors with yield and uniformity high enough toward very-large-scale-integrated circuits composed of tens of billions of devices, as what we can do with silicon. In this section, I will discuss the major manufacturability challenges for nanotube transistors, the latest progress to address these issues, and the remaining obstacles.

3.1 Obtaining purity level high enough for logic device applications

The first requirement is simply all the nanotubes in the fabricated transistors need to be semiconducting. The presence of any metallic nanotubes will short the source-drain electrodes and lead to unacceptably high I_{off}. Semiconducting nanotubes can be selectively prepared by chemical-vapor deposition (CVD) through adjusting the catalysts and the growth conditions [54–56]. However, the achieved purity level by direct synthesis so far has been far away from what is required for logic electronic applications, where the metallic nanotube counts need to below at least 0.1–1 part-per-billion (ppb) to ensure a reasonable circuit yield, even with some levels of redundancies built into the circuit designs [57]. In comparison, the post-synthesis separation, which has made significant progress during the past decade, is more mature and promising toward the purity target. The enrichment of different electronic types of nanotubes can be realized based on three major different mechanisms. (1) Charger transfer between nanotubes and the dispersion surfactants: As-synthesized nanotubes can be non-covalently functionalized by some surfactants and become well-dispersed in aqueous or organic solvents. For certain types of surfactants, charge transfer could happen between the more electron-rich metallic nanotubes and the surfactant molecules. This selective charge-transfer reaction leads to lower linear-charge density and higher packing density, which further translate into higher buoyant density and viscous drag, of the formed metallic nanotube-surfactant complexes compared to their semiconducting counterparts. These differences can be utilized for their separation into different bands after the density-gradient ultracentrifugation (Fig. 7(a)) [58–60] or column chromatography [61]. These processes represent the first successful approach to get bulk quantities of high-purity semiconducting nanotubes, but their major drawback is the relatively low throughput with the involvement of complex processing steps such as multiple extended ultracentrifugations. (2) Molecular recognition between nanotubes and semiflexible conjugated polymers: Some semiflexible conjugated polymers, such as aromatic polyfluorenes [62], polycarbazoles [63], polythiophenes [64], and even different strains of deoxyribonucleic acids (DNAs) [65], can wrap around nanotubes with nanotube discriminations by their diameters, chirality, and electronic types (Fig. 7(b)). Such selective dispersion of semiconducting nanotubes assisted by conjugated polymers has better manufacturing scalability and throughput, as it merely involves sonicating the nanotube powders in a solution of the conjugated polymers followed by a simple low-speed centrifugation step to remove the sediments. However, our understanding about the influence of polymer backbone

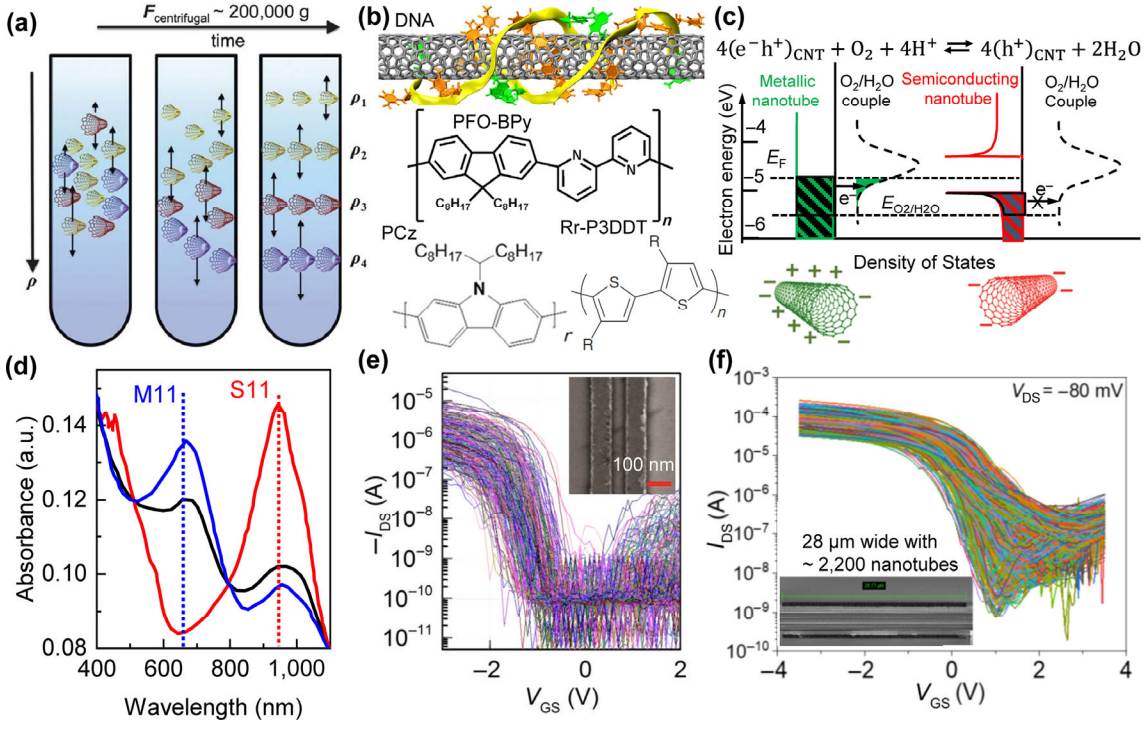


Figure 7 (a) Schematic illustration of the density-gradient-ultracentrifugation method to separate nanotubes based on electronic types (ρ: buoyant density; $F_{centrifugal}$: centrifugal force). Reproduced with permission from Ref. [136], © The Royal Society of Chemistry 2013. (b) Polymers that can selectively wrap around nanotubes with different electronic types and/or chirality, including DNA, 9,9-dioctylfluorenyl-2,7-diyl and bipyridine (PFO-BPy), PCz, and region-regular poly(3-dodecylthiophene) (Rr-P3DDT). Reproduced with permission from Ref. [65], © Macmillan Publishers Limited 2009. (c) Schematic illustration of the differential charge-transfer redox reaction between O_2/H_2O couples and nanotubes with different electronic structures as explained by the Marcus–Gerischer theory. (d) Optical absorption spectra of unsorted (black), semiconducting-enriched (red), and metallic-enriched (black) nanotube aqueous suspensions. Blue and red dashed lines serve as visual guide to mark the positions of M11 and S22 transitions, respectively. Reproduced with permission from Ref. [61], © American Chemical Society 2013. (e) A collection of subthreshold plots of transfer curves for transistors built with purified individual semiconducting nanotubes. Inset: SEM image of a device with a single nanotube spanning the source and drain electrodes. Applied V_{DS} is –0.5 V. Reproduced with permission from Ref. [70], © American Chemical Society 2016. (e) A collection of subthreshold plots of transfer curves for transistors built with aligned arrays of purified semiconducting nanotubes. Inset: SEM image of a device with ~ 2,200 nanotubes spanning the source and drain electrodes. Reproduced with permission from Ref. [73], © Lu, L. J. et al. 2020.

structures, the length of their alkyl side chains, and solvent properties on the selectivity of nanotube sorting is still very limited, and thus the rational design of the polymer structures for better sorting efficacy has not yet been realized [66, 67]. (3) Selective redox reactions: The sorting of nanotubes based on electronic types can also be accomplished based on the difference in their chemical reactivity in redox reactions [68, 69]. Since redox reactions involve the transfer of electrons, their reaction kinetics and equilibrium are sensitive to the different electronic structures of metallic versus semiconducting nanotubes. The heterogeneous oxidation of nanotubes suspended in water by the dissolved oxygen/water couple is a good example [70]. For this charge-transfer reaction to take place, there must be both available electrons on the nanotubes and available unoccupied states of the O_2/H_2O couple at the same energy level. Since the energy states of the O_2/H_2O couple can be precisely tuned by adjusting the pH following the Nernst equation, at optimized pH, this redox reaction only proceeds appreciably for metallic nanotubes but at a much slower rate for semiconducting nanotubes, where the overlap in energy states between nanotubes and the O_2/H_2O couple is drastically reduced or even eliminated due to the presence of the semiconducting nanotube bandgap (Fig. 7(c)). As a result, the metallic nanotubes are more easily oxidized to bear a positive surface charge, while the semiconducting nanotubes retain their initial negative surface charge. Such surface-charge polarity difference therefor enables their effective separation using simple column chromatography. This method can prepare high-purity semiconducting nanotubes with high throughput, with its sorting mechanism quantitatively described by the classical Marcus–Gerischer model as guidance for rational process optimizations.

Despite their different mechanisms, techniques based on either selective charge transfer, molecular recognition, or redox reactions can in fact all produce high-purity semiconducting nanotubes. However, a significant challenge is to accurately and quantitatively assess the purity level. Optical spectroscopy has been widely utilized for its simplicity [71]. The metallic and semiconducting nanotubes have different absorption peaks associated with the interband transitions between the van Hove singularities in their electronic-state densities. The suppression of the metallic nanotube peak intensity indicates the enrichment of semiconducting nanotubes and *vice versa* (Fig. 7(d)). However, this method only has the capability to quantify the semiconducting nanotube purity up to 99%, as metallic nanotubes with abundance less than 1% cannot be detected in the absorption spectra [61]. Electrical testing as the second option can quantify the purity of semiconducting nanotubes above 99% (Fig. 7(e)). However, the process is tedious. It involves the fabrication of hundreds of thousands of transistors each nominally incorporating one or two nanotubes as the channel, followed by the characterization of their transfer characteristics. The numbers of metallic and semiconducting devices are then counted and compared to determine the nanotube purity. The purity of semiconducting nanotubes sorted based on all three mechanisms has been quantified by this approach, and all of them can produce semiconducting nanotubes with purity approaching or above 99.99% [61, 70]. However, this method has poor scalability toward determining the portion of metallic nanotubes in the required part-per-million (ppm) or ppb level where millions or even billions of nanotube transistors will have to be fabricated and tested. One approach to increase the throughput is to have multiple nanotubes in the form of aligned arrays as the channel instead of individual nanotubes (Fig. 7(f)) [72, 73]. For example, 1,000 transistors each incorporating ~ 2,200 nanotubes were fabricated to verify the purity of semiconducting nanotubes sorted via the selective dispersion by poly[9-(1-octylonoyl)-9H-carbazole-2,7-diyl] (PCz) molecules to give a record-high purity level above 99.9999% [73]. However, the throughput is still low as the routine material metrology method in the future manufacturing process. In addition, the defects associated with either the nanotube assembly or the device fabrication process will compromise the results, leading to inaccurate estimation of the metallic nanotube abundance.

In summary, various approaches have been developed recently to enrich the semiconducting nanotubes to purity above 99.99%. Since they are based on different mechanisms, combining them orthogonally together is expected to achieve even higher purity, potentially close or even surpass the 0.1–1 ppb target. However, the lack of reliable and high-throughput approaches to quantitatively assess the purity of sorted nanotubes at these levels is a significant bottleneck, currently for the refinement of the sorting techniques and in the future for the material-quality control in the manufacturing process.

3.2 Placement of purified nanotubes into high-density aligned arrays with uniform pitch

After obtaining high-purity semiconducting nanotubes, which are typically suspended as random coils in solvents, the next step is to organize them into aligned arrays on solid substrates suitable for subsequent device fabrications. The generated nanotube arrays must have high density to score high device on-state current density and low width-normalized contact resistance, and at the same time exhibit uniform pitch with low defect density to ensure that roughly the same number of nanotubes will be incorporated into each transistor with the same channel width for the required device-performance uniformity. With these constrains, the approaches where well aligned arrays of nanotubes are directly grown on single-crystalline substrates by CVD followed by the selective removal of metallic nanotubes by chemical, optical, electrical, or thermal methods (assembly-first-separation-last) become less attractive [74–77]. Because unless 100% pure semiconducting nanotubes can be directly synthesized, which might be realized in the future but the current progress is still far from this target, the removal of metallic nanotubes post-synthesis from arrays will inevitably lead to the local change of nanotube density and cause performance variability for transistors incorporating them as the channel [78]. Therefore, here I mainly focus on the assembly of purified semiconducting nanotubes from solution into aligned arrays (separation-first-assembly-last), which is more technology-relevant in near term given current progress, and discuss three most promising strategies.

(1) Templated assembly on chemically patterned substrates: In templating methods, the surface of the substrate can be modified chemically and/or geometrically to assist the selective placement of nanotubes on specific regions for the generation of partially aligned arrays with uniform pitch (Fig. 8(a)). For example, the surface of the silicon-wafer substrate can be firstly patterned using standard micro-fabrication techniques into narrow HfO_2 trenches separated by SiO_2 barriers. Due to their different isoelectric points (1.7–3.5 for SiO_2 and 6–7 for HfO_2), the HfO_2 surface can be selectively functionalized with 4-(N-hydroxycarboxamido)-1-methylpyridinium iodide (NMPI) self-assembled monolayer, based on the selective reaction between the hydroxamic-acid end groups of NMPI molecules and the protonated-HfO_2 surface at nearly neutral pH. The positive charge on the opposite pyridinium-salt end group can then electrostatically bind with the negatively charged sulfonate groups of the sodium dodecyl sulfate (SDS) surfactants wrapping around the nanotubes in their aqueous suspension,

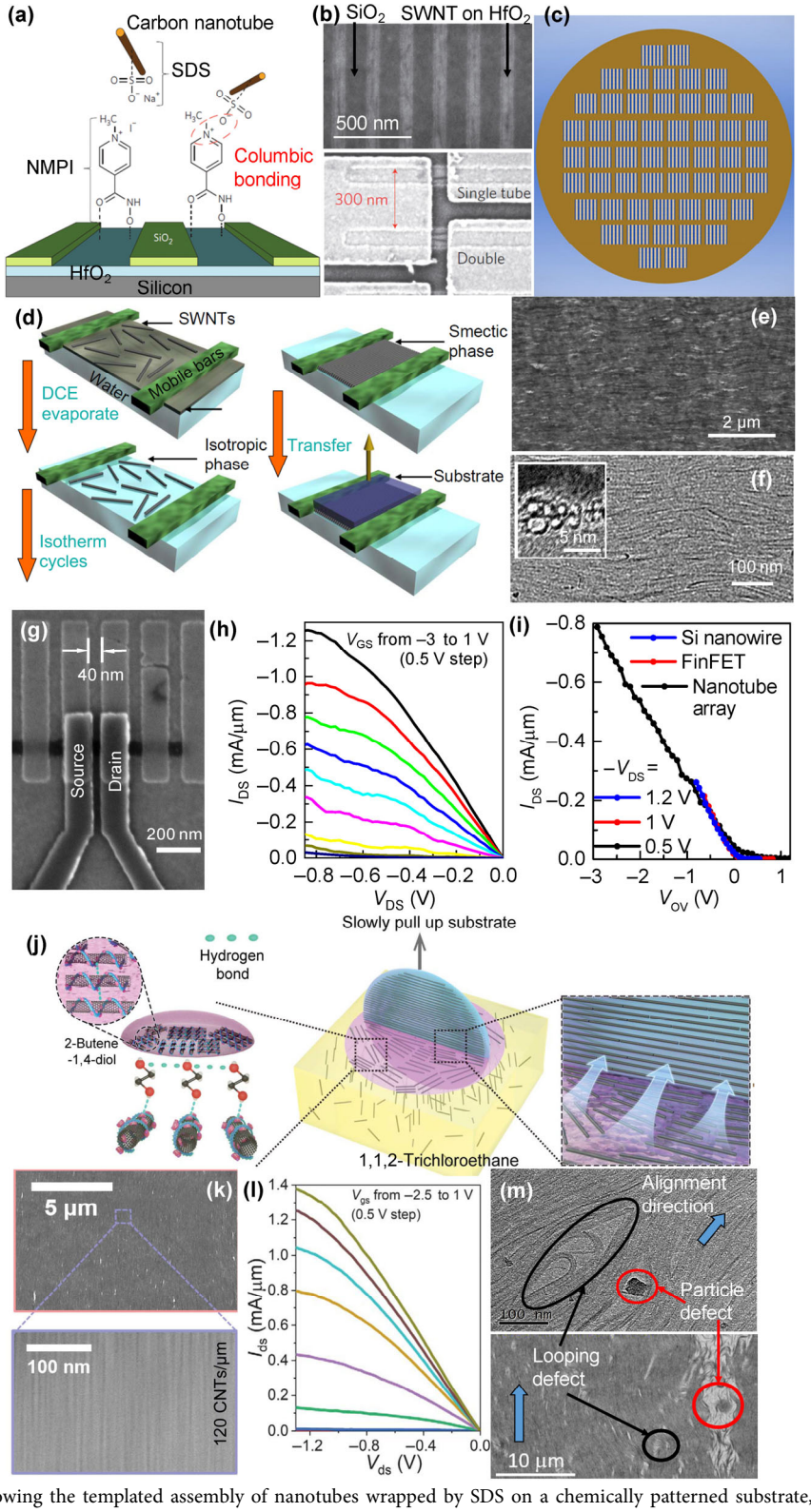


Figure 8 (a) Schematic showing the templated assembly of nanotubes wrapped by SDS on a chemically patterned substrate, where HfO_2 trenches are selectively functionalized by NMPI. (b) SEM images showing the nanotube arrays assembled with 200 nm pitch on a selectively functionalized HfO_2/SiO_2 substrate (top), which are further integrated into device arrays with the nanotube placed within one 70 nm-wide trench as the channel for each transistor (bottom). Reproduced with permission from Ref. [79], © Macmillan Publishers Limited 2012. (c) Schematic showing the placement of nanotubes over wafer scale based on the templated assembly. (d) Schematic showing the Langmuir-Schaefer assembly of nanotubes. (e) SEM image showing the Langmuir film composed of high-density aligned arrays of nanotubes over large area. (f) Top-view TEM of the same nanotube array. Inset: high-resolution cross-sectional TEM showing the circumference of assembled nanotubes. Reproduced with permission from Ref. [84], © Macmillan Publishers Limited 2013. (g)–(i) SEM image (g) and current–voltage characteristics (h) of a 40 nm-footprint transistor (20 nm L_g) incorporating the nanotube arrays assembled by the Langmuir-Schaefer process as the device channel. Its transfer characteristics (i) measured under a low V_{DS} of –0.5 V (black) are benchmarked with those of ultrathin-body silicon finFET (red, applied V_{DS} = –1 V) and nanowire FET (blue, applied V_{DS} = –1.2 V). Reproduced with permission from Ref. [33], © Cao, Q. et al. 2017. (j) Schematic illustration of the evaporation-driven assembly of nanotubes at the liquid-solid-air contact line. (k) SEM images showing the formed nanotube arrays at low (top frame) and high (bottom frame) magnifications. (l) Current–voltage characteristics of a 100 nm-L_g transistor incorporating such aligned arrays as the channel. Reproduced with permission from Ref. [73], © Liu, L. J. et al. 2020. (m) SEM micrographs highlighting the common defects observed in nanotube arrays assembled by the Langmuir-Schaefer process (top frame) and the dimension-limited self-alignment (DLSA) procedure (bottom frame), respectively.

and place the nanotube-surfactant complexes within the HfO_2 trenches [79]. For nanotubes suspended in organic solvents, 11-(4 amino-phenoxy)-1-undecylhydroxamic acid (AMUHA) monolayer can be used instead of NMPI, where the amino end groups of AMUHA can be converted to diazonium salts *in situ* to bind with the nanotube-polymer complexes through the diazo-coupling reaction [80]. In both cases, if the trench width is much smaller than the tube length, placed nanotubes can be partially aligned due to the spatial confinement as revealed in the scanning-electron microscopy (SEM) micrographs of the trench arrays post-assembly (Fig. 8(b)) [79]. The shear force can be utilized to improve the degree of alignment [81]. The nanotubes deposited into each trench can be incorporated into transistors, giving an electrically verified placement yield above 90% across the wafer. The major benefit of the templating method is its high throughput and good scalability into wafer scale, as both the surface-functionalization and nanotube-assembly steps only involve simply submerging the fabricated wafer substrates into the appropriate solution or suspension (Fig. 8(c)). However, its capability to generate high-density nanotube arrays required for high-performance nanotube transistors has not been verified in experiment. It is expected to be quite challenging and expensive to pattern the surface structures with sub-10 nm pitch over the entire wafer. And even if it can be accomplished based on the state-of-the-art extreme ultraviolet (EUV) lithography or block-copolymer nanolithography, the shrinkage of the pitch inevitably reduces the barrier width for the spatial confinement, which will likely increase the nanotube crossing defect density.

(2) Self-assembly by Langmuir–Blodgett/Langmuir–Schaefer techniques: Langmuir–Blodgett/Langmuir–Schaefer techniques are commonly used to fabricate closely packed superstructures of nanomaterials [82], and are also applicable toward the self-assembly of carbon nanotubes into highly anisotropic aligned arrays [83, 84]. The assembly process (Fig. 8(d)) is carried out in a water-filled trough equipped with a pair of mobile barriers and a pressure sensor, with ideally the whole setup placed inside a high-efficiency particulate air (HEPA) filtered glove box on top of an air-suspension table to minimize the impacts from particles and vibrations on the assembly quality [84]. Sorted semiconducting nanotubes suspended in organic solvent are then dispensed on top of the water surface. The key requirement here is that there is no non-volatile species within the suspension except the nanotube-polymer complexes, as any other non-volatile content will be incorporated as defects in the assembled nanotube arrays. As driven by the high surface tension of water, the suspension spreads to cover the whole air–water interface. The organic solvent will quickly evaporate, leaving behind the randomly orientated nanotubes floating on water surface. The mobile bars then laterally compress the nanotubes at a controlled speed. A typical assembly process involves multiple compression-release cycles. When compressed to high surface pressure, the nanotubes are forced to condense into a solid, cohesive film to minimize the inter-tube spacings. When released to low surface pressure, the nanotubes are allowed to re-orient themselves for the reduction of curving or looping defects. At the end of the compression-release cycles, the highest packing density will be achieved when the nanotubes are aligned one next to another in parallel with the mobile bars. The formed arrays can be transferred onto solid substrates by either vertical (Langmuir–Blodgett) or horizontal (Langmuir–Schaefer) dip-coating. The films are composed of high-density, well-aligned arrays of nanotubes over large area as shown in the SEM image (Fig. 8(e)). Higher-resolution TEM top-view and cross-sectional micrographs (Fig. 8(f)) further reveal that the nanotubes assembled are one next to another with the pitch about 2 nm as self-limited by the nanotube diameter plus their van der Waals separation. The nanotube arrays can be incorporated as the channel of scaled transistors to characterize their electrical properties (Fig. 8(g)) [33]. These nanotube-array devices exhibited excellent performance with on-state current density above 1 mA/μm (Fig. 8(h)). Their transfer characteristics under the same gate-overdrive voltage (V_{OV}, gate voltage applied above V_T) were benchmarked with advanced silicon finFETs and nanowire transistors, illustrating the capability of nanotube arrays to deliver the same performance in term of I_{on} density under a much lower power consumption with more than halved V_{DS} (Fig. 8(i)). Note here that these transistors adopted the end-bonded contact scheme to reduce the parasitic resistance, which otherwise would be pretty high in devices with Pd side-bonded contacts because of insufficient metal contact area for each individual nanotube inside such high-nanotube-density arrays [84, 85]. Compared to the templating method, the Langmuir–Blodgett/Langmuir–Schaefer techniques produce nanotube arrays with the ultimate density of uniform 2 nm nanotube pitch. However, the throughput is lower with the multiple isothermal compression-release cycles required, and their scalability toward the wafer-scale assembly could be limited.

(3) Directed assembly at the liquid-solid-air contact line: The evaporation-driven assembly of nanotubes at the liquid-solid-air contact line has potential to achieve both high nanotube density and high-throughput, wafer-scale deposition at the same time. As illustrated schematically in Fig. 8(j) [73], in this process, the substrate is immersed into the nanotube suspension, and slowly pulled out vertically. The nanotubes, which are randomly oriented in the bulk of the suspension, will be pinned along the horizontal direction at the liquid–solid–air interface, and get deposited onto the wafer surface with the slow evaporation of solvent. Convection flow during evaporation then brings additional nanotubes to the moving contact line to form continuous arrays. The interface properties, the evaporation rate of solvent, the pulling rate of the substrate, the nanotube concentration in the suspension, and the nanotube-surface interactions are all critical parameters determining the nanotube density and array structures. This process has been drastically improved during the past decade. In the initial demonstration, bare SiO_2/Si substrates were pulled out of the aqueous suspension of nanotubes to produce arrays with density limited to 10–20 nanotubes per micron [86]. Confining the nanotubes within a thin layer of fast-evaporating organic solvent floating on top of the water base helped to increase the density up to about 50–60 tubes/μm [87–89]. In the most recent design, high-purity semiconducting nanotubes were dispersed in trichloroethane and covered by an immiscible low-density top-layer organic solvent such as 2-butene-1,4-diol. This ultrathin top layer not only helps to limit nanotubes' rotational degree of freedom for better alignment, but also can readily spread on the receiving substrate surface with its low surface tension to increase the evaporation and deposition rate, which further improved the assembled nanotube density up to 100–200 nanotubes per micron (Fig. 8(k)) [73]. Transistors incorporating such nanotube arrays as the channel exhibited similar on-state current density compared to those based on the nanotube Langmuir films (Fig. 8(l)), while the process is more scalable with wafer-scale deposition already demonstrated in experiment.

In summary, the assembly of the high-purity semiconducting nanotubes into well-ordered aligned arrays up to 200–500 nanotubes per micron with uniform 2–5 nm pitch has been accomplished in experiment, and these formed semiconducting nanotube arrays can be utilized to fabricate transistors with I_{on}

density and transconductance comparable to, or even exceeding, those of current silicon transistors. However, for all these processes, the density of structural defects in the assembled nanotube arrays, caused by either nanotube looping or particulate impurities in the nanotube suspensions, still need to be reduced (Fig. 8(m)) [73, 84]. Replacing individual nanotubes with small-scale arrays as the initial building blocks, which are pre-assembled in suspension with the help of carefully designed DNA oligomers, might be helpful with their increased dimensions and structural rigidity [90–92]. Protocols to visualize and quantify the defect density with high throughput over wafer scale are required to assist further process optimizations.

3.3 Constructing reliable and high-performance n-channel nanotube transistors

With high-density semiconducting nanotube arrays becoming available, the next step is to integrate them into transistors and circuits. High-performance p-channel transistors are display in Figs. 8(h) and 8(l). For complementary logic circuits, n-channel transistors with reasonably matched performances are also required. The electron and hole mobilities in carbon nanotubes are comparable to each other, which is a significant advantage compared to III-V semiconductors [4]. However, with a work function of ~ 4.7 eV, as an intrinsic semiconductor with an average bandgap of 0.6 eV, nanotubes, when incorporated into transistors employing common air-stable high-work-function metals as contacts, typically exhibit p-channel operations. To realize n-channel nanotube transistors, a low-work-function contact has to be utilized to allow the injection of electrons with minimal barrier to the nanotube conduction band (Fig. 9(a)). Several rare-earth metals, including yttrium (Y), scandium (Sc), erbium (Er), and gadolinium (Gd), have been suggested as suitable low-work-function contacts to enable n-channel nanotube transistors with performance comparable to their p-channel counterparts (Fig. 9(b)) [93–97]. The major challenges for this approach are two-folds. First, most of these low-work-function metals are air-sensitive. Therefore, ultra-high-vacuum

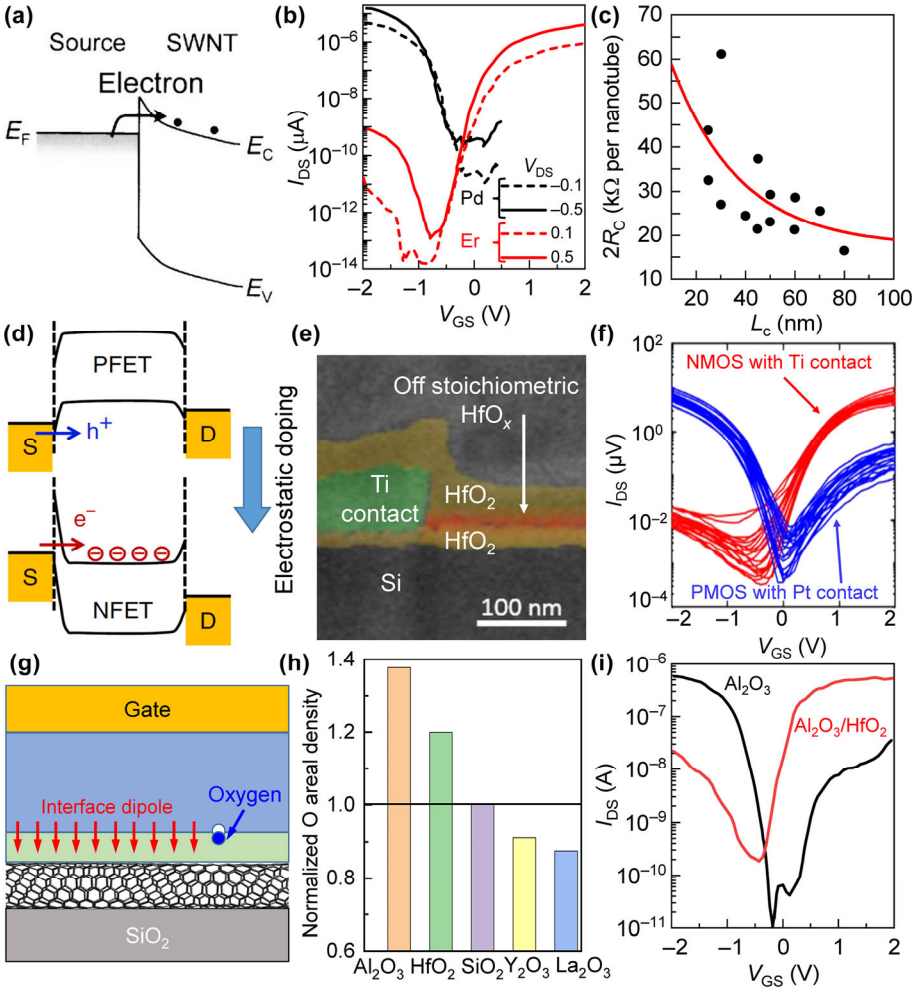


Figure 9 (a) Band-diagram illustration of n-channel nanotube transistors enabled by a low-work-function metal contact. (b) Transfer characteristics of individual nanotube transistors with Pd (black) and Er (red) contacts under low (0.1 V, dashed lines) and high (0.5 V solid lines) source-drain bias. Reproduced with permission from Ref. [94], © IEEE 2011. (c) L_c-scaling characteristics of the low-work-function Sc contact to nanotubes. Reproduced with permission from Ref. [32], © American Association for the Advancement of Science 2017. (d) Band-diagram illustration of n-channel nanotube transistors enabled by electrostatic doping, showing that the carrier type tunneling through the Schottky barrier at contacts converts from holes to electrons. Reproduced with permission from Ref. [101], © Macmillan Publishers Limited, part of Springer Nature 2018. (e) Cross-sectional TEM image showing the n-channel nanotube FET with the channel covered by the nonstoichiometric doping oxide of HfO_x. (f) Transfer characteristics of n-channel nanotube transistors with nonstoichiometric HfO_x encapsulation and Ti contacts (red) and p-channel nanotube transistors with Pd contacts (blue). Applied V_{DS} is |2| V. Reproduced with permission from Ref. [102], © American Chemical Society 2018. (g) Schematic showing the n-channel nanotube transistor with dipoles formed across the interface between two oxides within its gate stack. (h) Normalized areal density of oxygen atoms in various oxides. Reproduced with permission from Ref. [106], © American Institute of Physics 2009. (i) Transfer curves of a p-channel nanotube transistor with Al_2O_3 gate oxide (black) and an n-channel nanotube transistor with Al_2O_3/HfO_2 bilayer gate stack (red). Applied V_{DS} is |0.5| V. Reproduced with permission from Ref. [36], © American Chemical Society 2013.

chambers are required for their deposition, and careful passivation of both the top surface and the sidewalls of the metal electrodes are required to ensure their stability when exposed to ambient environment, raising concerns over the device reliability, fabrication throughput, and device scalability [98, 99]. Second, similar as the side-bonded Pd contacts, the parasitic $2R_c$ of these low-work-function metal contacts to nanotubes also increases sharply with the reduction of L_c, and the trend could be even worse due to their poor wetting on the nanotube surface (Fig. 9(c)) [32]. The end-bonded n-contact scheme will be difficult to realize as low-work-function metals will prefer to reacting with the gate oxides under annealing, instead of with carbon nanotubes to form carbide bonds.

Another approach is to electrostatically dope the nanotube channel by fixed charges in its vicinity. These fixed charges help to greatly bend down the energy bands. As a result, the Schottky barrier width for electrons to tunnel from high-work-function contacts into the conduction band of nanotubes is dramatically reduced under further applied positive gate voltage, which turns-on the device as an n-channel transistor (Fig. 9(d)). These fixed charges could be introduced by employing a thin layer of non-stoichiometric dielectrics, such as AlO_x, SiN_x and HfO_x, as part of the device gate stack (Fig. 9(e)) [100–102]. Here, air-stable metals with intermediate work function, such as titanium (Ti), can be used as the source-drain contacts to afford n-channel nanotube transistors with performance comparable to their p-channel counterparts built with Pd contacts (Fig. 9(f)) [102]. The major benefit of this approach is the drastically improved device air stability and simplified fabrication process, and therefore it is widely used in the construction of large-scale complementary nanotube logic circuits [103, 104]. In addition, the end-bonded contact scheme with high-work-function, carbide-forming metals as contacts can also be potentially utilized here instead of side-bonded Ti contacts to address the L_c-scaling issue [47]. The major concern is that the intentional introduction of these fixed charges within gate dielectrics could adversely affect the transistor reliability by increasing the time-dependent device V_T instability.

A promising alternative option is to electrostatically dope the nanotubes using the dipoles formed at the interface between oxides with different oxygen areal density (Fig. 9(g)), instead of fixed charges caused by vacancies. When two oxides are in contact, oxygen ions tend to diffuse following the density gradient. Such movement creates a charge imbalance across the interface, and thus a dipole moment [105]. The formed dipole moments then electrostatically dope the nanotubes similar as fixed charges and thus facilitate the electron conduction. However, since this process is driven by structural stabilization of the interface, the formed stable Frenkel defects have much less impact on the device reliability compared to vacancies (a similar strategy has been widely used to tune the device V_T for silicon transistors since 28 nm technology node) [106, 107]. In this approach, the gate stack is composed of two layers of stoichiometric oxides. The first layer is only a few atomic-layer thick, and is composed of oxides with low oxygen areal density (Fig. 9(h)). Another layer of oxide with higher oxygen density will then be formed on top as the bulk of the gate dielectric [106]. In experiment, because of these dipole moments, nanotube transistors employing the Al_2O_3/HfO_2 bilayer gate dielectric behaved as n-channel transistors, while those with only Al_2O_3 gate dielectric exhibited normal p-channel operations, both with high-work-function Pd as the source-drain contacts (Fig. 9(i)). In addition to realize reliable n-channel nanotube transistors, this approach can also be applied to dope the spacer—or source/drain extension—regions of self-aligned nanotube transistors.

In summary, various approaches have been established to fabricate n-channel nanotube transistors with performance commensurable with their p-channel counterparts. Complementary-type nanotube circuits have been demonstrated, from high-performance ring oscillators with gigahertz oscillating frequency to fully functional microprocessors comprising more than 14,000 nanotube transistors [73, 99, 103]. However, the device reliability could be a concern due to the adoption of either the reactive low-work-function metal contacts or the off-stoichiometric gate dielectrics with high fixed-charge density, and research to evaluate the nanotube transistor reliability is still very limited [108].

3.4 Reducing device variability for integrated circuits

Probably the biggest remaining challenge for the nanotube transistor technology is the large device variability. The random defects and the pitch variations in the nanotube-assembly processes (Fig. 8(m)), as well as the wide distribution of the parasitic $2R_c$ (Fig. 4(j)), could lead to large device I_{on} variability, which I have discussed in previous sections. Moreover, large variations are also observed for nanotube transistors in terms of both V_T and SS. Transfer curves from a collection of nanotube transistors built on the same substrate, each with 1–2 nanotubes as the channel, are plotted in Fig. 10(a) to illustrate the severity of the problem [33]: In addition to the spanning of the device I_{on} over two orders of magnitude, the V_T varied over a range above 1 V, and the device SS changed from 60 up to 800 mV/dec. Having multiple nanotubes in parallel as the channel helps to narrow the I_{on} and V_T distributions due to averaging effects [109], but it degrades the device SS which will be dominated by those nanotubes having weak response to the applied gate electric field. For example, by randomly selecting 25 single-nanotube measurements out of the 600 displayed in Fig. 10(a) and adding their output currents as if they were all connected in parallel, the SS of the simulated nanotube-array transistors all become larger than 400 mV/dec (Fig. 10(b)), which agrees with what was observed in experiment for transistors based on high-density nanotube arrays assembled by the Langmuir-Schaefer method as the device channel (Fig. 10(c)).

Therefore, to improve the uniformity and performance of nanotube-array transistors for integrated circuit applications, the variability has to be minimized on individual nanotube basis, and a critical question to ask first is where the variability of single-nanotube transistors as displayed in Fig. 10(a) comes from. The diameter distribution of the nanotubes accounts for only a very small portion of the observed variations, but the fixed charges randomly placed at the gate-oxide surface, as schematically illustrated in Fig. 10(d), could be a major contributor [110]. Since nanotubes are a quasi-one-dimensional nanomaterial where carriers cannot pass around local barriers as in devices based on conventional bulk semiconductors, even a single charge in their vicinity can dramatically affect the device operation, both shifting the V_T by several hundred millivolts and drastically degrading the SS. A microscopic Monte-Carlo model was built to quantitatively evaluate the impact of such stray charges on the carrier transport through a ballistic nanotube channel [111]. With a surface-charge density of 5×10^{12} cm^{-2}, which is comparable to the value reported for pristine SiO_2/air interface as measured by electrostatic-force microscopy, this simple model can successfully describe the changes of both average V_T (Fig. 10(e)) and V_T variability (Fig. 10(f)) as a function of t_{ox} observed in experiment. The ~ 50% smaller V_T standard deviation obtained in simulation is largely because the nanotubes in simulation are fully passivated by SiO_2, which more effectively screens these stray charges

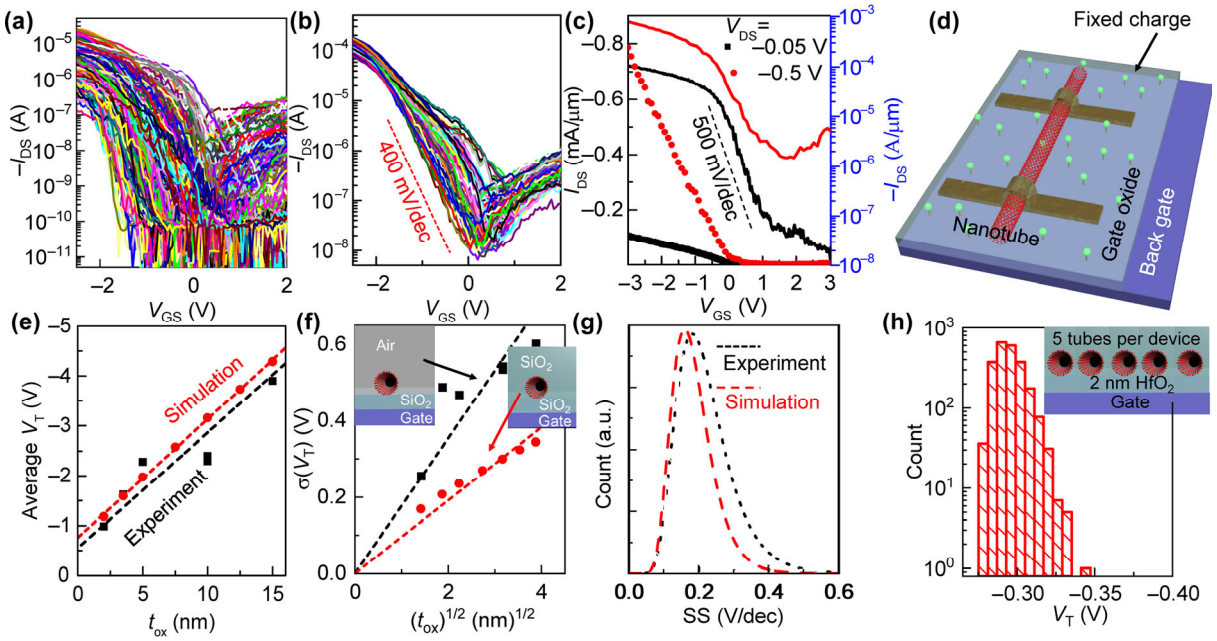


Figure 10 (a) Transfer characteristics of 635 transistors each built on nominally a single nanotube as the channel. Applied V_{DS} is −0.5 V. (b) Simulated transfer curves of nanotube-array transistors where the carrier transport through 25 nanotubes randomly selected out of the collection displayed in (a) are added as they are connected in parallel. The red dashed line serves as a visual guide to illustrate the minimal SS obtained as above 400 mV/dec. (c) Measured transfer characteristics of a nanotube-array transistor with 50–100 nanotubes as the channel displayed in both linear (symbols, left axis) and logarithmic (lines, right axis) scales under low (−0.05 V) and high (−0.5 V) source-drain bias. The black dashed line serves as a visual guide to illustrate the SS obtained at around 500 mV/dec. Reproduced with permission from Ref. [33], © Cao, Q. et al. 2017. (d) Schematic illustrating the randomness of stray charges at the oxide surface affecting the performance variability of nanotube transistors. (e) and (f) The change of the average V_T (e) and the standard deviation of V_T ($\sigma(V_T)$) as a function of the SiO_2 gate dielectric thickness (t_{ox}) for single-nanotube transistors as determined by experiment (black) and predicted in simulation (red) where only the contribution from the fixed random distribution of charges at the oxide surface is considered. Inset: device cross-sectional schematics for the devices in experiment versus in simulation. (g) The distribution of SS for nanotube transistors with t_{ox} = 2 nm as determined in experiment (black) and predicted in simulation (red). Reproduced with permission from Ref. [111], © American Physical Society 2015. (h) Simulated V_T distribution for planar nanotube transistors with 5 nanotubes in parallel as the channel, and 2 nm HfO_2 as the gate dielectric. The oxide surface-charge density is assumed to be 5×10^{11} cm^{-2}. Inset: device cross-sectional schematic.

with its higher dielectric constant of 3.9, while in experiment the nanotubes were exposed to air whose relative dielectric constant is around 1 instead. In addition to V_T variability, the simulation further proved that these randomly placed fixed charges on the oxide surface are also an important factor responsible for the large SS variations (Fig. 10(g)). These results raised a critical question for technology development: Will nanotubes' intrinsic susceptibility to stray charges make it impossible for us to control the V_T and SS variability of nanotube transistors to the level required for very-large-scale integration? To answer this question, we ran a simulation to predict the V_T and SS variations of hypothetical transistors composed of five nanotubes in parallel with 2 nm HfO_2 as the gate dielectric. The fixed charge density on the gate-oxide surface is assumed to be 5×10^{11} cm^{-2}, which is 10× smaller than what we currently have for unpassivated SiO_2 surface but typical for the optimized HfO_2–Si interface, and the nanotubes are fully passivated by HfO_2. The reduction of the fixed-charge density, the increase of the gate capacitance, and the adoption of multiple nanotubes as channel effectively reduce the device V_T variability with the range of V_T shift staying below 75 mV even in extreme cases (Fig. 10(h)). It is sufficient to meet the current design rule where the maximum V_T shift from the average is required to be smaller than 150 mV. The worst SS is lower than 110 mV/dec with the average at about 90 mV/dec, still better compared to the best silicon transistors at 10 nm L_g.

Guided by such mechanistic understanding of nanotube transistor variations, the passivation of the oxide surface by organic self-assembled monolayers to reduce the surface-charge density and the adoption of electrolyte gating to better screen these stray charges have both been demonstrated in experiment as effective approaches to reduce the variability of nanotube devices and improve their SS [73, 112]. But we are still far away from the target. For nantube transistors, the gate-all-around configuration, where the gate electrode will screen external charges, might be particularly attractive, not just for better electrostatic coupling but also for minimizing the device variability caused by the randomness of stray charges. In parallel, we will also need new approaches to conformlly deposit high-quality gate oxides on nanotubes with low density of traps at the nanotube–oxide interfaces [39].

3.5 Qualifying nanotubes in conventional semiconductor foundry

Looking further down the road, even if we have all the processes to remove metallic nanotubes below ppb level, align nanotubes uniformly into aligned arrays on wafer scale, realize reliable p-channel and n-channel nanotube transistors with comparably high performance, and control the device variability to an acceptable level for integration, we will still face a critical challenge of transferring the manufacturing of the nanotube transistors and circuits to a commercial semiconductor foundry where chips based on silicon transistors will be fabricated in parallel: Considering the inertia around silicon and the cost of the leading-edge semiconductor fabs, nanotubes must make themselves compatible with the silicon foundry otherwise the risk would be too high for any commercial company to absorb. Semiconductor fabs have extremely stringent limits on chemical and particulate contaminations to ensure the device yield, which is critical to their commercial success [113, 114].

It therefore requires that the deposition of nanotube arrays cannot introduce virtually any particles or metal ions to the wafer surface, which will not only affect the yield and uniformity of nanotube transistors but also contaminate the processing tools. However, it is extremely difficult for carbon nanotubes: The growth of nanotubes involves the use the metal nanoparticles as the catalyst. The dispersion of nanotubes in solvents involves the high-power ultrasonication where the metal ions from the sonication tip could leach into the nanotube suspensions. The various chemicals used in the purification and dispersion of nanotubes, especially those organic surfactants, are not electronic grade and could contain various metal ions and organic/inorganic particles. It requires a lot of careful engineering work to firstly identify the sources of all these contaminations, and then remove them from the process flow of nanotube transistors. However, we currently lack appropriate venues to support associated research and development efforts: It is clear beyond the scope of conventional academic research, but without establishing the compatibility, nanotubes will not be welcome even at semi-production lines in industry research labs.

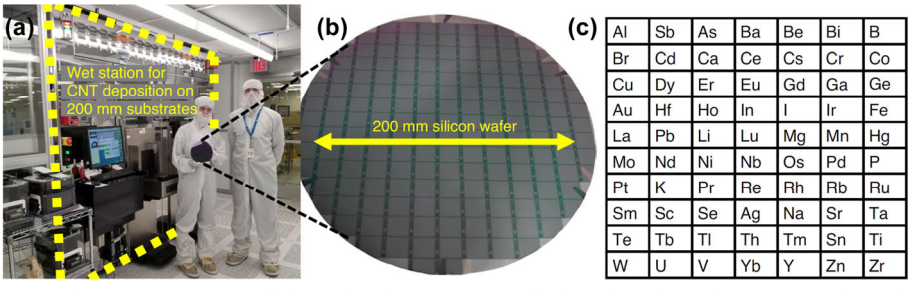
Despite the difficulty, we have started to see some promising initial efforts. Shulaker group at Massachusetts Institute of Technology teamed up with Analog Devices and SkyWater Technology [115]. They have successfully qualified the deposition of nanotubes and the fabrication of nanotube transistors within existing silicon manufacturing facilities (Fig. 11(a)). Nanotube transistors with high yield and uniformity can be fabricated across 200 mm-diameter wafers (Fig. 11(b)). Under optimized conditions, none of the contaminations listed in Fig. 11(c) was identified by inductively coupled plasma mass spectrometry on the wafer surface after nanotube deposition. In the future, high throughput inline contamination measurement techniques, such as the energy dispersive X-ray spectroscopy, surface photo-voltage analysis, and total X-ray reflection fluorescence, need to be integrated with the processing of nanotubes, which requires close collaboration among research teams, semiconductor foundries, and tool companies. In addition, appropriate channels need to be established so that the best practices can be shared in the field but the trade secrets and intellectual properties of industry partners can also be adequately protected.

4 Outlook

Since their first report in 1998 [9], carbon nanotube transistors have been acclaimed as one of the most viable options to take torch from silicon-based MOSFETs as the "next switch" in extremely scaled, high-performance logic chips. After 20 years of sustained efforts by the globe scientific community, the scaling advantages of nanotubes, resulting from their intrinsically high carrier saturation velocity, nanometer thickness, and favorable contact-scaling characteristics, have all been well established in experiment [31, 32, 46, 50]. High-performance nanotube transistors with merely 40 nm overall footprint have been fabricated and they experimentally confirmed the attractive properties of nanotubes at extremely small device dimensions compared to what is possible with silicon [33]. 2–3 times better performance under 2–3 times lower energy consumption is expected by taking such devices as drop-in replacement of silicon MOSFETs in current microprocessors [53]. Regarding manufacturability, semiconducting nanotubes with purity above 99.9999% have been prepared in bulk quantity, and they can be placed as parallel arrays with uniform 2–5 nm pitch [73, 84]. Both p- and n-channel nanotube transistors are fabricated in yield high enough for the realization of complex complementary-type large-scale integrated circuits [103]. Nanotubes have entered the door to some commercial semiconductor foundries [115]. The device variability remains a critical issue, but the key contributor has been identified with various options on the table to keep it under control [111]. It is clear that nanotube transistors have successfully evolved from toys in research labs to a serious device technology, and therefore they are attracting interest and investment from not only federal agencies but also the semiconductor industry as led by companies including IBM, Taiwan Semiconductor Manufacturing Company (TSMC), and Analog Devices, as well as many startups such as Nantero and Carbonics.

Despite the impressive progress made by the whole community, it is still a grand challenge with many loose ends to tie up. Many are correlated with the metrology and quality control. For example, high-throughput and high-sensitivity techniques to detect metallic nanotubes and the structural defects in assembled nanotube arrays have yet to be developed. A combination of advanced optical spectroscopy, high-throughput electron microscopy, and machine-learning-assisted image analysis are likely required. Others are mainly about the device uniformity and reliability. The $2R_c$, I_{on}, V_T, and SS variations of nanotube transistors need to be dramatically reduced, possibility based on material-device-process co-optimizations. Existing works on nanotube device reliability are very limited [108], and it is expected to be a significant problem with current largely unoptimized nanotube–metal and nanotube–dielectric interfaces. Silicon-compatible processing protocols of carbon nanotubes need to be standardized. Many of these problems are beyond the capability of individual research groups and even single sectors of the semiconductor industry.

Given these challenges, totally replacing silicon with carbon nanotubes in our logic chips may not happen immediately. However, nanotube transistors are indeed getting ready for some niche applications, where their unique properties have already given them significant advantages over most of the competitors while their remaining challenges can be accommodated. First, because nanotubes can be deposited at room temperature from their suspensions, and fabricated into



Al	Sb	As	Ba	Be	Bi	B
Br	Cd	Ca	Ce	Cs	Cr	Co
Cu	Dy	Er	Eu	Gd	Ga	Ge
Au	Hf	Ho	In	I	Ir	Fe
La	Pb	Li	Lu	Mg	Mn	Hg
Mo	Nd	Ni	Nb	Os	Pd	P
Pt	K	Pr	Re	Rh	Rb	Ru
Sm	Sc	Se	Ag	Na	Sr	Ta
Te	Tb	Tl	Th	Tm	Sn	Ti
W	U	V	Yb	Y	Zn	Zr

Figure 11 (a) and (b) Nanotube-processing station (a) housed within a commercial silicon foundry to deposit nanotube random networks with good uniformity over 200 mm-diameter wafers (b). (c) A partial list of potential contaminants that need to be removed from the nanotube suspensions. Reproduced with permission from Ref. [115], © Bishop, M. D. et al. 2020.

high-performance transistors within a low thermal budget, nanotube transistors are especially suitable for the monolithic 3D integration on top of silicon chips with nanotubes' better electrical properties compared to competitors, such as low-temperature-deposited polycrystalline silicon and semiconductor metal oxides (Fig. 12(a)) [116–118]. These monolithic 3D-integrated circuits offer multiple advantages over the conventional planar-2D architecture, which include increasing the communication bandwidth through high-density inter-layer vias, reducing the interconnect power dissipation through more efficient vertical routing, and achieving higher device integration density without further scaling down the size of individual transistors [119–121]. Here, for instance, the nanotube transistors may not be used for constructing digital circuits for computing but driving the resistive-random-access memory cells integrated together at the back-end-of-line stack (Fig. 12(b)) [116, 117]. Therefore, the device performance and uniformity requirements could be somewhat relaxed. Second, in addition to their low-temperature processing, nanotubes' intrinsic ultrathin body and strong mechanical strength make them an ideal material candidate for building mechanically robust flexible electronics on plastic substrates. These flexible nanotube transistors can be integrated into complex digital circuits (Fig. 12(c)) and complementary-type ring oscillators with sub-10 nanosecond stage delay (Fig. 12(d)) [101, 122–124]. Their performances, even based on nanotube random networks, are substantially better compared to flexible electronic circuits built on conventional materials such as amorphous silicon and organic semiconductors, and therefore they are particularly attractive for the emerging edge-computing and Internet-of-things (IoT) applications. Moreover, with the solution-based process, the deposition of nanotube films can be easily scaled up to cover large area. Therefore, in addition to high-performance electronics, carbon nanotubes are also attractive for flexible macroelectronics

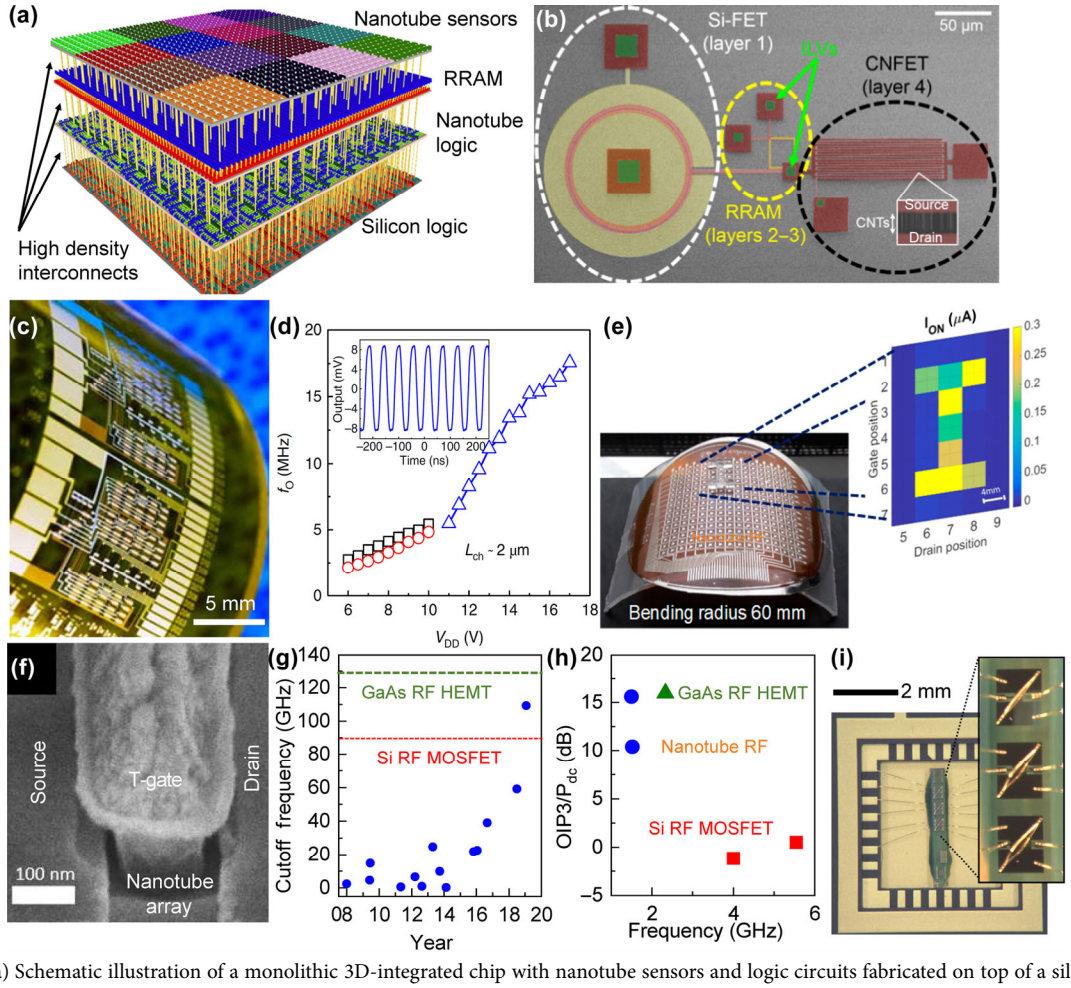


Figure 12 (a) Schematic illustration of a monolithic 3D-integrated chip with nanotube sensors and logic circuits fabricated on top of a silicon chip with high-density interconnects for high bandwidth interlayer communications. Reproduced with permission from Ref. [116], © Macmillan Publishers Limited, part of Springer Nature 2017. (b) False-colored SEM micrograph showing the nanotube select transistor integrated with two resistive random-access memory (RRAM) cells and a silicon pass transistor on different layers, which are connected through interlayer vias (ILVs). Reproduced with permission from Ref. [117], © IEEE 2014. (c) Optical image of a medium-scale integrated circuit composed of 88 nanotube thin-film transistors on a flexible kapton substrate. (d) Oscillation frequency (f_o) as a function of the applied drive voltage V_{DD} for a five-stage ring oscillator based on flexible nanotube thin-film transistors. Inset: output waveform with the oscillation frequency of 17.6 MHz. Reproduced with permission from Ref. [101], © Macmillan Publishers Limited, part of Springer Nature 2018. (e) Optical image of an artificial "electronic skin" with carbon nanotube active matrix for pressure sensing and mapping over large area on curved surfaces. Reproduced with permission from Ref. [126], © American Chemical Society 2018. (f) Tilted-view SEM micrograph showing the nanotube RF transistor with semiconducting nanotube arrays as the channel and T-shaped metal gate to minimize parasitics. Reproduced with permission from Ref. [134], © Rutherglen, C. et al. 2019. (g) The improvement of the cutoff frequency for nanotube RF transistors (blue dots) over the past decade, in comparison with the performance of silicon RF MOSFETs (red dashed line) and GaAs HEMTs (green dashed line). (h) The low-frequency linearity ratio (OIP3/P_{dc}) of state-of-the-art nanotube RF transistors (blue dots) in comparison with that of commercial silicon RF MOSFETs (red squares) and GaAs HEMTs (green triangles). Reproduced with permission from Ref. [127], © Cao, Q. 2019. (i) Optical image of three individually diced nanotube RF transistors wire-bonded into dual in-line package to realize a radio together with some other passive components. Reproduced with permission from Ref. [135], © The National Academy of Sciences of the USA 2008.

where the circuits need to spread cost-effectively over large area as interface with environment and even human body [125]. For example, nanotube films deposited from solution on large-area (5" square) plastic substrates can be used together with pressure-sensitive rubber to realize a type of flexible artificial "electronic skin" with high tactile spatial resolution and fast response to detect complex objects with excellent accuracy as future human–machine interface (Fig. 12(e)) [126]. Since those devices will likely be fabricated using dedicated large-area-applicable printing or shadow evaporation equipment, the process compatibility of carbon nanotubes with silicon will be of less a concern. In addition, the large device size helps to improve the device uniformity by averaging effect with each transistor employing thousands of nanotubes as the device channel. Finally, the exceptional electrical properties and the favorable scaling characteristics of carbon nanotubes, as well as their intrinsically linear amplification of input signals ensured by their 1D charge transport, make nanotube transistors especially suitable for radio-frequency (RF) applications, particularly in the high-frequency regime beyond the capability of conventional silicon-based RF devices [127, 128]. With the continuing efforts to improve the nanotube–metal contacts and increase the density of aligned semiconducting-enriched nanotube arrays as what I described in detail in previous sections, as well as the improvement of the gate electrode design to minimize the parasitic resistance and capacitance (Fig. 12(f)), the performance of nanotube RF transistors have been increasing steadily during the past decade [129–133]. Both their cut-off frequency (Fig. 12(g)) and the signal linearity in term of the output third-order intercept point (Fig. 12(h)) have surpassed those of silicon RF MOSFETs and approached the metrics of commercial GaAs HEMTs [134]. Compared to logic transistors, the performance of RF devices is relatively insensitive to I_{off}, and therefore the requirement on the semiconducting nanotube purity is greatly relaxed. In addition, compared to logic circuits composed of billions of transistors, these RF devices in many applications are individually tested, diced, and packaged. Therefore, the current problems associated with the variability and yield of nanotube transistors can be tolerated. Figure 12(i) shows a system where three individually packaged nanotube RF transistors were used as RF amplifier, RF mixer, and audio amplifier to receive, demodulate, and amplify signals broadcasted by commercial radio stations [135]. I believe that the commercialization of carbon nanotube transistor technology will likely start from one of these niche applications. Only after it happens, people can then begin gathering experiences and know-how on the production of nanotube transistors on industry scale, and learn how to reduce the device variability and improve the device yield in a manufacturing setting, which will eventually pave the road toward the adoption of carbon nanotube transistors in the much more demanding high-performance logic chips.

To summarize this review, carbon nanotube transistors have gone a long way toward commercialization. The long-term "Holy Grail" goal is to replace silicon in MOSFETs and help us sustaining Moore's law beyond the physical limits of silicon. The recent progress has suggested that it is not only scientifically attractive with the unique intrinsic properties of nanotubes but also technologically feasible. In the meantime, we expect to see sooner adoption of nanotube transistors in other niche applications in near term, which would be essential to build the critical mass around this technology toward commercialization. In recently years, we have witnessed the resurgence of interests in nanotubes from academia together with the increasing research and development activities around nanotubes from the semiconductor industry. For nanotube transistors to safely traverse the last miles in the valley of innovation death, and for the society to see the decade-long investment in this technology to finally pay off, a strong public-private partnership and collaborations across different sectors of the semiconductor industry are more important than ever.

Acknowledgement

This work is supported by the National Science Foundation (NSF) through grant ECCS-1950182.

References

[1] Thompson, N. C.; Greenewald, K.; Lee, K.; Manso, G. F. The computational limits of deep learning. 2020, arXiv: 2007.05558. arXiv.org e-Print archive. https://arxiv.org/abs/2007.05558v1 (accessed Apr 15th, 2021).

[2] Liu, Y. Q.; Qian, K.; Yu, J. L.; Wang, K.; He, L. Effective scaling of blockchain beyond consensus innovations and Moore's law. 2020, arXiv: 2001.01865. arXiv.org e-Print archive. https://arxiv.org/abs/2001.01865 (accessed Apr 15th, 2021)

[3] Leiserson, C. E.; Thompson, N. C.; Emer, J. S.; Kuszmaul, B. C.; Lampson, B. W.; Sanchez, D.; Schardl, T. B. There's plenty of room at the top: What will drive computer performance after Moore's law? *Science* **2020**, *368*, eaam9744.

[4] del Alamo, J. A. Nanometre-scale electronics with III–V compound semiconductors. *Nature* **2011**, *479*, 317–323.

[5] Li, M. Y.; Su, S. K.; Wong, H. S. P.; Li, L. J. How 2D semiconductors could extend Moore's law. *Nature* **2019**, *567*, 169–170.

[6] Bethune, D. S.; Kiang, C. H.; de Vries, M. S.; Gorman, G.; Savoy, R.; Vazquez, J.; Beyers, R. Cobalt-catalysed growth of carbon nanotubes with single-atomic-layer walls. *Nature* **1993**, *363*, 605–607.

[7] Iijima, S.; Ichihashi, T. Single-shell carbon nanotubes of 1-nm diameter. *Nature* **1993**, *363*, 603–605.

[8] Hamada, N.; Sawada, S. I.; Oshiyama, A. New one-dimensional conductors: Graphitic microtubules. *Phys. Rev. Lett.* **1992**, *68*, 1579–1581.

[9] Tans, S. J.; Verschueren, A. R. M.; Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **1998**, *393*, 49–52.

[10] Martel, R.; Schmidt, T.; Shea, H. R.; Hertel, T.; Avouris, P. Single- and multi-wall carbon nanotube field-effect transistors. *Appl. Phys. Lett.* **1998**, *73*, 2447–2449.

[11] Javey, A.; Guo, J.; Wang, Q.; Lundstrom, M.; Dai, H. J. Ballistic carbon nanotube field-effect transistors. *Nature* **2003**, *424*, 654–657.

[12] McEuen, P. L.; Fuhrer, M. S.; Park H. Single-walled carbon nanotube electronics. *IEEE Trans. Nanotechnol.* **2002**, *1*, 78–85.

[13] Collins, P. G.; Avouris, P. Nanotubes for electronics. *Sci. Am.* **2000**, *283*, 62–69.

[14] Bohr, M. A 30 year retrospective on Dennard's MOSFET scaling paper. *IEEE Solid-State Circuits Soc. Newsl.* **2007**, *12*, 11–13.

[15] Meyerson, B. Innovation: The future of silicon technology. In *Semico IMPACT Conference*, Scottsdale, AZ, USA, 2004.

[16] Liu, Y.; Luisier, M.; Majumdar, A.; Antoniadis, D. A.; Lundstrom, M. S. On the interpretation of ballistic injection velocity in deeply scaled MOSFETs. *IEEE Trans. Electron Devices* **2012**, *59*, 994–1001.

[17] Jeong, C.; Antoniadis, D. A.; Lundstrom, M. S. On backscattering and mobility in nanoscale silicon MOSFETs. *IEEE Trans. Electron Devices* **2009**, *56*, 2762–2769.

[18] Theis, T. N.; Solomon, P. M. It's time to reinvent the transistor! *Science* **2010**, *327*, 1600–1601.

[19] Lee, C. S.; Pop, E.; Franklin, A. D.; Haensch, W.; Wong, H. S. P. A compact virtual-source model for carbon nanotube FETs in the sub-10-nm regime—Part I: Intrinsic elements. *IEEE Trans. Electron Devices* **2015**, *62*, 3061–3069.

[20] Xu, L.; Qiu, C. G.; Zhao, C. Y.; Zhang, Z. Y.; Peng, L. M. Insight into ballisticity of room-temperature carrier transport in carbon nanotube field-effect transistors. *IEEE Trans. Electron Devices* **2019**, *66*, 3535–3540.

[21] Yan, R. H.; Ourmazd, A.; Lee, K. F. Scaling the Si MOSFET: From bulk to SOI to bulk. *IEEE Trans. Electron Devices* **1992**, *39*, 1704–1710.

[22] Ferain, I.; Colinge, C. A.; Colinge, J. P. Multigate transistors as the future of classical metal–oxide–semiconductor field-effect transistors. *Nature* **2011**, *479*, 310–316.
[23] Yao, J. X.; Li, J.; Luo, K.; Yu, J. H.; Zhang, Q. Z.; Hou, Z. Z.; Gu, J.; Yang, W.; Wu, Z. H.; Yin, H. X. et al. Physical insights on quantum confinement and carrier mobility in Si, $Si_{0.45}Ge_{0.55}$, Ge gate-all-around NSFET for 5 nm technology node. *IEEE J. Electron Devices Soc.* **2018**, *6*, 841–848.
[24] Takagi, S.; Koga, J.; Toriumi, A. Subband structure engineering for performance enhancement of Si MOSFETs. In *International Electron Devices Meeting. IEDM Technical Digest*, Washington, DC, USA, 1997, pp 219–222.
[25] Brunner, T. A.; Chen, X. M.; Gabor, A.; Higgins, C.; Sun, L.; Mack, C. A. Line-edge roughness performance targets for EUV lithography. In *Proceedings of SPIE 10143, Extreme Ultraviolet (EUV) Lithography VIII*, San Jose, CA, USA, 2017, p 101430E.
[26] Uchida, K.; Saitoh, M.; Kobayashi, S. Carrier transport and stress engineering in advanced nanoscale transistors from (100) and (110) transistors to carbon nanotube FETs and beyond. In *2008 IEEE International Electron Devices Meeting*, San Francisco, CA, USA, 2008, pp 1–4.
[27] Choi, J. H.; Park, Y. J.; Min, H. S. Electron mobility behavior in extremely thin SOI MOSFET's. *IEEE Electron Device Lett.* **1995**, *16*, 527–529.
[28] Gamiz, F.; Lopez-Villanueva, J. A.; Roldan, J. B.; Carceller, J. E.; Cartujo, P. Monte Carlo simulation of electron transport properties in extremely thin SOI MOSFET's. *IEEE Trans. Electron Devices* **1998**, *45*, 1122–1126.
[29] Doris, B.; Ieong, M.; Zhu, T.; Zhang, Y.; Steen, M.; Natzle, W.; Callegari, S.; Narayanan, V.; Cai, J.; Ku, S. H. et al. Device design considerations for ultra-thin SOI MOSFETs. In *IEEE International Electron Devices Meeting 2003*, Washington, DC, USA, 2003, pp 27.3.1–27.3.4.
[30] Chau, R.; Datta, S.; Doczy, M.; Doyle, B.; Jin, B.; Kavalieros, J.; Majumdar, A.; Metz, M.; Radosavljevic, M. Benchmarking nano-technology for high-performance and low-power logic transistor applications. *IEEE Trans. Nanotechnol.* **2005**, *4*, 153–158.
[31] Franklin, A. D.; Luisier, M.; Han, S. J.; Tulevski, G.; Breslin, C. M.; Gignac, L.; Lundstrom, M. S.; Haensch, W. Sub-10 nm carbon nanotube transistor. *Nano Lett.* **2012**, *12*, 758–762.
[32] Qiu, C. G.; Zhang, Z. Y.; Xiao, M. M.; Yang, Y. J.; Zhong, D. L.; Peng, L. M. Scaling carbon nanotube complementary transistors to 5-nm gate lengths. *Science* **2017**, *355*, 271–276.
[33] Cao, Q.; Tersoff, J.; Farmer, D. B.; Zhu, Y.; Han, S. J. Carbon nanotube transistors scaled to a 40-nanometer footprint. *Science* **2017**, *356*, 1369–1372.
[34] Li, M.; Yeo, K. H.; Suk, S. D.; Yeoh, Y. Y.; Kim, D. W.; Chung, T. Y.; Oh, K. S.; Lee, W. S. Sub-10 nm gate-all-around CMOS nanowire transistors on bulk Si substrate. In *2009 Symposium on VLSI Technology*, Kyoto, Japan, 2009, pp 94–95.
[35] Yu, B.; Chang, L.; Ahmed, S.; Wang, H. H.; Bell, S.; Yang, C. Y.; Tabery, C.; Ho, C.; Xiang, Q.; King, T. J. et al. FinFET scaling to 10 nm gate length. In *Digest. International Electron Devices Meeting*, San Francisco, CA, USA, 2002, pp 251–254.
[36] Franklin, A. D.; Koswatta, S. O.; Farmer, D. B.; Smith, J. T.; Gignac, L.; Breslin, C. M.; Han, S. J.; Tulevski, G. S.; Miyazoe, H.; Haensch, W. et al. Carbon nanotube complementary wrap-gate transistors. *Nano Lett.* **2013**, *13*, 2490–2495.
[37] Farmer, D. B.; Gordon, R. G. Atomic layer deposition on suspended single-walled carbon nanotubes via gas-phase noncovalent functionalization. *Nano Lett.* **2006**, *6*, 699–703.
[38] Kwak, I.; Kavrik, M.; Park, J. H.; Grissom, L.; Fruhberger, B.; Wong, K. T.; Kang, S.; Kummel, A. C. Low interface trap density in scaled bilayer gate oxides on 2D materials via nanofog low temperature atomic layer deposition. *Appl. Surf. Sci.* **2019**, *463*, 758–766.
[39] Pitner, G.; Zhang, Z.; Lin, Q.; Su, S. K.; Gilardi, C.; Kuo, C.; Kashyap, H.; Weiss, T.; Yu, Z.; Chao, T. A. et al. Sub-0.5 nm interfacial dielectric enables superior electrostatics: 65 mV/dec top-gated carbon nanotube FETs at 15 nm gate length. In *2020 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2020, pp 3.5.1–3.5.4.
[40] Lee, C. S.; Pop, E.; Franklin, A. D.; Haensch, W.; Wong, H. S. P. A Compact virtual-source model for carbon nanotube FETs in the Sub-10-nm regime—part II: Extrinsic elements, performance assessment, and design optimization. *IEEE Trans. Electron Devices* **2015**, *62*, 3070–3078.
[41] Franklin, A. D.; Han, S. J.; Tulevski, G. S.; Luisier, M.; Breslin, C. M.; Gignac, L.; Lundstrom, M. S.; Haensch, W. Sub-10 nm carbon nanotube transistor. In *2011 International Electron Devices Meeting*, Washington, DC, USA, 2011, pp 23.7.1–23.7.3.
[42] Zhang, Z.; Koswatta, S. O.; Bedell, S. W.; Baraskar, A.; Guillorn, M.; Engelmann, S. U.; Zhu, Y.; Gonsalves, J.; Pyzyna, A.; Hopstaken, M. et al. Ultra low contact resistivities for CMOS beyond 10-nm node. *IEEE Electron Device Lett.* **2013**, *34*, 723–725.
[43] Waldron, N.; Kim, D. H.; del Alamo, J. A. A self-aligned InGaAs HEMT architecture for logic applications. *IEEE Trans. Electron Devices* **2010**, *57*, 297–304.
[44] Berger, H. H. Models for contacts to planar devices. *Solid State Electron.* **1972**, *15*, 145–158.
[45] Solomon, P. M. Contact resistance to a one-dimensional quasi-ballistic nanotube/wire. *IEEE Electron Device Lett.* **2011**, *32*, 246–248.
[46] Cao, Q.; Han, S. J.; Tersoff, J.; Franklin, A. D.; Zhu, Y.; Zhang, Z.; Tulevski, G. S.; Tang, J. S.; Haensch, W. End-bonded contacts for carbon nanotube transistors with low, size-independent resistance. *Science* **2015**, *350*, 68–72.
[47] Tang, J. S.; Cao, Q.; Farmer, D. B.; Tulevski, G.; Han, S. J. Carbon nanotube complementary logic with low-temperature processed end-bonded metal contacts. In *2016 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2016, pp 5.1.1–5.1.4.
[48] Franklin, A. D.; Chen, Z. H. Length scaling of carbon nanotube transistors. *Nat. Nanotechnol.* **2010**, *5*, 858–862.
[49] Franklin, A. D.; Farmer, D. B.; Haensch, W. Defining and overcoming the contact resistance challenge in scaled carbon nanotube transistors. *ACS Nano* **2014**, *8*, 7333–7339.
[50] Pitner, G.; Hills, G.; Llinas, J. P.; Persson, K. M.; Park, R.; Bokor, J.; Mitra, S.; Wong, H. S. P. Low-temperature side contact to carbon nanotube transistors: Resistance distributions down to 10 nm contact length. *Nano Lett.* **2019**, *19*, 1083–1089.
[51] Yeap, G.; Lin, S. S.; Chen, Y. M.; Shang, H. L.; Wang, P. W.; Lin, H. C.; Peng, Y. C.; Sheu, J. Y.; Wang, M.; Chen, X. et al. 5nm CMOS production technology platform featuring full-fledged EUV, and high mobility channel FinFETs with densest 0.021μm^2 SRAM cells for mobile SoC and high performance computing applications. In *2019 IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2019, pp 36.7.1–36.7.4.
[52] Cao, Q.; Han, S. J. Single-walled carbon nanotubes for high-performance electronics. *Nanoscale* **2013**, *5*, 8852–8863.
[53] Tulevski, G. S.; Franklin, A. D.; Frank, D.; Lobez, J. M.; Cao, Q.; Park, H.; Afzali, A.; Han, S. J.; Hannon, J. B.; Haensch, W. Toward high-performance digital logic technology with carbon nanotubes. *ACS Nano* **2014**, *8*, 8730–8745.
[54] Wang, J. T.; Jin, X.; Liu, Z. B.; Yu, G.; Ji, Q. Q.; Wei, H. M.; Zhang, J.; Zhang, K.; Li, D. Q.; Yuan, Z. et al. Growing highly pure semiconducting carbon nanotubes by electrotwisting the helicity. *Nat. Catal.* **2018**, *1*, 326–331.
[55] Yang, F.; Wang, X.; Zhang, D. Q.; Yang, J.; Luo, D.; Xu, Z. W.; Wei, J.; K. Wang, J. Q.; Xu, Z.; Peng, F. et al. Chirality-specific growth of single-walled carbon nanotubes on solid alloy catalysts. *Nature* **2014**, *510*, 522–524.
[56] Zhang, S. C.; Kang, L. X.; Wang, X.; Tong, L. M.; Yang, L. W.; Wang, Z. Q.; Qi, K.; Deng, S. B.; Li, Q. W.; Bai, X. D. et al. Arrays of horizontal carbon nanotubes of controlled chirality grown using designed catalysts. *Nature* **2017**, *543*, 234–238.
[57] Franklin, A. D. The road to carbon nanotube transistors. *Nature* **2013**, *498*, 443–444.
[58] Arnold, M. S.; Green, A. A.; Hulvat, J. F.; Stupp, S. I.; Hersam, M. C. Sorting carbon nanotubes by electronic structure using density differentiation. *Nat. Nanotechnol.* **2006**, *1*, 60–65.
[59] Ghosh, S.; Bachilo, S. M.; Weisman, R. B. Advanced sorting of single-walled carbon nanotubes by nonlinear density-gradient ultracentrifugation. *Nat. Nanotechnol.* **2010**, *5*, 443–450.
[60] Green, A. A.; Hersam, M. C. Nearly single-chirality single-walled

carbon nanotubes produced via orthogonal iterative density gradient ultracentrifugation. *Adv. Mater.* **2011**, *23*, 2185–2190.
[61] Tulevski, G. S.; Franklin, A. D.; Afzali, A. High purity isolation and quantification of semiconducting carbon nanotubes via column chromatography. *ACS Nano* **2013**, *7*, 2971–2976.
[62] Nish, A.; Hwang, J. Y.; Doig, J.; Nicholas, R. J. Highly selective dispersion of single-walled carbon nanotubes using aromatic polymers. *Nat. Nanotechnol.* **2007**, *2*, 640–646.
[63] Gu, J. T.; Han, J.; Liu, D.; Yu, X. Q.; Kang, L. X.; Qiu, S.; Jin, H. H.; Li, H. B.; Li, Q. W.; Zhang, J. Solution-processable high-purity semiconducting SWCNTs for large-area fabrication of high-performance thin-film transistors. *Small* **2016**, *12*, 4993–4999.
[64] Lee, H. W.; Yoon, Y.; Park, S.; Oh, J. H.; Hong, S.; Liyanage, L. S.; Wang, H. L.; Morishita, S.; Patil, N.; Park, Y. J. et al. Selective dispersion of high purity semiconducting single-walled carbon nanotubes with regioregular poly(3-alkylthiophene)s. *Nat. Commun.* **2011**, *2*, 541.
[65] Tu, X. M.; Manohar, S.; Jagota, A.; Zheng, M. DNA sequence motifs for structure-specific recognition and separation of carbon nanotubes. *Nature* **2009**, *460*, 250–253.
[66] Samanta, S. K.; Fritsch, M.; Scherf, U.; Gomulya, W.; Bisri, S. Z.; Loi, M. A. Conjugated polymer-assisted dispersion of single-wall carbon nanotubes: The power of polymer wrapping. *Acc. Chem. Res.* **2014**, *47*, 2446–2456.
[67] Lei, T.; Pochorovski, I.; Bao, Z. N. Separation of semiconducting carbon nanotubes for flexible and stretchable electronics using polymer removable method. *Acc. Chem. Res.* **2017**, *50*, 1096–1104.
[68] Wang, C. J.; Cao, Q.; Ozel, T.; Gaur, A.; Rogers, J. A.; Shim, M. Electronically selective chemical functionalization of carbon nanotubes: Correlation between Raman spectral and electrical responses. *J. Am. Chem. Soc.* **2005**, *127*, 11460–11468.
[69] Gui, H.; Streit, J. K.; Fagan, J. A.; Hight Walker, A. R.; Zhou, C. W.; Zheng, M. Redox sorting of carbon nanotubes. *Nano Lett.* **2015**, *15*, 1642–1646.
[70] Wang, J.; Nguyen, T. D.; Cao, Q.; Wang, Y. L.; Tan, M. Y. C.; Chan-Park, M. B. Selective surface charge sign reversal on metallic carbon nanotubes for facile ultrahigh purity nanotube sorting. *ACS Nano* **2016**, *10*, 3222–3232.
[71] Lefebvre, J.; Ding, J. F.; Li, Z.; Finnie, P.; Lopinski, G.; Malenfant, P. R. L. High-purity semiconducting single-walled carbon nanotubes: A key enabling material in emerging electronics. *Acc. Chem. Res.* **2017**, *50*, 2479–2486.
[72] Lei, T.; Shao, L. L.; Zheng, Y. Q.; Pitner, G.; Fang, G. H.; Zhu, C. X.; Li, S. C.; Beausoleil, R.; Wong, H. S. P.; Huang, T. C. et al. Low-voltage high-performance flexible digital and analog circuits based on ultrahigh-purity semiconducting carbon nanotubes. *Nat. Commun.* **2019**, *10*, 2161.
[73] Liu, L. J.; Han, J.; Xu, L.; Zhou, J. S.; Zhao, C. Y.; Ding, S. J.; Shi, H. W.; Xiao, M. M.; Ding, L.; Ma, Z. et al. Aligned, high-density semiconducting carbon nanotube arrays for high-performance electronics. *Science* **2020**, *368*, 850–856.
[74] Zhang, G. Y.; Qi, P. F.; Wang, X. R.; Lu, Y. R.; Li, X. L.; Tu, R. Y.; Bangsaruntip, S.; Mann, D.; Zhang, L.; Dai, H. J. Selective etching of metallic carbon nanotubes by gas-phase reaction. *Science* **2006**, *314*, 974–977.
[75] Song, J. Z.; Li, Y. H.; Du, F.; Xie, X.; Huang, Y. G.; Rogers, J. A. Thermal analysis for laser selective removal of metallic single-walled carbon nanotubes. *J. Appl. Phys.* **2015**, *117*, 165102.
[76] Shulaker, M. M.; Van Rethy, J.; Wu, T. F.; Suriyasena Liyanage, L.; Wei, H.; Li, Z. Y.; Pop, E.; Gielen, G.; Wong, H. S. P.; Mitra, S. Carbon nanotube circuit integration up to sub-20 nm channel lengths. *ACS Nano* **2014**, *8*, 3434–3443.
[77] Jin, S. H.; Dunham, S. N.; Song, J. Z.; Xie, X.; Kim, J. H.; Lu, C. F.; Islam, A.; Du, F.; Kim, J.; Felts, J. et al. Using nanoscale thermocapillary flows to create arrays of purely semiconducting single-walled carbon nanotubes. *Nat. Nanotechnol.* **2013**, *8*, 347–355.
[78] Jie, Z.; Patil, N.; Hazeghi, A.; Mitra, S. Carbon Nanotube circuits in the presence of carbon nanotube density variations. In *2009 46th ACM/IEEE Design Automation Conference*, San Francisco, CA, USA, 2009, pp 71–76.
[79] Park, H.; Afzali, A.; Han, S. J.; Tulevski, G. S.; Franklin, A. D.; Tersoff, J.; Hannon, J. B.; Haensch, W. High-density integration of carbon nanotubes via chemical self-assembly. *Nat. Nanotechnol.* **2012**, *7*, 787–791.
[80] Kumar, B.; Falk, A. L.; Afzali, A.; Tulevski, G. S.; Oida, S.; Han, S. J.; Hannon, J. B. Spatially selective, high-density placement of polyfluorene-sorted semiconducting carbon nanotubes in organic solvents. *ACS Nano* **2017**, *11*, 7697–7701.
[81] Jinkins, K. R.; Chan, J.; Jacobberger, R. M.; Berson, A.; Arnold, M. S. Substrate-Wide confined shear alignment of carbon nanotubes for thin film transistors. *Adv. Electron. Mater.* **2019**, *5*, 1800593.
[82] Tao, A. R.; Huang, J. X.; Yang, P. D. Langmuir–blodgettry of nanocrystals and nanowires. *Acc. Chem. Res.* **2008**, *41*, 1662–1673.
[83] Li, X. L.; Zhang, L.; Wang, X. R.; Shimoyama, I.; Sun, X. M.; Seo, W. S.; Dai, H. J. Langmuir−blodgett assembly of densely aligned single-walled carbon nanotubes from bulk materials. *J. Am. Chem. Soc.* **2007**, *129*, 4890–4891.
[84] Cao, Q.; Han, S. J.; Tulevski, G. S.; Zhu, Y.; Lu, D. D.; Haensch, W. Arrays of single-walled carbon nanotubes with full surface coverage for high-performance electronics. *Nat. Nanotechnol.* **2013**, *8*, 180–186.
[85] Perebeinos, V.; Tersoff, J. Wetting transition for carbon nanotube arrays under metal contacts. *Phys. Rev. Lett.* **2015**, *114*, 085501.
[86] Engel, M.; Small, J. P.; Steiner, M.; Freitag, M.; Green, A. A.; Hersam, M. C.; Avouris, P. Thin film nanotube transistors based on self-assembled, aligned, semiconducting carbon nanotube arrays. *ACS Nano* **2008**, *2*, 2445–2452.
[87] Joo, Y.; Brady, G. J.; Arnold, M. S.; Gopalan, P. Dose-controlled, floating evaporative self-assembly and alignment of semiconducting carbon nanotubes from organic solvents. *Langmuir* **2014**, *30*, 3460–3466.
[88] Brady, G. J.; Way, A. J.; Safron, N. S.; Evensen, H. T.; Gopalan, P.; Arnold, M. S. Quasi-ballistic carbon nanotube array transistors with current density exceeding Si and GaAs. *Sci. Adv.* **2016**, *2*, e1601240.
[89] Jinkins, K. R.; Chan, J.; Brady, G. J.; Gronski, K. K.; Gopalan, P.; Evensen, H. T.; Berson, A.; Arnold, M. S. Nanotube alignment mechanism in floating evaporative self-assembly. *Langmuir* **2017**, *33*, 13407– 13414.
[90] Han, S. P.; Maune, H. T.; Barish, R. D.; Bockrath, M.; Goddard, W. A. DNA-linker-induced surface assembly of ultra dense parallel single walled carbon nanotube arrays. *Nano Lett.* **2012**, *12*, 1129– 1135.
[91] Zhao, M. Y.; Chen, Y. H.; Wang, K. X.; Zhang, Z. X.; Streit, J. K.; Fagan, J. A.; Tang, J. S.; Zheng, M.; Yang, C. Y.; Zhu, Z. et al. DNA-directed nanofabrication of high-performance carbon nanotube field-effect transistors. *Science* **2020**, *368*, 878–881.
[92] Sun, W.; Shen, J.; Zhao, Z.; Arellano, N.; Rettner, C.; Tang, J. S.; Cao, T. Y.; Zhou, Z. Y.; Ta, T.; Streit, J. K. et al. Precise pitch-scaling of carbon nanotube arrays within three-dimensional DNA nanotrenches. *Science* **2020**, *368*, 874–877.
[93] Zhang, Z. Y.; Liang, X. L.; Wang, S.; Yao, K.; Hu, Y. F.; Zhu, Y. Z.; Chen, Q.; Zhou, W. W.; Li, Y.; Yao, Y. G. et al. Doping-free fabrication of carbon nanotube based ballistic CMOS devices and circuits. *Nano Lett.* **2007**, *7*, 3603–3607.
[94] Shahrjerdi, D.; Franklin, A. D.; Oida, S.; Tulevski, G. S.; Han, S. J.; Hannon, J. B.; Haensch, W. High device yield carbon nanotube NFETs for high-performance logic applications. In *2011 International Electron Devices Meeting*, Washington, DC, USA, 2011, pp 23.3.1–23.3.4.
[95] Ding, L.; Wang, S.; Zhang, Z. Y.; Zeng, Q. S.; Wang, Z. X.; Pei, T.; Yang, L. J.; Liang, X. L.; Shen, J.; Chen, Q. et al. Y-contacted high-performance n-type single-walled carbon nanotube field-effect transistors: Scaling and comparison with Sc-contacted devices. *Nano Lett.* **2009**, *9*, 4209–4214.
[96] Wang, C.; Ryu, K.; Badmaev, A.; Zhang, J. L.; Zhou, C. W. Metal contact engineering and registration-free fabrication of complementary metal-oxide semiconductor integrated circuits using aligned carbon nanotubes. *ACS Nano* **2011**, *5*, 1147–1153.
[97] Ding, L.; Zhang, Z. Y.; Liang, S. B.; Pei, T.; Wang, S.; Li, Y.; Zhou, W. W.; Liu, J.; Peng, L. M. CMOS-based carbon nanotube pass-transistor logic integrated circuits. *Nat. Commun.* **2012**, *3*, 677.
[98] Tang, J. S.; Farmer, D.; Bangsaruntip, S.; Chiu, K. C.; Kumar, B.; Han, S. J. Contact engineering and channel doping for robust carbon nanotube NFETs. In *2017 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA)*, Hsinchu, China, 2017, pp 1–2.

[99] Han, S. J.; Tang, J. S.; Kumar, B.; Falk, A.; Farmer, D.; Tulevski, G.; Jenkins, K.; Afzali, A.; Oida, S.; Ott, J. et al. High-speed logic integrated circuits with solution-processed self-assembled carbon nanotubes. *Nat. Nanotechnol.* **2017**, *12*, 861–865.

[100] Ha, T. J.; Chen, K.; Chuang, S.; Yu, K. M.; Kiriya, D.; Javey, A. Highly uniform and stable n-type carbon nanotube transistors by using positively charged silicon nitride thin films. *Nano Lett.* **2015**, *15*, 392–397.

[101] Tang, J. S.; Cao, Q.; Tulevski, G.; Jenkins, K. A.; Nela, L.; Farmer, D. B.; Han, S. J. Flexible CMOS integrated circuits based on carbon nanotubes with sub-10 ns stage delays. *Nat. Electron.* **2018**, *1*, 191–196.

[102] Lau, C.; Srimani, T.; Bishop, M. D.; Hills, G.; Shulaker, M. M. Tunable *n*-type doping of carbon nanotubes through engineered atomic layer deposition HfO_x films. *ACS Nano* **2018**, *12*, 10924–10931.

[103] Hills, G.; Lau, C.; Wright, A.; Fuller, S.; Bishop, M. D.; Srimani, T.; Kanhaiya, P.; Ho, R.; Amer, A.; Stein, Y. et al. Modern microprocessor built from complementary carbon nanotube transistors. *Nature* **2019**, *572*, 595–602.

[104] Kanhaiya, P. S.; Lau, C.; Hills, G.; Bishop, M. D.; Shulaker, M. M. Carbon Nanotube-based CMOS SRAM: 1 kbit 6T SRAM arrays and 10T SRAM cells. *IEEE Trans. Electron Devices* **2019**, *66*, 5375–5380.

[105] Kita, K.; Toriumi, A. Intrinsic origin of electric dipoles formed at high-k/SiO_2 interface. In *2008 IEEE International Electron Devices Meeting*, San Francisco, CA, USA, 2008, pp 1–4.

[106] Kita, K.; Toriumi, A. Origin of electric dipoles formed at high-*k*/SiO_2 interface. *Appl. Phys. Lett.* **2009**, *94*, 132902.

[107] McMahon, W.; Tian, C.; Uppal, S.; Kothari, H.; Jin, M.; LaRosa, G.; Nigam, T.; Kerber, A.; Linder, B. P.; Cartier, E. et al. Intrinsic dielectric stack reliability of a high performance bulk planar 20nm replacement gate high-k metal gate technology and comparison to 28nm gate first high-k metal gate process. In *2013 IEEE International Reliability Physics Symposium (IRPS)*, Monterey, CA, USA, 2013, pp 4C.4.1–4C.4.4.

[108] Noyce, S. G.; Doherty, J. L.; Cheng, Z. H.; Han, H.; Bowen, S.; Franklin, A. D. Electronic stability of carbon nanotube transistors under long-term bias stress. *Nano Lett.* **2019**, *19*, 1460–1466.

[109] Hu, Z. Y.; Tulevski, G. S.; Hannon, J. B.; Afzali, A.; Liehr, M.; Park, H. Variability and reliability analysis in self-assembled multichannel carbon nanotube field-effect transistors. *Appl. Phys. Lett.* **2015**, *106*, 243106.

[110] Cao, Q.; Han, S. J.; Penumatcha, A. V.; Frank, M. M.; Tulevski, G. S.; Tersoff, J.; Haensch, W. E. Origins and characteristics of the threshold voltage variability of quasiballistic single-walled carbon nanotube field-effect transistors. *ACS Nano* **2015**, *9*, 1936–1944.

[111] Cao, Q.; Tersoff, J.; Han, S. J.; Penumatcha, A. V. Scaling of device variability and subthreshold swing in ballistic carbon nanotube transistors. *Phys. Rev. Appl.* **2015**, *4*, 024022.

[112] Franklin, A. D.; Tulevski, G. S.; Han, S. J.; Shahrjerdi, D.; Cao, Q.; Chen, H. Y.; Wong, H. S. P.; Haensch, W. Variability in carbon nanotube transistors: Improving device-to-device consistency. *ACS Nano* **2012**, *6*, 1109–1115.

[113] Osburn, C.; Berger, H.; Donovan, R.; Jones, G. The effects of contamination on semiconductor manufacturing yield. *J. IEST* **1988**, *31*, 45–57.

[114] Gorodokin, V.; Zemlyanov, D. Metallic contamination in silicon processing. In *2004 23rd IEEE Convention of Electrical and Electronics Engineers in Israel*, Tel-Aviv, Israel, 2004, pp 157–160.

[115] Bishop, M. D.; Hills, G.; Srimani, T.; Lau, C.; Murphy, D.; Fuller, S.; Humes, J.; Ratkovich, A.; Nelson, M.; Shulaker, M. M. Fabrication of carbon nanotube field-effect transistors in commercial silicon manufacturing facilities. *Nat. Electron.* **2020**, *3*, 492–501.

[116] Shulaker, M. M.; Hills, G.; Park, R. S.; Howe, R. T.; Saraswat, K.; Wong, H. S. P.; Mitra, S. Three-dimensional integration of nanotechnologies for computing and data storage on a single chip. *Nature* **2017**, *547*, 74–78.

[117] Shulaker, M. M.; Wu, T. F.; Pal, A.; Zhao, L.; Nishi, Y.; Saraswat, K.; Wong, H. P.; Mitra, S. Monolithic 3D integration of logic and memory: Carbon nanotube FETs, resistive RAM, and silicon FETs. In *2014 IEEE International Electron Devices Meeting*, San Francisco, CA, USA, 2014, pp 27.4.1–27.4.4.

[118] Liu, Y.; Wang, S.; Liu, H. P.; Peng, L. M. Carbon nanotube-based three-dimensional monolithic optoelectronic integrated system. *Nat. Commun.* **2017**, *8*, 15649.

[119] Batude, P.; Vinet, M.; Previtali, B.; Tabone, C.; Xu, C.; Mazurier, J.; Weber, O.; Andrieu, F.; Tosti, L.; Brevard, L. et al. Advances, challenges and opportunities in 3D CMOS sequential integration. In *2011 International Electron Devices Meeting*, Washington, DC, USA, 2011, pp 7.3.1–7.3.4.

[120] Vinet, M.; Batude, P.; Fenouillet-Beranger, C.; Clermidy, F.; Brunet, L.; Rozeau, O.; Hartmannn, J. M. ; Billoint, O.; Cibrario, G.; Previtali, B. et al. Monolithic 3D integration: A powerful alternative to classical 2D scaling. In *SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, Millbrae, CA, USA, 2014, pp 1–3.

[121] Aly, M. M. S.; Wu, T. F.; Bartolo, A.; Malviya, Y. H.; Hwang, W.; Hills, G.; Markov, I.; Wootters, M.; Shulaker, M. M.; Wong, H. S. P. et al. The N3XT approach to energy-efficient abundant-data computing. *Proc. IEEE* **2019**, *107*, 19–48.

[122] Cao, Q.; Kim, H. S.; Pimparkar, N.; Kulkarni, J. P.; Wang, C. J.; Shim, M.; Roy, K.; Alam, M. A.; Rogers, J. A. Medium-scale carbon nanotube thin-film integrated circuits on flexible plastic substrates. *Nature* **2008**, *454*, 495–500.

[123] Hu, Y. F.; Peng, L. M.; Xiang, L.; Zhang, H. Flexible integrated circuits based on carbon nanotubes. *Acc. Mater. Res.* **2020**, *1*, 88–99.

[124] Chen, H. T.; Cao, Y.; Zhang, J. L.; Zhou, C. W. Large-scale complementary macroelectronics using hybrid integration of carbon nanotubes and IGZO thin-film transistors. *Nat. Commun.* **2014**, *5*, 4097.

[125] Zhu, C. X.; Chortos, A.; Wang, Y.; Pfattner, R.; Lei, T.; Hinckley, A. C.; Pochorovski, I.; Yan, X. Z.; To, J. W. F.; Oh, J. Y. et al. Stretchable temperature-sensing circuits with strain suppression based on carbon nanotube transistors. *Nat. Electron.* **2018**, *1*, 183–190.

[126] Nela, L.; Tang, J. S.; Cao, Q.; Tulevski, G.; Han, S. J. Large-area high-performance flexible pressure sensor with carbon nanotube active matrix for electronic skin. *Nano Lett.* **2018**, *18*, 2054–2059.

[127] Cao, Q. Better radio-frequency transistors with nanotubes. *Nat. Electron.* **2019**, *2*, 495–496.

[128] Rutherglen, C.; Jain, D.; Burke, P. Nanotube electronics for radiofrequency applications. *Nat. Nanotechnol.* **2009**, *4*, 811–819.

[129] Che, Y. C.; Lin, Y. C.; Kim, P.; Zhou, C. W. T-gate aligned nanotube radio frequency transistors and circuits with superior performance. *ACS Nano* **2013**, *7*, 4343–4350.

[130] Zhong, D. L.; Shi, H. W.; Ding, L.; Zhao, C. Y.; Liu, J. X.; Zhou, J. S.; Zhang, Z. Y.; Peng, L. M. Carbon nanotube film-based radio frequency transistors with maximum oscillation frequency above 100 GHz. *ACS Appl. Mater. Interfaces* **2019**, *11*, 42496–42503.

[131] Cao, Y.; Che, Y. C.; Seo, J. W. T.; Gui, H.; Hersam, M. C.; Zhou, C. W. High-performance radio frequency transistors based on diameter-separated semiconducting carbon nanotubes. *Appl. Phys. Lett.* **2016**, *108*, 233105.

[132] Wang, Z. X.; Liang, S. B.; Zhang, Z. Y.; Liu, H. G.; Zhong, H.; Ye, L. H.; Wang, S.; Zhou, W. W.; Liu, J.; Chen, Y. B. et al. Scalable fabrication of ambipolar transistors and radio-frequency circuits using aligned carbon nanotube arrays. *Adv. Mater.* **2014**, *26*, 645–652.

[133] Kocabas, C.; Dunham, S.; Cao, Q.; Cimino, K.; Ho, X.; Kim, H. S.; Dawson, D.; Payne, J.; Stuenkel, M.; Zhang, H. et al. High-frequency performance of submicrometer transistors that use aligned arrays of single-walled carbon nanotubes. *Nano Lett.* **2009**, *9*, 1937–1943.

[134] Rutherglen, C.; Kane, A. A.; Marsh, P. F.; Cain, T. A.; Hassan, B. I.; AlShareef, M. R.; Zhou, C. W.; Galatsis, K. Wafer-scalable, aligned carbon nanotube transistors operating at frequencies of over 100 GHz. *Nat. Electron.* **2019**, *2*, 530–539.

[135] Kocabas, C.; Kim, H. S.; Banks, T.; Rogers, J. A.; Pesetski, A. A.; Baumgardner, J. E.; Krishnaswamy, S. V.; Zhang, H. Radio frequency analog electronics based on carbon nanotube transistors. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 1405–1409.

[136] Hároz, E. H.; Duque, J. G.; Tu, X. M.; Zheng, M.; Hight Walker, A. R.; Hauge, R. H.; Doorn, S. K.; Kono, J. Fundamental optical processes in armchair carbon nanotubes. *Nanoscale* **2013**, *5*, 1411–1439.