Technology Projection Using Simple Compact Models

H.-S. Philip Wong, Lan Wei, Saeroonter Oh, Albert Lin, Jie Deng, Soogine Chong, Kerem Akarvardar
Department of Electrical Engineering and Center for Integrated Systems,
Stanford University, Stanford, CA 94305, U.S.A.

***Abstract*— We review recent efforts to capture the device non-idealities for circuit-level technology projection for Si CMOS. We also give some examples of simple compact model development for assessing the circuit-level performance of exploratory devices such as III-V FET, carbon nanotube transistor, and nanoelectromechanical (NEM) transistors and relays.**

Keywords – Technology projection; compact model; III-V FET; CNT; carbon nanotube transistor; nanoelectromechanical relay; NEMS; parasitic capacitance; parasitic resistance; Si CMOS

I. INTRODUCTION

Technology projection used to be a device-centric activity. Simple metrics such as *CV/I* (where *C* is the intrinsic gate capacitance, *V* is the power supply voltage, and *I* is the saturation on-current of a MOSFET) can predict circuit level performance and have served the industry well for a long time. As technology scaling reaches into the sub-22 nm regime and beyond, simple performance metrics are beginning to lose their effectiveness for predicting performance at the circuit level [1, 2]. Contributing to this trend is the increasing role played by the parasitic capacitance and parasitic resistance for Si CMOS. Furthermore, the transistor is never fully turned-on or fully turned-off. Therefore, the entire switching trajectory of a logic gate must be taken into account to predict circuit level speed/power performance [3, 4]. The *CV/I* metric more accurately predicts circuit performance if *C* is C_{tot} (the total gate capacitance including parasitic capacitance), and *I* is I_{eff} (the effective current that accounts for the switching trajectory of the logic gate), leading to $\tau=C_{tot}V/I_{eff}$.

When considering new devices beyond Si CMOS, for which various new channel materials and new logic switches have been proposed, compact models are required for a first order assessment of circuit level performance. As these "new" switches are based on different device physics from the Si MOSFET, these simple compact models must capture the essential device behavior to enable look-ahead technology assessment at the circuit level while maintaining a direct correlation with physical device parameters.

In this paper, we review the recent efforts to capture the device non-idealities for circuit-level technology projection for Si CMOS [5, 6, 7]. The focus is on properly modeling the parasitic elements. We will also give some examples of simple compact model development for assessing exploratory devices beyond Si CMOS such as the III-V FET, carbon nanotube transistor, and nanoelectromechanical transistor and relay. A synopsis of the learning from the use of such compact models to assess technology options is presented.

II. SI CMOS SCALING

A. *Current Trends*

While gate length scaling has been very effective in reducing the "*C*" of the *CV/I* metric in previous technology generations, as device physical gate length is reduced below 20 nm, gate length scaling becomes less effective or impractical due to several factors: (1) the inability to scale the equivalent oxide thickness (EOT) to maintain device electrostatics to control short channel effects, and (2) the increasing contribution of the parasitic capacitance and parasitic resistance. To compensate for the slowdown in gate length scaling, various carrier transport enhancing techniques based on band structure engineering using strain/stress are used [2,5]. Alternative channel materials such as Ge and III-V are explored for future generations to enlarge the band structure engineering design space. Device scaling is as much about density improvement as it is about energy-delay improvement. In the limit of zero gate length and completely ballistic transport, the device performance is determined by the parasitic capacitances [7] and parasitic resistances. Many parasitic elements do not scale down with device dimensions. Therefore the parasitic elements will become increasingly important [6] and technology projection going forward must necessarily pay close attention to the modeling of parasitic capacitances and resistances (Fig. 1).

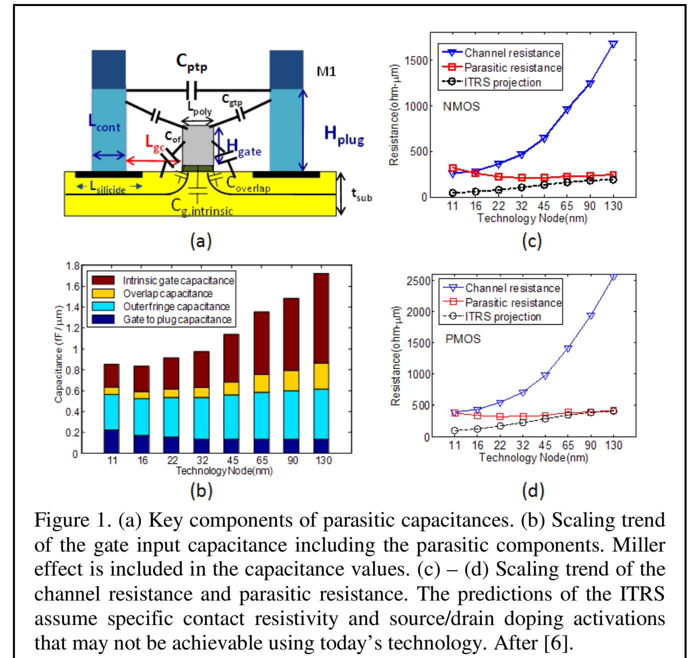


Figure 1. (a) Key components of parasitic capacitances. (b) Scaling trend of the gate input capacitance including the parasitic components. Miller effect is included in the capacitance values. (c) – (d) Scaling trend of the channel resistance and parasitic resistance. The predictions of the ITRS assume specific contact resistivity and source/drain doping activations that may not be achievable using today's technology. After [6].

B. *Selective Device Structure Scaling*

Even with the gate length remaining the same, selectively scaling the device structure aggressively will provide significant device- and circuit-level performance improvement from technology generation to technology generation. The main reason is that engineering the device structure to reduce device footprint reduces the device parasitic resistance and the interconnect length and therefore improves speed and power efficiency. This concept of selective device structure scaling [8,9] is analogous to aggressive selective scaling of the gate length introduced in the 0.35 µm era. Selectively scaling down the device structure requires a careful examination of the parasitic capacitance and parasitic resistance, because bringing device structures closer involves the opposing tradeoff of resistance vs capacitance.

As examples, consider the selective device structure scaling scenarios in Fig. 2. Reducing the contact size (Fig. 2a) and the contact to gate distance (Fig. 2d) brings the device pitch closer and reduces interconnect capacitance by reducing the interconnect length. However, this may increase the contact resistance sharply at very advanced technology nodes, if the transfer length of the contact becomes comparable to the size of the contacts. Lowering the contact plug height (Fig. 2b) reduces the source/drain node capacitance (C_{sd}) by reducing $C_{plug\text{-}plug}$, but does not significantly reduce the total gate node capacitance (C_{gg}). Lowering the gate height (Fig. 2c) is effective in reducing both C_{gg} and C_{sd} [9].

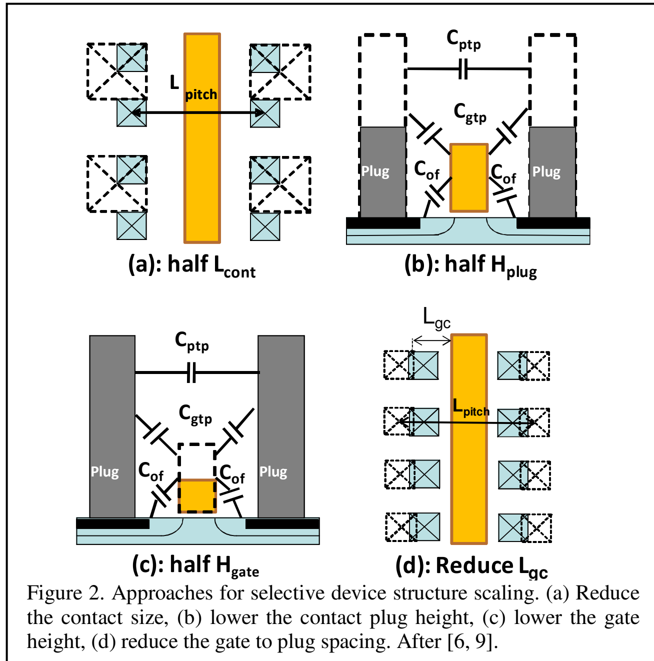


Figure 2. Approaches for selective device structure scaling. (a) Reduce the contact size, (b) lower the contact plug height, (c) lower the gate height, (d) reduce the gate to plug spacing. After [6, 9].

Khakifirooz & Antoniadis [5] examined various options for CMOS scaling into the 32 nm node, including (1) even more aggressive gate length scaling, (2) relaxed gate pitch, (3) thinner gate oxide, (4) lower parasitic resistance, (5) higher power dissipation, and (6) reduced fringing capacitance. A clear conclusion from [5] is that reducing the fringing capacitance has the strongest effect. Experimental evidence of the effectiveness of such approach is presented recently by Ren et al. [10] where it is shown that by reducing the gate height from 100 nm to 80 nm (20% reduction), the ring oscillator speed is improved by 3% (comparing matched DC device characteristics such as SCE and I_{on}-I_{off}). Simulations show a performance gain from 4% [10] to 16% [7] for a 50% reduction in gate height depending on other assumptions. Further downscaling of the gate length is difficult and selectively scaling the device structure to reduce parasitic effects is perhaps the most scalable approach and has the added benefit of reduced active power consumption [5, 8, 9]. Fig. 3 shows a possible roadmap for scaling down to 11 nm node by selective device structure scaling without scaling the gate length [9].

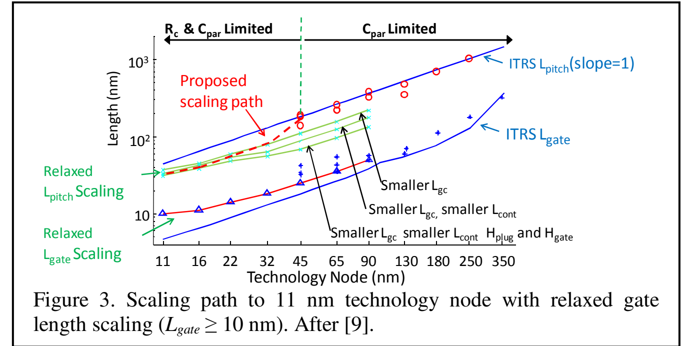


Figure 3. Scaling path to 11 nm technology node with relaxed gate length scaling ($L_{gate} \geq 10$ nm). After [9].

C. *Parasitic Capacitance Modeling*

Recognizing the importance of the parasitic capacitance, it is instructive to develop simple models of the parasitic capacitance and use the models to gain insights into device design options.

Wei et al. (Table 1 in [7]) derived simple analytical expressions for the various parasitic capacitances for a typical bulk, FDSOI, and double-gate FET. Briefly, the parallel plate capacitance model is used to model charge-sheet capacitances, such as the gate-to-channel capacitance (C_{gc}) and overlap capacitance (C_{ov}). Conformal mapping is applied for fringe type capacitances, such as outer-fringe capacitance (C_{of}), inner-fringe capacitance (C_{if}) and corner capacitance (C_{corner}) [11-12].

It is necessary to be aware of which capacitance components affect the delay. Fig. 4 illustrates an inverter stage driving an identical load stage. The capacitance load changes during the switching event as the output and input nodes of the driver and load stages make the transition from one logic level to another. The relevant time period is from the 50% input level to the 50% output level. The current from the FET charges/discharges the capacitances looking into the drain node (C_d) of the driving stage, and those looking into the gate node of the loading stage, which switches between on and off (C_{g_on} and C_{g_off}). For a first approximation, during the rise-to-fall (driver rising, load falling) transition, the nFET is always on, and the pFET is on half the time. Similarly, during the fall-to-rise transition, the nFET is on half the time and the pFET is always on. Averaging over the fall-to-rise and rise-to-fall delays, the effective load capacitance is expressed by (1). Here,

M is the Miller effect coefficient. Interconnect capacitance can be simply included as an additional term.

$$C_{tot} = C_d + \left(0.25C_{g_off} + 0.75C_{g_on}\right)\cdot FO \quad (1)$$
$$= \left(C_{ov} + C_{of} + C_{pcca} + C_{corner}\right)\cdot M + C_j + \begin{bmatrix} 0.25\left(C_{gb_off} + 2C_{ov} + 2C_{if} + 2C_{of} + 2C_{pcca} + 2C_{corner}\right) \\ +0.75\left(C_{gc} + 2C_{ov} + 2C_{of} + 2C_{pcca} + 2C_{corner}\right)\end{bmatrix}\cdot FO$$

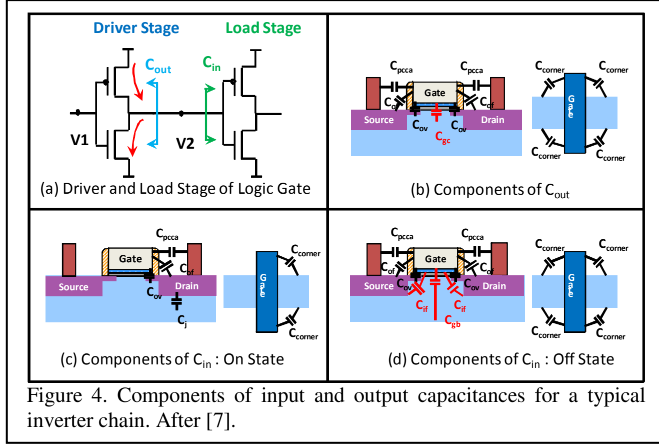


Figure 4. Components of input and output capacitances for a typical inverter chain. After [7].

As an illustration of the use of the capacitance models for technology projection, we consider two aspects of device design that are of current interest: (1) corner fringing capacitance from narrow device width, and (2) nFET/pFET relative strengths.

1) Corner Capacitance: Traditionally, the device width statistical distribution is divided into two main families: “large” width (~0.5µm) in logic cells and small width (<0.1µm) in SRAM cells. As logic cell size is scaled down, the average “large” width is becoming smaller and is now close to 0.1µm. As a consequence, the C_{corner} becomes significant. As shown in Fig. 5a, wide devices, where C_{corner} is negligible, is 35% (45nm) – 58% (11nm) faster than those with minimum widths.

2) nFET/pFET Strength: Conventionally, for static circuit design, the ratio of the width of PMOS over NMOS is about 2. There has been intense effort to boost PMOS driving capability [13-14], which can potentially reduce the PMOS device width and thus reduce the total capacitance. Fig 5b shows that 11% (11nm) – 20% (45nm) delay reduction is obtained by boosting PMOS driving capability by 2x, (i.e. same driving capability as NMOS). Boosting PMOS driving capability by 1.5x achieves 2/3 of the improvement of 2x, (7% (11nm) – 14% (45nm)). This improvement diminishes towards more advanced technology because of the disproportionally scaled parasitic capacitances.

D. One-Dimensional Channels

Beyond the planar channel, FinFET and one-dimensional channels such as semiconductor nanowires and carbon nanotubes are attractive options because of the excellent electrostatics offered by the very small body thickness and multi-dimensional gating [15, 16]. The gate delay metric $\tau = C_{tot}V/I_{eff}$ can be expanded into four terms: C_{gc}/I_{off}, I_{off}/I_{eff}, $(1+C_{par}/C_{gc})$ and V as in (2),

$$\tau = \frac{C_{tot}V}{I_{eff}} = \frac{(C_{gc}+C_{par})V}{I_{eff}} = \frac{C_{gc}}{I_{off}}\cdot\frac{I_{off}}{I_{eff}}\left(1+\frac{C_{par}}{C_{gc}}\right)\cdot V \quad (2)$$

Here, $C_{tot}=C_{gc}+C_{par}$. I_{eff} and I_{off} are the effective current and off-state current, respectively.

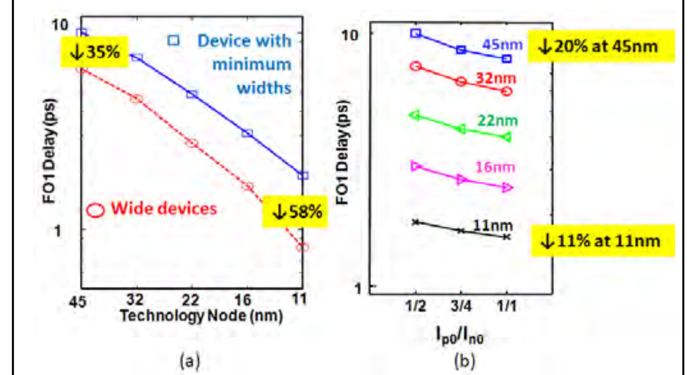


Figure 5. (a) C_{corner} is an important parasitic capacitance for narrow devices. Wide devices are much faster, especially at very advanced technology. At advanced technology nodes where devices are narrower, C_{corner} is increasingly important and contributes to 58% delay reduction for wide devices. Wide device = 1 µm width. Narrow device = minimum width for the technology. $W_p = 2W_n$. (b) Delay dependence on the ratio of pFET/nFET current drive strength. Boosting PMOS current driving capability becomes less effective in improving delay at advanced technology nodes. After [7].

Performance estimation for these one-dimensional (1-D) channels is often based on device metrics that are normalized to a single channel (e.g. a single fin, a single nanowire, or a single carbon nanotube). Normalizing the current to the physical device width introduces difficulty because of the non-planar geometry. In addition, in order to provide adequate current to drive load capacitances, it is necessary to have many channels in parallel [17]. In this case, the channel screening effect degrades the current carried per channel [18, 19]. To fairly compare among different technologies, i.e. 1D, 2D and 3D, it is necessary to compare devices for the same I_{off}/C_{gc} because C_{gc} is proportional to the channel charge at the on-state. For a constant I_{off}/C_{gc}, the gate delay τ is determined by both I_{eff}/I_{off} and $(1+C_{par}/C_{gc})$.

For cylindrical channel geometry, gate capacitance (C_{gc}) and outer fringing parasitic capacitance (C_{par}) models have been developed [19, 20]. Modeling the effect of inter-channel screening and the corner capacitance (Fig. 4 and Fig. 6(c)) are particularly important. The impact of the parasitic capacitance can be substantial especially for sparse 1D channels. Fig. 6(d) plots the $(1+C_{par}/C_{gc})$ for various 1D channel densities. The C_{gc} is the useful capacitance, while the C_{par} loads down the circuit. The delay improvement can be degraded to about 50% of the ideal, 2D assumptions for typical situations [20]. The impact of inter-channel screening suggests that the gate-all-around device structure may be a better solution even though the conduction channel density for gate-all-around devices is lower due to the need to put a gate electrode between the

channels. Further optimization studies are required to fully understand the tradeoff.

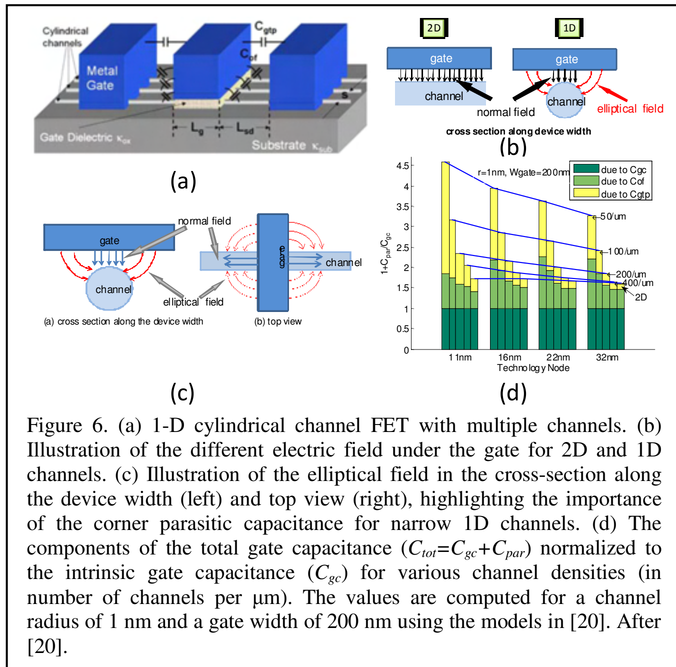


Figure 6. (a) 1-D cylindrical channel FET with multiple channels. (b) Illustration of the different electric field under the gate for 2D and 1D channels. (c) Illustration of the elliptical field in the cross-section along the device width (left) and top view (right), highlighting the importance of the corner parasitic capacitance for narrow 1D channels. (d) The components of the total gate capacitance ($C_{tot}=C_{gc}+C_{par}$) normalized to the intrinsic gate capacitance (C_{gc}) for various channel densities (in number of channels per µm). The values are computed for a channel radius of 1 nm and a gate width of 200 nm using the models in [20]. After [20].

E. III-V FET

Improving the carrier transport (mobility, injection velocity) by engineering the band structure has been the dominant means for performance gains since the 90 nm node. Strain engineering is currently the primary means for engineering the band structure [21, 22]. Beyond strain engineering, the use of alternative channel materials such as Ge and III-V compound semiconductor materials has been actively researched [23-26].

Performance estimation for III-V FET is particularly interesting and non-trivial because of the complex band structure with multiple valleys and various proposed device structures (HEMT, MOSFET-type with a surface channel, and HEMT with a gate dielectric, to name a few). Experimentally, In-compound (e.g., InGaAs, InAs) quantum well devices have demonstrated I_{ON}/I_{OFF} ratios of above 10^5, gate delay metrics (CV/I) in the sub-ps range [25], and subthreshold slope (SS) under 70mV/dec at sub-50nm gate lengths [26].

To enable digital circuit design and performance estimation for future technology nodes, a compact model of III-V FET targeted for digital application is required. Compact modeling in the III-V community has been mostly oriented towards RF analog applications [27]. For digital applications, the compact model needs to account for effects such as: field- and spatially-confined subband energy levels, short channel effects, intrinsic capacitances with dependence on source and drain fields, and parasitic capacitance and parasitic resistance for devices with small device footprint. Since most experimental III-V FET devices demonstrated to date have relatively large footprint, modeling the parasitic components for a III-V FET that has the same footprint as state-of-the-art silicon CMOS is extremely important for performance estimation.

To illustrate the importance of the device footprint, a simple approach is taken in [28]. In [28], experimental *I-V* characteristics of a micron-scale III-V HEMT (with device footprint of about 80λ) was fitted to a simple IV model in a TCAD tool by adjusting the carrier transport parameters. Then a hypothetical device with a device footprint appropriate for digital CMOS at an advanced technology node (~12λ with the same gate length) is constructed. The appropriate parasitic elements (capacitances, resistances) are automatically included in the TCAD model. Mixed-mode TCAD simulation is used to predict inverter chain delay. This exercise highlights several important observations (Fig. 7): (1) by bringing the device footprint in line with the digital application requirement, the parasitic resistance can be reduced significantly, at the expense of an increase in the parasitic capacitance, (2) the on-current is increased by 40% and the delay is reduced by 20% for the small footprint device despite this resistance/capacitance tradeoff, (3) the experimental device suffers from a large series resistance, which will be one of the eventual performance limiters of III-V FETs.

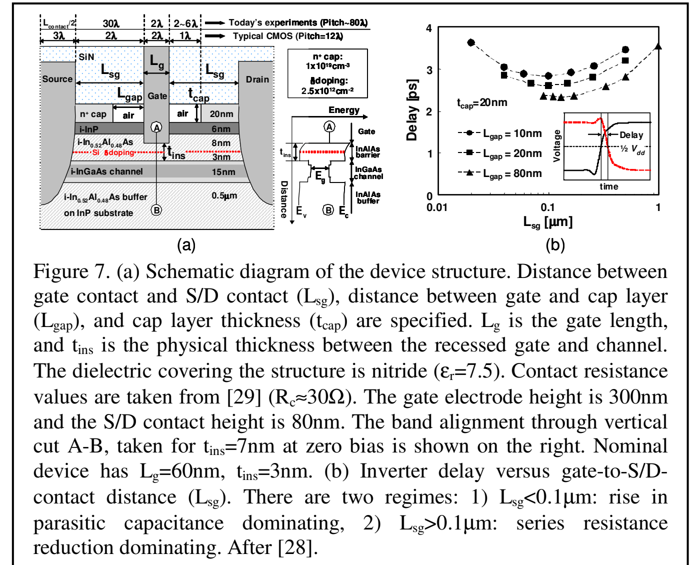


Figure 7. (a) Schematic diagram of the device structure. Distance between gate contact and S/D contact (L_{sg}), distance between gate and cap layer (L_{gap}), and cap layer thickness (t_{cap}) are specified. L_g is the gate length, and t_{ins} is the physical thickness between the recessed gate and channel. The dielectric covering the structure is nitride (ε_r=7.5). Contact resistance values are taken from [29] ($R_c\approx 30\Omega$). The gate electrode height is 300nm and the S/D contact height is 80nm. The band alignment through vertical cut A-B, taken for t_{ins}=7nm at zero bias is shown on the right. Nominal device has L_g=60nm, t_{ins}=3nm. (b) Inverter delay versus gate-to-S/D-contact distance (L_{sg}). There are two regimes: 1) L_{sg}<0.1µm: rise in parasitic capacitance dominating, 2) L_{sg}>0.1µm: series resistance reduction dominating. After [28].

In order to provide further insights into device design (e.g. choice of materials, layer structures for the III-V heterostructure channel, source/drain contact geometry, gate dielectric thickness vs short channel effect control tradeoff), a compact model that takes into account the effects outlined in the previous paragraphs will be valuable. A first attempt at such a model is presented in [30]. Briefly, the energy levels of the III-V channel, which are important for obtaining the channel charge, are obtained by an analytical fit to the numerical solution of the Schrödinger equation. The approximation must account for the finite energy barrier of the heterostructure (the infinite well approximation introduces significant errors) and the quantization of the energy levels due to the applied normal electric field and the spatial confinement. The two dimensional potential profile of the channel is important for an accurate description of the short channel effect and the intrinsic capacitances. It is modeled using the approximate Poisson solutions of the scale length theory for Si MOSFETs [31, 32]. The 2D potential profile enables the calculation of the intrinsic capacitances as a function of the

applied biases. Fig. 8 shows the agreement of model to experimental data [25] and projections of L_g=20 nm inverter chain delay using this model. The model was developed in MATLAB and subsequently ported to VerilogA that can be executed in a typical SPICE circuit simulation environment.

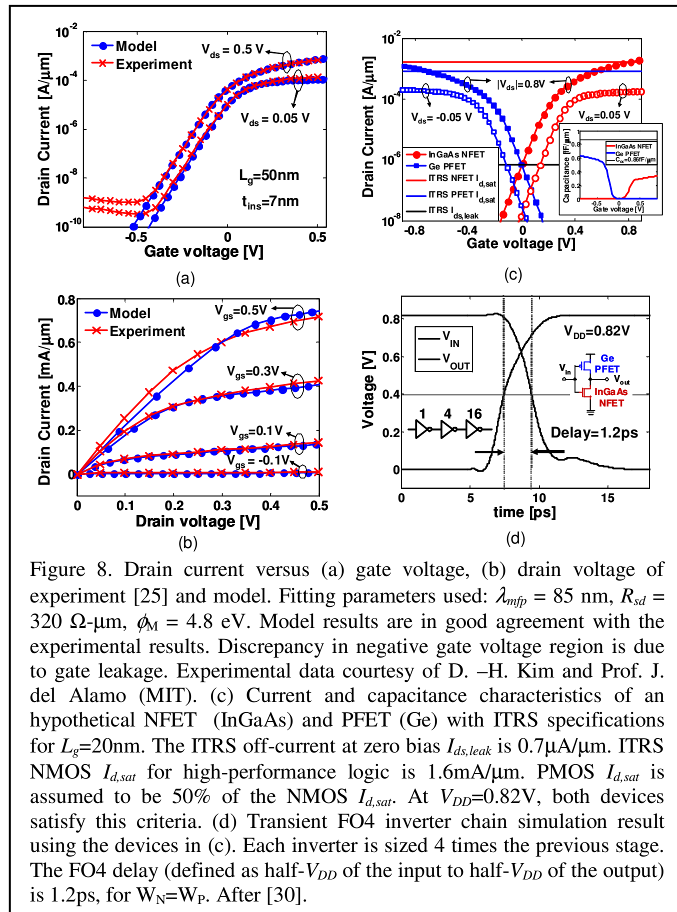


Figure 8. Drain current versus (a) gate voltage, (b) drain voltage of experiment [25] and model. Fitting parameters used: λ_{mfp} = 85 nm, R_{sd} = 320 Ω-μm, ϕ_M = 4.8 eV. Model results are in good agreement with the experimental results. Discrepancy in negative gate voltage region is due to gate leakage. Experimental data courtesy of D. –H. Kim and Prof. J. del Alamo (MIT). (c) Current and capacitance characteristics of an hypothetical NFET (InGaAs) and PFET (Ge) with ITRS specifications for L_g=20nm. The ITRS off-current at zero bias $I_{ds,leak}$ is 0.7μA/μm. ITRS NMOS $I_{d,sat}$ for high-performance logic is 1.6mA/μm. PMOS $I_{d,sat}$ is assumed to be 50% of the NMOS $I_{d,sat}$. At V_{DD}=0.82V, both devices satisfy this criteria. (d) Transient FO4 inverter chain simulation result using the devices in (c). Each inverter is sized 4 times the previous stage. The FO4 delay (defined as half-V_{DD} of the input to half-V_{DD} of the output) is 1.2ps, for W_N=W_P. After [30].

III. BEYOND SI CMOS SCALING

Beyond Si CMOS scaling, there are options that include the use of nanomaterials for the channel such as carbon nanotube (CNT) and graphene nanoribbon (GNR), and switching devices that operate on principles different from a field effect transistor. Two prominent examples of such device options are the tunnel FET [33-36], and nanoelectromechanical (NEM) transistor [37] and relay [38].

In the following subsections, we review some of the recent compact modeling work on carbon nanotube transistor and nanoelectromechanical transistor and relay, with a view to providing estimates of performance and insights for improving design design.

A. *Carbon Nanotube Transistor (CNFET)*

The carbon nanotube transistor (CNFET) is a promising extension to Si CMOS because of the excellent transport properties and the ultra-thin body of the channel which leads to excellent electrostatic control by the gate electrode [17]. In addition, design infrastructure for Si CMOS can be utilized for the CNFET because the device characteristics and layout geometries are similar to Si CMOS [39]. The initial estimates of performance based on single carbon nanotube intrinsic transport properties show that CNFET can significantly outperform silicon CMOS [40, 41] when normalized to the width of the CNT. Therefore, there is a strong interest to project the circuit level performance of the CNFET as a digital VLSI logic switch. To accomplish this goal, it is necessary to develop a compact model suitable for circuit simulation (e.g. in SPICE) that includes all the device structural non-idealities and parasitic elements. Examples of these considerations include: channel-to-channel screening for transistors with multiple CNTs per device, the various geometric capacitances due to the gate and source/drain electrodes, parasitic resistance of the ungated source/drain extension regions, contact resistance, and the quantum capacitance that arises from the low density of states of the carbon nanotube.

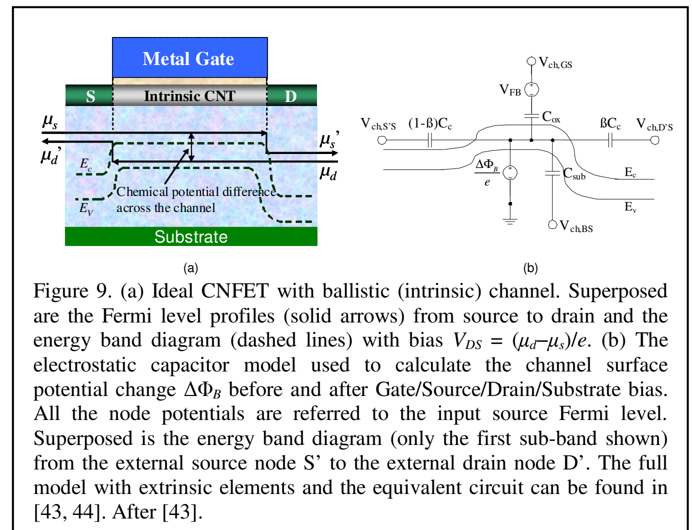


Figure 9. (a) Ideal CNFET with ballistic (intrinsic) channel. Superposed are the Fermi level profiles (solid arrows) from source to drain and the energy band diagram (dashed lines) with bias V_{DS} = $(\mu_d-\mu_s)/e$. (b) The electrostatic capacitor model used to calculate the channel surface potential change $\Delta\Phi_B$ before and after Gate/Source/Drain/Substrate bias. All the node potentials are referred to the input source Fermi level. Superposed is the energy band diagram (only the first sub-band shown) from the external source node S' to the external drain node D'. The full model with extrinsic elements and the equivalent circuit can be found in [43, 44]. After [43].

Fig. 9 shows the ideal intrinsic CNFET with a ballistic channel and the electrostatic capacitor model used to calculate the surface potential as a function of applied biases. The device structure (Fig. 6(a)) is a top-gated CNFET with multiple CNT channels per device, with the source/drain contacted by metals connected to ungated, doped source/drain extension regions. A conventional MOSFET-like equivalent circuit for the CNFET can be used to account for the intrinsic trans-capacitances [43] of the device. The screening and geometric capacitances can be modeled using the methodology described in Section II-D [19, 20]. The quantum capacitance and the voltage-to-charge relationship in the channel can be approximated by using a tight binding approximation of the band structure [42] and iteratively solving the Poisson's equation [43]. Carrier transport can be described by a ballistic model [40] coupled with a phenomenological description of various scattering processes (acoustic phonon, optical phonon) using a mean-free-path parameter. Modeling of the contact resistance remains an unsolved challenge [44] as the physics of the metal to CNT contact is still unclear [45]. The CNFET compact model was developed in HSPICE and later ported to VerilogA [46, 47].

The circuit performance of various circuit structures (simple logic gates, inverter chains) have been simulated using the CNFET model to gain insights into how various device design parameters will affect circuit performance [17, 46].

Device variations and other imperfections can be simulated in a conventional way [17]. The energy-delay improvement over silicon CMOS can be substantial [17] (>10×) only if the density of carbon nanotube is more than 100 CNT/μm (Fig. 10). The CNT density that are typically achieved experimentally today is about 3 – 5 CNT/μm [48, 49]. At this CNT density, the circuit performance of CNFET is substantially lower than silicon CMOS because the parasitic capacitance of the device dominates over the current drive provided by the few number of CNT channels in a device even though the current drive of the CNT channel is excellent when normalized to the width of a single CNT. Additionally, the benchmarking exercise points out two aspects of CNFET technology that are important for technology projection in future technology nodes: (1) the ideal subthreshold slope of the CNFET is identical to conventional MOSFETs with a good electrostatic control; thus, the power supply for CNFET must be similar to that of silicon CMOS in order to maintain gate over-drive, (2) the contact and series resistance for CNFET are extremely important because of the low channel resistance of the CNFET provided by the excellent carrier transport.

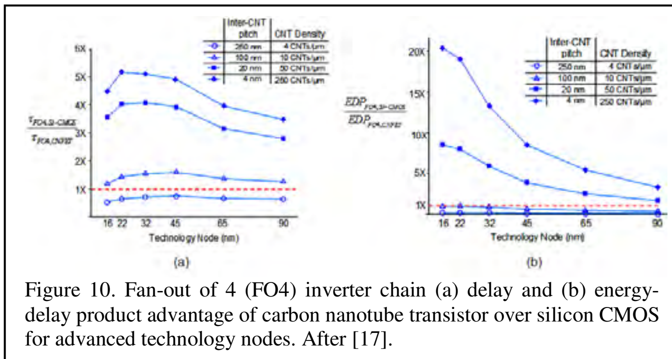


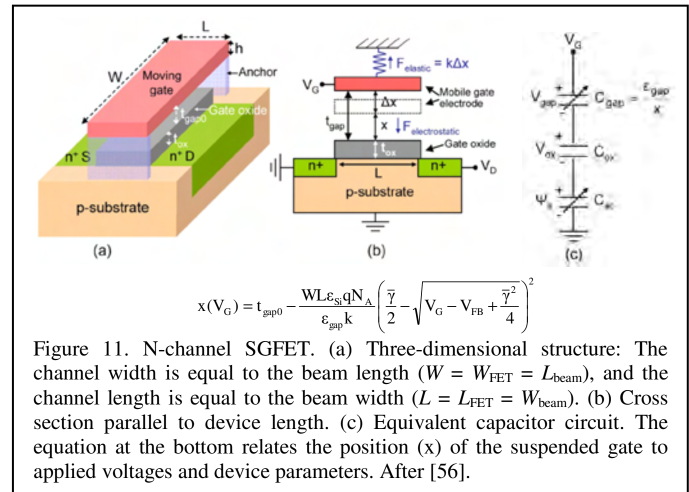
Figure 10. Fan-out of 4 (FO4) inverter chain (a) delay and (b) energy-delay product advantage of carbon nanotube transistor over silicon CMOS for advanced technology nodes. After [17].

In order to further improve the computation speed of the model, the iterative solution for the surface potential can be replaced by a segmented linear fit [50] or an analytical approximation of the density of states of the CNT [51, 52] with reasonable accuracy. A faster model will enable exploration of optimized device design points (e.g. CNT diameter, gate dielectric thickness, gate length, gate workfunction) via a system level optimizer [53].

B. Nanoelectromechanical(NEM) FET and Relay

The use of mechanical motion for logic switching introduces new opportunities for substantially lowering the off-state leakage current and the power supply voltage.

1) Suspended gate FET (SGFET): An early adaptation of nanoelectromechanical structures to the FET is the suspended gate FET (SGFET) [37, 54]. The SGFET structure is similar to that of a MOSFET except that there is an air gap between the gate insulator and the gate electrode (Fig. 11). The device was initially conceived for resonator applications, where mechanical gate vibrations could efficiently be converted to electrical signals via FET source-drain current [54]. Recently, the idea of using the SGFET as an abrupt current switch was also introduced [37, 38] and experimentally demonstrated [55]. In SGFETs (sometimes also called NEMFETs) optimized for logic applications [38], once the "pull-in instability" occurs, the gate electrode collapses onto the gate oxide. This leads to an abrupt increase in gate capacitance and abrupt reduction in threshold voltage, resulting in turn-on of the device with an infinitely sharp slope [56].



$$x(V_G) = t_{gap0} - \frac{WL\varepsilon_{Si}qN_A}{\varepsilon_{gap}k}\left(\frac{\bar{\gamma}}{2} - \sqrt{V_G - V_{FB} + \frac{\bar{\gamma}^2}{4}}\right)^2$$

Figure 11. N-channel SGFET. (a) Three-dimensional structure: The channel width is equal to the beam length ($W = W_{FET} = L_{beam}$), and the channel length is equal to the beam width ($L = L_{FET} = W_{beam}$). (b) Cross section parallel to device length. (c) Equivalent capacitor circuit. The equation at the bottom relates the position (x) of the suspended gate to applied voltages and device parameters. After [56].

The SGFET current–voltage characteristics can be simply obtained starting from any MOSFET compact model just by replacing the oxide capacitance C_{ox} in the original model equations with a series equivalent of C_{ox} and C_{gap} where C_{gap} is the capacitance of the air gap (Fig. 11(c)). Exact numerical and approximate analytical C_{gap} models based on parallel-plate assumption are available in [54] and [56]. Even such a simple integration of the C_{gap} into a MOSFET model can provide useful insights into the operation and utility of the SGFET as a circuit element. This is illustrated in Fig. 12, where the simple "compact SGFET model" reveals that the complementary SGFET inverter exhibits hysteresis, similar to the characteristics of a Schmitt-trigger. This feature is proposed to be exploited for designing SRAM cells based on SGFETs [56]. Other promising applications of the SGFETs include the header and footer switches for power management and SRAM configuration switches for field-programmable-gate-arrays.

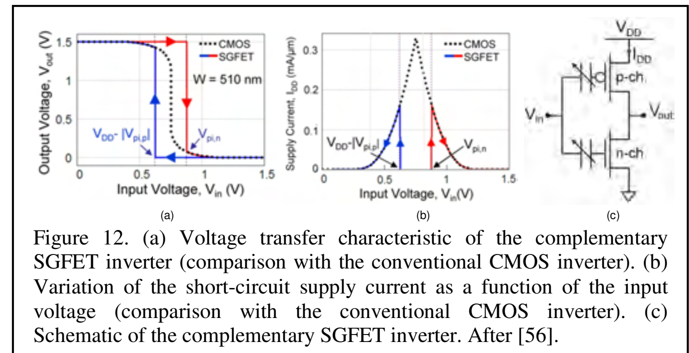


Figure 12. (a) Voltage transfer characteristic of the complementary SGFET inverter (comparison with the conventional CMOS inverter). (b) Variation of the short-circuit supply current as a function of the input voltage (comparison with the conventional CMOS inverter). (c) Schematic of the complementary SGFET inverter. After [56].

2) Nanoelectromechanical (NEM) Relay: Early computers were made of macroscopic electromechanical switches. Today, electromechanical (EM) relays (Fig. 13) are being reconsidered for computation and memory applications. The motivation behind this recently growing interest for EM devices is twofold: (1) the EM relays have inherent advantages for logic

and memory operations, such as zero leakage and sharp hysteresis; (2) they are now believed to be fabricated small enough to provide an acceptable footprint, speed, and operating voltage for VLSI digital applications.

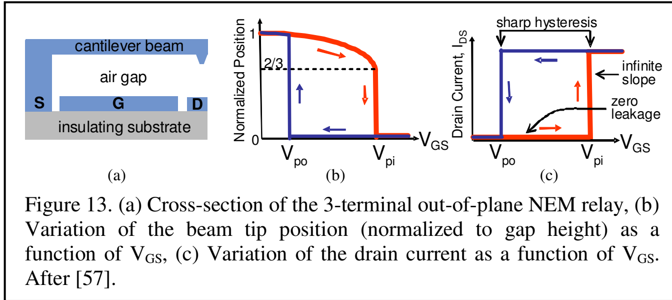
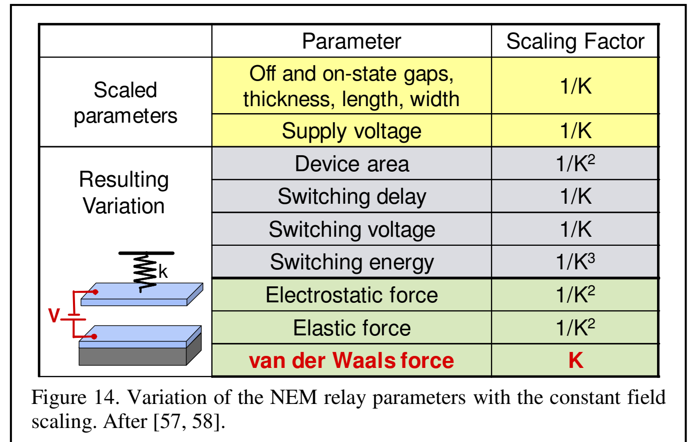


Figure 13. (a) Cross-section of the 3-terminal out-of-plane NEM relay, (b) Variation of the beam tip position (normalized to gap height) as a function of V_{GS}, (c) Variation of the drain current as a function of V_{GS}. After [57].

The operation of NEM relays is based on physics that is significantly different from MOSFETs; yet, NEM relays are proposed as candidate devices to replace MOSFETs. Thus, even a very simple model (e.g. 1D model) of NEM relays is valuable in obtaining insights of how different physics involved affect different aspects of device performances in comparison to MOSFETs.

A simple 1D electrostatic model of the electromechanical system can already provide insights into the device performance [57, 58]. Fig. 14 shows the scaling properties of NEM relay device parameters under a constant field scaling (CFS) scenario. It is seen that the surface adhesion forces (in particular the van der Waals force, F_{vdw}), which increase with the Constant-Field scaling (CFS), affect the ideal scaling trends drastically (Fig. 15(a)) and make an improved NEM relay scaling scheme necessary. Fig. 15(a) suggests that F_{vdw} can, in principle, be used to decrease the switching voltage if the relay dimensions are such that the F_{vdw} is comparable to the electrostatic force (F_e). However from the manufacturability point of view, exploitation of F_{vdw} in order to reduce the switching voltage is restricted by the maximum affordable sensitivity to the dimensions. This sensitivity can be quantified with the slope, S, of the plot in Fig. 15(a). Physically, S is a measure of how much F_{vdw} is involved in switching. As S increases, the F_e to F_{vdw} ratio at $V_{GS} = V_{pi}$ decreases. As an example, it can be shown that for S = 2, $F_e/F_{vdw} = 2.4$, while for S = 10, $F_e/F_{vdw} = 0.4$ (meaning that in this latter case the switching work is mostly done by F_{vdw}) [57]. The design curve in Fig. 15(a) is similar to the short channel effect design curve provided by the scale length theory [32] where the immunity to short channel effect of a MOSFET is uniquely characterized by the ratio of the gate length to the scale length [59].

In order to properly take into account F_{vdw} and the resulting sensitivity issue for the NEM relays, the "Constant-Sensitivity Scaling" (CSS) [57] is proposed. CSS consists of scaling the beam aspect ratio (beam length divided by the beam thickness) together with the dimensions such that the increase in adhesion forces is compensated by the restoring elastic force and S is kept constant. As a result, voltage scaling for NEM relay is challenging because of the presence of the surface adhesion forces.



	Parameter	Scaling Factor
Scaled parameters	Off and on-state gaps, thickness, length, width	1/K
	Supply voltage	1/K
Resulting Variation	Device area	$1/K^2$
	Switching delay	1/K
	Switching voltage	1/K
	Switching energy	$1/K^3$
	Electrostatic force	$1/K^2$
	Elastic force	$1/K^2$
	van der Waals force	**K**

Figure 14. Variation of the NEM relay parameters with the constant field scaling. After [57, 58].

According to Fig. 15(b), for S = 2 and silicon surfaces, the NEM relay achieves 1 ns delay for $V_{GS} = V_{pi} = 1$ V (for 7.5 nm gap and beam thickness [58]). If the silicon surfaces are coated with teflon, which reduces the adhesion force by 6x [60], a voltage-delay performance similar to that of the LSTP NMOSFET is obtained. Finally, if the S is further increased from 2 to 10, the NEM relay achieves 1 ns at only 200 mV (for 4 nm gap, 4 nm beam thickness, and 100 nm beam length [58]). Relative to 45 nm LSTP MOSFET, this last scenario presents 2.25x lower voltage at the same delay or 125x faster switching at the same voltage. It can be shown that the dynamic energy dissipation at 1 ns is only 1.5 aJ, presenting 2.7x reduction relative to that of the MOSFET with the same width (100 nm). These results suggest that the future NEM research should primarily pursue the decrease of surface adhesion with material and process-based solutions. Also, minimizing the process variability would enable tolerating a higher sensitivity, which will decrease the operating voltage and thus improve the performance.

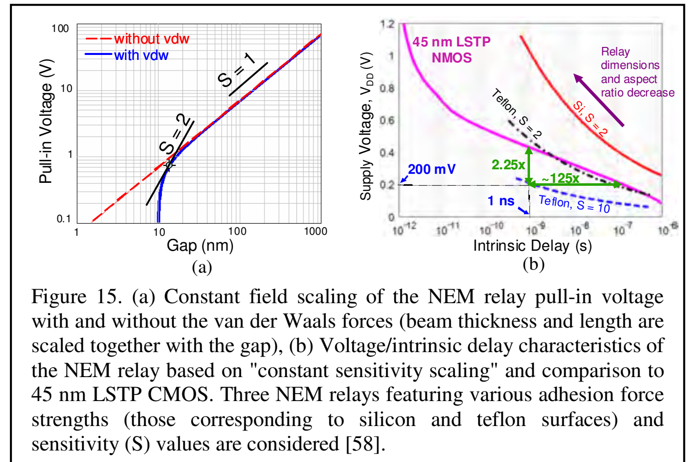


Figure 15. (a) Constant field scaling of the NEM relay pull-in voltage with and without the van der Waals forces (beam thickness and length are scaled together with the gap), (b) Voltage/intrinsic delay characteristics of the NEM relay based on "constant sensitivity scaling" and comparison to 45 nm LSTP CMOS. Three NEM relays featuring various adhesion force strengths (those corresponding to silicon and teflon surfaces) and sensitivity (S) values are considered [58].

IV. SUMMARY

Technology projections at advanced technology nodes necessarily require the inclusion of the parasitic resistances and parasitic capacitances. This is because in the limit of zero gate length and completely ballistic transport, the device performance is determined by the parasitic capacitances [61] and parasitic resistances. Simple analytical model or compact device model for circuit simulation are invaluable tools for

estimating performance at the circuit level. We illustrate the development and application of these models with selected examples from III-V FETs, carbon nanotube transistors, and nanoelectromechanical transistors and relays. These models, while they may be crude approximations of the rigorous device physics, do provide insights for device design and point to directions for technology development since they reflect performance at the circuit level while maintaining a direct link to the physical device parameters. A hierarchy of modeling tools that span the gamut from materials to devices and circuits are required to enable research and development of new technologies.

ACKNOWLEDGMENT

Discussions and collaborations with D.A. Antoniadis (MIT), F. Boeuf (ST Microelectronics), L. Chang (IBM), C.T. Chuang (IBM and National Chiao Tung University), J. del Alamo (MIT), D. Frank (IBM), W. Haensch (IBM), R.T. Howe (Stanford), C. Hu (UC Berkeley), K. Kim (IBM), S. Mitra (Stanford), N. Patil (Stanford), and T. Skotnicki (ST Microelectronics) are gratefully acknowledged.

The work described in this paper is sponsored in part by the SRC, the Focus Center Research Program (FCRP) C2S2, MSD, and FENA, the National Science Foundation, and the member companies of the Stanford Initiative for Nanoscale Materials and Processes (INMP). L. Wei and S. Oh are additionally supported by the Stanford Graduate Fellowship. S. Oh and S. Chong are additionally supported by the Samsung Scholarship. J. Deng was with Stanford University. He is currently with IBM SRDC.

REFERENCES

[1] J. Deng, H.-S. P. Wong, T-*ED*, p. 1317 (2006).
[2] A. Khakifirooz, D.A. Antoniadis, *T-ED*, p. 1391 (2008).
[3] M.H. Na, E. J. Nowak, W. Haensch, J. Cai, *IEDM*, p. 121 (2002).
[4] K. von Arnim, C. Pacha, K. Hofmann, T. Schulz, K. Schrulfer, and J. Berthold, *IEDM*, p. 483 (2007).
[5] A. Khakifirooz, D.A. Antoniadis, *T-ED*, p. 1401 (2008).
[6] H.-S. P. Wong, L. Wei, J. Deng, *ICSICT,* pp. 21 – 24, 2008.
[7] L. Wei, F. Boeuf, T. Skotnicki, H.-S. P. Wong, *VLSI-TSA*, p. 78, 2009.
[8] J. Deng, K. Kim, C.-T. Chuang, H.-S. P. Wong, *T-ED*, p. 1148 (2007).
[9] L. Wei, J. Deng, L.-W. Chang, K. Kim, C.T. Chuang, H.-S. P. Wong, *T-ED*, p. 312 (2009).
[10] Z. Ren, K. T. Schonenberg, V. Ontalus, I. Lauer, and S. A. Butt, *ICSICT*, paper A1.6, 2008.
[11] K. Suzuki, *T-ED*, p. 1895 (1999).
[12] A. Bansal, B. Paul, K. Roy, *T-ED*, p. 256 (2005).
[13] H. C.-H. Wang *et al*., *IEDM,* p.67, 2006.
[14] F. Boeuf, M. Sellier, A. Farcy, T. Skotnicki, *T-ED*, p. 1433 (2008).
[15] H.-S. P. Wong, *IBM J. Res & Dev.*, p. 133 (2002).
[16] W. Haensch et al., *IBM J. Res. & Dev*., p. 339 (2006).
[17] N. Patil, J. Deng, S. Mitra, H.-S. P. Wong, *T-Nano*,p. 37 (2009).
[18] X. Wang, H.-S. P. Wong, P. Oldiges and R.J. Miller, *SISPAD*, p. 163 (2003).
[19] J. Deng, H.-S. P. Wong, *T-ED*, p. 2377 (2007).
[20] L. Wei, J. Deng, H.-S. P. Wong, *T-Nano,* p. 720 (2008).
[21] S.E. Thompson et al., *T-ED*, p. 1790 (2004).
[22] D. A. Antoniadis et al., *IBM J. Res. & Develop.*, p. 363 (2006).
[23] K. C. Saraswat, C. O. Chui, D. Kim, T. Krishnamohan and A. Pethe, *IEDM Tech. Dig*. (2006).
[24] H. Shang et al., *IEDM,* p. 441(2002).
[25] D. H. Kim, J. A. del Alamo, J. H. Lee, K. S. Seo, *IEDM,* p. 767 (2005).
[26] D. -H. Kim, J. A. del Alamo, *IEDM,* p. 629 (2007).
[27] D. Delagebeaudeuf, N. T. Linh, *T-ED*, p. 955, 1982.
[28] S. Oh and H.-S. P. Wong, *T-ED*, p. 1161 (2008).
[29] D. H. Kim and J. A. del Alamo, *CS Mantech*., p. 251(2006).
[30] S. Oh and H.-S. P. Wong, *IEDM*, p. 883 (2008).
[31] K. Suzuki, Y. Tosaka, T. Sugii, *T-ED*, p. 1166 (1996).
[32] D. Frank, Y. Taur, H.-S. P. Wong, *EDL*, p. 385 (1998).
[33] J. Appenzeller, Y.-M. Lin, J. Knoch, and P. Avouris, *Phys. Rev. Lett.,* p. 196805 (2004).
[34] Q. Zhang. W. Zhao, A. Seabaugh, *EDL*, p. 297 (2006).
[35] W. Y. Choi, B.-G. Park, J. D. Lee, and T.-J. King Liu, *EDL,* p. 743 (2007).
[36] S.H. Kim, H. Kam, C. Hu and T.-J. King-Liu, *Symp. VLSI Tech.*, p. 178 (2009).
[37] A. M. Ionescu, V. Pott, R. Fritschi, K. Banerjee, M. J. Declerq, P. Renaud, C. Hibert, P. Fluckiger, and G. A. Racine, *Proc. ISQED,* p. 496 (2002).
[38] H. Kam, D.T. Lee, R.T. Howe, and T.-J. King, *IEDM*, p. 463 (2005).
[39] N. Patil, J. Deng, A. Lin, H.-S. P. Wong, S. Mitra, *T-CAD,* p. 1725 (2008).
[40] J. Guo, A. Javey, H. Dai, and M. Lundstrom, *IEDM*, p. 703 (2004).
[41] A. Javey, J. Guo, D. B. Farmer, Q. Wang, D. Wang, R. G. Gordon, M. Lundstrom, and H. Dai, *Nano Lett.*, p. 447 (2004).
[42] J. W. Mintmire and C. T. White, *Phys. Rev. Lett.*, p. 2506 (1998).
[43] J. Deng, H.-S. P. Wong, *T-ED*, p. 3186 (2007).
[44] J. Deng, H.-S. P. Wong, *T-ED*, p. 3195 (2007).
[45] C. Lan, P. Srisungsitthisunti, P. B. Amama, T.S. Fisher, X. Xu, and R. G. Reifenberger, *Nanotechnology,* **19,** p. 125703 (2008).
[46] J. Deng, A. Lin, G.C. Wan, H.-S. P. Wong, *ACM JETC,* **4**, Issue. 2, pp. 7:1 – 7:20, April, 2008.
[47] Available at http://nano.stanford.edu/models.php
[48] C. Kocabas, M. Shim and J.A. Rogers, *J. Am. Chem. Soc.*, p. 4540 (2006).
[49] N. Patil, A. Lin, E. Myers, K. Ryu, A. Badmaev, C. Zhou, H.-S. P. Wong, S. Mitra, *T-Nano*, p. 498 (2009).
[50] L. Wei, D. Frank, L. Chang, H.-S. P. Wong, *ESSDERC*, p. 222 (2008).
[51] D. Akinwande, J. Liang, S. Chong, Y. Nishi, H.-S. P. Wong, *J. Appl. Phys.*, p. 124514 (2008).
[52] J. Liang, D. Akinwande, H.-S. P. Wong, *J. Appl. Phys.*, p. 064515 (2008).
[53] D. J. Frank, W. Haensch, G. Shahidi, O. Dokumaci, "Optimizing CMOS technology for maximum performance," *IBM J. Res. & Dev*, p. 419 (2006).
[54] H. C. Nathanson, W. E. Newell, R. A. Wickstrom, and J. R. Davis, *T-ED,* p. 117 (1967).
[55] N. Abele, R. Fritschi, K. Boucart, F. Casset, P. Ancey, and A. M. Ionescu, *IEDM,* p. 479 (2005).
[56] K. Akarvardar, C. Eggimann, D. Tsamados, Y. Chauan, G. C. Wan, A. M. Ionescu, R.T. Howe, and H.-S. P. Wong, *T-ED*, p.48 (2008).
[57] K. Akarvardar, D. Elata, R. Parsa, G. C. Wan, K. Yoo, J. Provine, P. Peumans, R. T. Howe, H.-S. P. Wong, *IEDM*, p. 299 (2007).
[58] K. Akarvardar, H.-S. P. Wong, *ECS Trans.* **19** (1), pp. 49-59, 2009.
[59] D.J. Frank, R. H. Dennard, E. J. Nowak, P.M. Solomon, Y. Taur, H.-S. P. Wong, *IEEE Proc.*, p. 259 (2001).
[60] C. M. Mate, in Tribology on the Small Scale, Oxford University Press (2008).
[61] P. Solomon, S. Laux, *IEDM*, p.95 (2001).