A SPICE-Compatible Model of Graphene Nano-Ribbon Field-Effect Transistors Enabling Circuit-Level Delay and Power Analysis Under Process Variation

Ying-Yu Chen[1], Artem Rogachev[1], Amit Sangai[1], Giuseppe Iannaccone[2], Gianluca Fiori[2] and Deming Chen[1]

[1]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL
[2]Dipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy
{chen150, arogach2, sangai2, dchen}@illinois.edu {g.iannaccone, g.fiori}@iet.unipi.it

***Abstract*—This paper presents the *first* parameterized, SPICE-compatible compact model of a Graphene Nano-Ribbon Field-Effect Transistor (GNRFET) with doped reservoirs that also supports process variation. The current and charge models closely match numerical TCAD simulations. In addition, process variation in transistor dimension, edge roughness, and doping level in the reservoir are accurately modeled. Our model provides a means to analyze delay and power of graphene-based circuits under process variation, and offers design and fabrication insights for graphene circuits in the future. We show that edge roughness severely degrades the advantages of GNRFET circuits; however, GNRFET is still a good candidate for low-power applications.**

I. Introduction

Field-effect transistors using carbon-based nano-materials have emerged as promising next-generation devices because of their outstanding electrical properties and integration capabilities via new fabrication techniques [1]–[3]. The most studied are carbon nanotube FETs (CNFETs) and graphene nanoribbon FETs (GNRFETs). Compared to cylindrical CNTs, GNRs can be grown through a silicon-compatible, transfer-free, and *in situ* process [2,4,5], thus having no alignment and transfer-related issues as encountered by CNT-based circuits [2]. However, graphene-based circuits face other types of challenges, including small band gap, degraded mobility, and unstable conductivity due to process variation [6,15,16,20,25]. Therefore, it is important to evaluate these effects and provide a general assessment about the potential and usability of graphene circuits under realistic settings.

Since fabrication technology of GNRFETs is still in an early stage, transistor modeling has been playing an important role for evaluating futuristic graphene circuits. GNRFET simulations based on non-equilibrium Green's function (NEGF) formalism have been published [8,9], which are the most accurate, but are also of the highest complexity. A semi-analytical model was developed in [10], but could not be straightforwardly used in circuit simulation since it still required numerical integrals. A lookup-table-based circuit-level simulator was implemented in [11], and an accurate physics-based compact model was developed in [12] using device-dependent curve-fitting. However, a major drawback of device-dependent models, either based on lookup tables or heavily-fitted equations, is that whenever the need to simulate a new device with a different design parameter arises, a complete set of device simulations are required to rebuild the model. This implies the infeasibility of using above models to perform design space exploration or evaluate the impact of process variation. In order to enable true exploration of graphene-based technology, a parameterized, SPICE-compatible model is required. This allows designers to input custom design parameters and quickly evaluate circuit functionality and performance. In our work, we developed our model based on a wide range of design parameters of sub-20-nm feature sizes, the scale in which GNRFETs are regarded as potential new devices. As a result, our model offers the same features as a typical compact model of a Si-CMOS transistor. Note that there has been research on modeling either CNFETs [13] or Graphene FETs (GFETs[1] [14]) in which such parameterized compact models are proposed, but we are the first to do so on GNRFETs. We plan to release this model to aid designers in exploring graphene-based circuits and evaluating their potentials.

In addition, most existing work regarding graphene circuits focuses either on logic gates [9,11,12] or on interconnects [3] without considering the entire system. We proposed a practical architecture that uses GNRs as both gates and local interconnects, and we discussed how GNRs and metal should be chosen as different interconnects to improve performance. We simulated digital circuits designed in this way by using our GNRFET SPICE model and compared their delay and power performance to that of the 16-nm Si-CMOS technology.

To summarize, the main contributions of our paper are as follows:

- Developing the first parameterizable SPICE-compatible GNRFET model.
- Modeling process variation in several design parameters as well as graphene-specific edge roughness.
- Proposing a GNR-based digital circuit architecture that integrates transistors and interconnects.
- Exploring the design space of GNRFET for desirable transistor-level properties.
- Comparing GNRFET circuits with Si-CMOS circuits.

The rest of the paper is organized as follows: Section II provides additional background on GNRFETs and discusses their use in logic gates; Section III presents our SPICE-compatible GNRFET model for the evaluation of GNRFET circuits; Section IV presents the experimental results; and Section V draws conclusions.

II. Building Circuits with GNRFETs

A. Graphene Properties and Fabrication Techniques

Graphene is a sheet of carbon atoms tightly packed into a two-dimensional honeycomb lattice. It is a zero-band-gap material, which makes it an excellent conductor by nature [2]. Graphene must be processed into narrow strips (GNRs) with widths below 10 nm in order to open a band gap and become semiconducting [2]. Theoretical work has shown that GNRs have band gaps inversely proportional to their widths [15]. Conductivity is also determined by the edge state [15]. GNRs with predominantly *armchair* edges are observed to be semiconducting, while GNRs with predominantly *zigzag* edges demonstrate metallic properties [2][2]. The width of a GNR (denoted W_{CH}) is commonly defined via the number of dimer lines N as illustrated in Figure 1, where $W_{CH} = (N+1) \cdot \sqrt{3} \times 0.144/2$ [17].

There are two varieties of GNRFETs: *SB-type* and *MOSFET-type* [2]. SB-type uses metal contacts and a graphene channel, which form Schottky barriers at junctions. In MOSFET-type GNRFETs, the

[1]A GFET is made of a zero-band-gap graphene sheet instead of GNRs, which are narrowed strips with finite band gaps. GFETs have a low I_{on}/I_{off} ratio and are more suitable in analog applications.

[2]Although zigzag GNRs with pristine edges have a zero band gap, studies showed that band gap could actually be opened for zigzag GNRs with rough edges or those passivated with hydrogen atoms [6,7]. In this work, we will focus on armchair GNRs.

978-3-9815370-0-0/DATE13/© 2013 EDAA

reservoirs are doped with donors or acceptors. Doping with donors (acceptors) results in a N-type (P-type) GNRFET, in which current is dominated by electron (hole) conduction. MOSFET-type GNRFETs demonstrate a higher I_{on}/I_{off} ratio and outperform SB-type ones in digital circuit applications [9]. Therefore, we choose to model MOSFET-type GNRFETs here.

GNR fabrication techniques include lithography, chemical synthesis, and unzipping of carbon nanotubes [18]–[22], etc. Due to limitation of resolution, lithography can only pattern GNRs down to 20 nm in width and tends to produce uneven edges [18]. In [19], a method to produce GNRs ∼4 nm was proposed, in which lithography is used to pattern GNRs and etching is used to narrow GNRs. Chemical synthesis can refine GNRs down to 2 nm in width [21]. Extreme ultraviolet (EUV) lithography is also promising [23]. Further improvement in fabrication technology is necessary to realize mass production of GNR circuits.

Mobility of GNRFETs have been studied [16,20]. In [20], mobility of a GNRFET with a 2.5 nm-wide GNR is reported to be 171-189 cm^2/Vs, calculated based on partial measurements and electrostatic simulations. In [16], GNRFET's mobility is estimated using full-band electron and phonon dispersion relations, and is reported to be ~ 500 cm^2/Vs for 1 nm-wide suspended GNR at room temperature. In our work, channel length is ~ 15 nm and channel width is ~ 1.5 nm. GNRs with this width have a mobility comparable to that of Si-CMOS [16]. Moreover, the mean free path is almost equal to the channel length for such a feature size, and carriers exhibit ballistic transport [16]. Therefore, mobility is less of a concern in this work.

B. Device Structure and Circuit-Level Architecture

Figure 2 shows the structure of the MOSFET-type GNRFET in our proposed design. In one GNRFET, multiple ribbons are connected in parallel to increase drive strength and to form wide, conducting contacts, as demonstrated in [19,22] and modeled in [11]. The ribbons are of armchair chirality. Each GNR is intrinsic (undoped) under the gate and is heavily doped with doping fraction f_{dop} between the gate and the wide contact. The doped parts are called *reservoirs*, and the intrinsic part is called the *channel*. The channel is turned on and off by the gate. L_{CH} is channel length, L_{RES} is the reservoir length, W_{CH} is the ribbon width, W_G is the gate width, and $2W_{sp}$ is the spacing between the ribbons.

For every graphene-metal contact, there is a high resistance introduced on the interface, severely degrading circuit performance [24]. As a result, we seek to minimize the number of graphene-metal contacts in our proposed architecture. The proposed circuit design has multiple metal (e.g. Cu) layers on top of a single graphene layer. Channels, drains, and sources of GNRFETs are located on the graphene layer, and gates of GNRFETs are located on the first metal layer. Connections within each logic gate are made on the graphene layer without the need of vias, and the logic gates are connected to each other on the metal layers. At widths above 20 nm, both zigzag and armchair GNRs serve as good conductors, so there is freedom in routing using GNRs as local interconnects on the graphene layer. Vias are assumed to be metal because vertical graphene vias have not been well studied. Note that the use of graphene-metal vias is inevitable because a logic gate output (source/drain) is on the graphene layer, while a logic gate input is on the metal layer; nevertheless, the proposed architecture minimizes its usage. Figure 3 demonstrates the proposed architecture by showcasing a NAND gate.

III. Modeling GNR Circuits

This section covers the modeling of GNRFET circuits. In III-A, the model of a single GNR ribbon is developed. In III-B, a model of a full GNRFET with multiple GNRs is developed, and modeling of vias and graphene interconnects is presented. Note that the discussion focuses on N-type transistors. Similar derivations can be done for P-type transistors.

A. Single GNR Model

Figure 4 (Left) shows the equivalent circuit of a single GNR, which is similar to the Si-CMOS SPICE model. Our main challenge is to define equations for all components. I_{DS} models the current flowing through the channel, while the capacitors $C_{CH,D}$, $C_{CH,S}$, $C_{G,CH}$, and $C_{SUB,CH}$ along with the voltage-controlled voltage source V_{CH} are included to model the transient currents that result when the channel charges and discharges. We will derive all the equations in the remainder of this subsection.

*1) **Computing the Subbands**:* A positive subband ε_α is given by (1) [10,25], where N is the number of dimer lines as defined in II-A, $t = 2.7$ eV is the tight-binding hopping parameter, α is the subband index ($1 \le \alpha \le N$), and $\delta\varepsilon_\alpha$ is the edge correction factor. A negative subband is computed similarly with a negative sign.

$$\varepsilon_\alpha = \left| t \cdot \left(1 + 2\cos\left(\frac{\pi\alpha}{N+1} \right) + \delta\varepsilon_\alpha \right) \right| \tag{1}$$

The lowest lying subbands dominate the electrostatic and conduction properties [10]. Our experiments show that at most two lowest subbands have a first-order effect on charge and current; hence, our model includes the two lowest subbands for both high accuracy and short computation time. Let α_1 and α_2 be the subband indices corresponding to the two lowest subbands. Let α_0 be a value of α such that $\varepsilon_\alpha = 0$, given by (2). Then, α_1 and α_2 correspond to the two integer values closest to α_0. Plugging α_1 and α_2 into (1) gives the subbands.

$$\alpha_0 = \frac{(N+1)\cos^{-1}(-0.5)}{\pi} = \frac{2N+2}{3} \tag{2}$$

*2) **Finding Channel Potential** Ψ_{CH}:* Let Q_{CH} be the channel charge and Q_{CAP} be the charge across the all capacitors that couple into the channel lumped together. Both Q_{CH} and Q_{CAP} are functions of Ψ_{CH} and have to be equal in magnitude. As a result, equating Q_{CH} and Q_{CAP} yields solution of Ψ_{CH}. In practice, an equation solver (Figure 4, right) is constructed in SPICE to solve for Ψ_{CH}. Note that a similar solver was used in the Stanford CNFET Model [13]. Next, we derive Q_{CH} and Q_{CAP}.

*3) **Finding Channel Charge** Q_{CH}: Q_{CH}* is derived from carrier density. Electron density n_α in subband ε_α is given by (3). Here, $f(E)$ given by (4) is the Fermi-Dirac distribution function, and $D_\alpha(E)$ given by (5) is the density of states (DOS) in a GNR based on [10]. E is the energy level relative to the conduction band E_C. This implies that $E_C = 0$. E_F is the Fermi level relative to E_C, $\hbar$ is the reduced Plank's constant, and M_α is the effective mass given by (6) [10], k is Boltzmann's constant, T is temperature, and $a = 2.46 \times 10^{-10}$ m is the lattice constant.

$$n_\alpha = \int_0^\infty f(E) \cdot D_\alpha(E) \mathrm{d}E \tag{3}$$

$$f(E) = \frac{1}{1 + e^{\frac{E-E_F}{kT}}} \tag{4}$$

$$D_\alpha(E) = \frac{2\sqrt{M_\alpha}}{\pi\hbar} \cdot \frac{\varepsilon_\alpha + E}{\sqrt{\varepsilon_\alpha E(E + 2\varepsilon_\alpha)}} \tag{5}$$

$$M_\alpha = \frac{2\hbar^2\varepsilon_\alpha}{3a^2t^2 \cdot \cos(\frac{\pi\alpha}{N+1})} \tag{6}$$

The integral in (3) has no closed-form solution. A closed-form approximation was derived in [10] by approximating $f(E)$ with

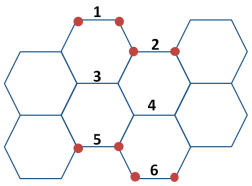

Fig. 1. Lattice structure of a armchair-type GNR with $N = 6$. N is the number of dimer lines in the armchair orientation.

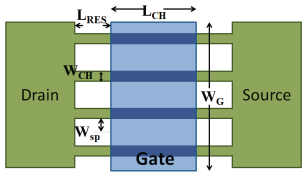

Fig. 2. The structure of a four-ribbon MOSFET-type GNRFET. A common drain and a common source are shared by the ribbons.

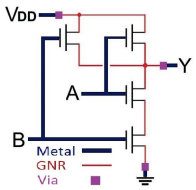

Fig. 3. Using metal and graphene connections in a NAND2 gate.

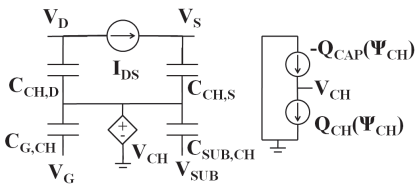

Fig. 4. **Left:** SPICE model of a single GNR. **Right:** SPICE setup for solving Ψ_{CH}. V_{CH} is set to be equal to the channel potential Ψ_{CH}.

Boltzmann distribution $\exp((E_F - E)/kT)$, which is valid when $E - E_F > 3kT$. Since GNRs may have a low subband, the approximation is not always accurate. Therefore, we need to derive an expression valid for all possible E. Since (3) cannot be solved directly, we approximate $f(E)$ with an exponential function when $E_F < E_C = 0$, a step function when $E_F - E_C > 3kT$, and a smoothing function in between.

a) ***Exponential Approximation****:* Here, $f(E)$ is approximated by a decaying exponential function $f'(E)$ [26] as follows:

$$f(E) \sim f'(E) = f(0) \cdot e^{\frac{-E}{\beta kT}} \tag{7}$$

where β is chosen such that $f(3kT) = f'(3kT)$ and is given by

$$\beta = \frac{3}{\ln(f(0)) - \ln(f(3kT))} \tag{8}$$

Note that we have $f(E) = f'(E)$ on the conduction band ($E = E_C = 0$) such that $f'(E)$ approximates $f(E)$ very well when $E \sim E_C$. Since DOS $D_\alpha(E)$ is highest near the conduction band, this gives an accurate estimation of n_α. Electron density computed with this approximation is denoted $n_{\alpha,exp}$ and is given by

$$n_{\alpha,exp}(E) = \frac{\sqrt{M_\alpha(\beta kT)^3}\left(1 + \frac{2\varepsilon_\alpha}{\beta kT}\right)}{2\pi\hbar\varepsilon_\alpha} \cdot e^{\frac{E_F}{\beta kT}} \tag{9}$$

b) ***Step Approximation****:* When $E_F > 3kT$, $f(E) \sim 1$ as $E \sim E_C$. Since DOS $D_\alpha(E)$ is highest in this region, approximating the Fermi-Dirac distribution as a step function (1 when $E \leq E_F$ and 0 when $E > E_F$) provides a good approximation of electron density. Electron density computed with this approximation is denoted $n_{\alpha,step}$ and is given by

$$\begin{aligned} n_{\alpha,step}(E) &= \int_0^{E_F} 1 \cdot D_\alpha(E)\mathrm{d}E \\ &= \frac{2\sqrt{M_\alpha}}{\pi\hbar}\sqrt{\max\left(\frac{E_{FC}(E_{FC} + 2\varepsilon_\alpha)}{\varepsilon_\alpha}, 0\right)} \end{aligned} \tag{10}$$

Note that for $E_F < E_C = 0$, the expression evaluates to 0.

c) ***Combined Approximation****:* We have derived two expressions that approximate electron density n_α under different conditions. To obtain a smooth, continuous charge function, n_α is expressed as a weighted sum of the two approximations as in (11), where m is the relative weight defined in (12). To make the expressions more general, E_{FC} is introduced, which is the difference between the Fermi level and the conduction band. Since $E_C = 0$, $E_{FC} = E_F$. Note that if $E_{FC} = kT$, both approximations are weighted equally. The exponential approximation dominates when $E_{FC} < 0$, while the step approximation dominates when $E_{FC} > 3kT$.

$$n_\alpha(E_{FC}) = m \cdot n_{\alpha,exp}(E_{FC}) + (1 - m)n_{\alpha,step} \tag{11}$$

$$m = \frac{1}{1 + e^{\frac{3(E_{FC} - kT)}{kT}}} \tag{12}$$

The effectiveness of (11) was tested and validated in the range $0.1 < \varepsilon_\alpha < 0.5$. The case where $\varepsilon_\alpha = 0.3$ eV (corresponding to $N = 12$) is shown in Figure 5, where *Numerical* was obtained

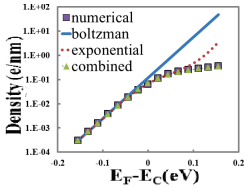

Fig. 5. Charge density n_α vs E_{FC} in the case of $\varepsilon_\alpha = 0.3eV$ ($N = 12$).

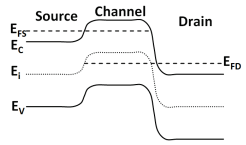

Fig. 6. A typical band diagram of a GNRFET.

by evaluating the integral in (3), *Boltzmann* was obtained from expressions in [10], *Exponential* was obtained from (9), and *Combined* was obtained from (11). All three expressions match *Numerical* when E_{FC} is small. However, as E_{FC} increases, both *Exponential* and *Boltzmann* fail, while *Combined* is accurate throughout the entire range. This is because the combined approximation gives an accurate Fermi level over the entire range, while the exponential and Boltzmann approximations do not.

d) ***Computing Channel Charge*** Q_{CH}*:* Total channel charge Q_{CH} is derived by analyzing the band diagram. Figure 6 shows a band diagram where GNRFET is biased at $V_{GS} > 0$ and $V_{DS} > 0$. Fermi levels at the source and the drain are denoted E_{FS} and E_{FD}, respectively. Since $V_{DS} > 0$, $E_{FD} < E_{FS}$. Because the source and the drain are heavily doped and have high electron densities, E_{FS} and E_{FD} are both above the conduction band.

Holes are negligible in the channel when V_{DS} is low. However, as V_{DS} increases, the conduction band on the drain side ($E_{C,D}$) goes below the valence band of the channel ($E_{V,CH}$), and holes tunnel from the drain into the channel. The tunneling probability $Tr(\Psi_{CH,D})$ is given by (13), where $\Psi_{CH,D}$ is the amount of band bending between channel and drain, $\eta_{0.5}$ is a fitting parameter adjusting the amount of band bending such that $Tr = 0.5$ when $\Psi_{CH,D} > E_G = E_C - E_V$, γ is another fitting parameter controlling how fast Tr increases as $\Psi_{CH,D}$ increases. Note that $\eta_{0.5}$ and γ only need to be obtained once and are valid throughout different devices at different biases. In our implementation, $\eta_{0.5} = 0.6$ and $\gamma = 1/6$.

$$Tr(\Psi_{CH,D}) = \left(1 + e^{\frac{(2+\eta_{0.5})\varepsilon_\alpha - \Psi_{CH,D}}{\gamma\varepsilon_\alpha}}\right)^{-1} \tag{13}$$

The final expression of Q_{CH} (14) is obtained by summing up electron and hole densities and multiplying by electron charge q. The channel potential Ψ_{CH} is the negative of the intrinsic energy level E_i. Therefore, the conduction band is $E_C = \varepsilon_\alpha - \Psi_{CH}$, and the valence band $E_V = -\varepsilon_\alpha - \Psi_{CH}$. Also, the Fermi level at source/drain equals to the applied voltage. Thus, $E_{FS} - E_C = -V_S - (\varepsilon_\alpha - \Psi_{CH})$.

$$\begin{aligned} Q_{CH}(\Psi_{CH}, V_D, V_S) = \frac{qL_{CH}}{2}\sum_\alpha [&-n_\alpha(\Psi_{CH} - \varepsilon_\alpha - V_S) \\ &-n_\alpha(\Psi_{CH} - \varepsilon_\alpha - V_D) \\ &+Tr(\Psi_{CH,D}) \cdot p_\alpha(V_D - \Psi_{CH} - \varepsilon_\alpha)] \end{aligned} \tag{14}$$

*4) **Finding** Q_{CAP}:* Q_{CAP} (15) is composed of several parts. $C_{G,CH}$ and $C_{SUB,CH}$ are physical capacitors that model the coupling between gate/channel and channel/substrate, respectively, empirically modeled by (16). $C_{DIBL,D}$ and $C_{DIBL,S}$ are effective capacitors that model the drain-induced barrier-lowering (DIBL) effect. They were empirically set to $0.15C_{G,CH} \cdot Tr$ and $0.05C_{G,CH}$, respectively. V_{FB} is the flat-band voltage, the work function difference between metal and graphene. ϵ_r is the relative permittivity of the material.

$$\begin{aligned} Q_{CAP} &= C_{G,CH}(V_G - V_{FB} - \Psi_{CH}) \\ &+ C_{SUB,CH}(V_{SUB} - V_{FB} - \Psi_{CH}) \\ &+ C_{DIBL,D}(V_D - \Psi_{CH}) + C_{DIBL,S}(V_S - \Psi_{CH}) \end{aligned} \tag{15}$$

$$C_{G(SUB),CH} = \frac{5.55 \times 10^{-11} \epsilon_r L_{CH}}{\left(1 + \frac{1.5T_{ox}}{W_G}\right) \ln\left(\frac{5.98W_{CH}}{0.8T_{ox}}\right)} \tag{16}$$

*5) **Intrinsic Capacitors**:* By definition, $C_{CH,D} = \partial Q_{CH}/\partial V_D$ and $C_{CH,S} = \partial Q_{CH}/\partial V_S$. They were implemented in SPICE as a voltage-controlled capacitor by defining the charge equation.

*6) **Current Modeling**:* Given Ψ_{CH}, the electron current I_e is computed from (17) based on the Landauer-Buttiker formalism [10,12]. Here, h is Plank's constant, and $f(\cdot)$ is the Fermi-Dirac distribution. $E_{FD,C}$ ($E_{FS,C}$) is the difference between the E_C in the channel and E_F on the drain (source) side, as in Figure 6. Essentially, the probability of electrons being injected into the conduction band from the source is subtracted from the probability of electrons being injected into the conduction band from the drain. By recognizing the Fermi-Dirac integral of order 0 [26], (17) can be evaluated analytically, which yields (18). In an N-type GNRFET, $I_{DS} = I_e$, while in a P-type GNRFET, $I_{DS} = I_h$, which is obtained similarly.

$$I_e = \frac{2q}{h} \sum_{\alpha} \int_0^{\infty} \left[f(E - E_{FS,C}) - f(E - E_{FD,C}) \right] \mathrm{d}E \tag{17}$$

$$\begin{aligned} I_e(\Psi_{CH}, V_D, V_S) = \frac{2qkT}{h} \sum_{\alpha} \Bigg[&\ln\left(1 + e^{\frac{q(\Psi_{CH} - V_S) - \varepsilon_\alpha}{kT}}\right) \\ &- \ln\left(1 + e^{\frac{q(\Psi_{CH} - V_D) - \varepsilon_\alpha)}{kT}}\right) \Bigg] \end{aligned} \tag{18}$$

*7) **Considering Edge Roughness**:* To date, fabrication technology cannot produce GNRs with perfectly smooth edges. The uneven edges result in a phenomenon called *edge roughness*, which affects the properties of GNRs. Edge roughness is characterized by p_r, the probability that any atom on the edges of a GNR is removed [8]. The removal of atoms has two effects: 1) Subbands (1) varies throughout the channel as N is no longer constant. 2) Ballistic transport is disrupted. These effects strongly depend on which atoms are removed [8]; hence, numerical simulations are required for the most accurate analysis. Nevertheless, we are able to model the trend as p_r varies and evaluate the effect of edge roughness on the circuit level.

To model the varying width, we introduce the concept of an effective subband $\varepsilon_{\alpha,eff}$ given by (19), where $\varepsilon_{\alpha,N}$ is the ε_α for a given N. In a unit segment of GNR, there are 8 atoms (shown as red dots in Figure 1) that would reduce N by 1 if removed. Therefore, the probability of N remaining unchanged is $(1-p_r)^8$. And $\varepsilon_{\alpha,eff}$ is the weighted average of $\varepsilon_{\alpha,N}$ and $\varepsilon_{\alpha,N-1}$, given by (19). The scattering coefficient A is introduced to account for the current reduction due to disrupted ballistic transport. It is empirically modeled as (20).

$$\varepsilon_{\alpha,eff} = (1-p_r)^8 \varepsilon_{\alpha,N} + 1(1-(1-p_r)^8)\varepsilon_{\alpha,N-1} \tag{19}$$

$$A = 0.98(1 - 4p_r)^6 + 0.02 \tag{20}$$

The current equation derived in III-A6 assumes ballistic transport and is denoted I_{bal}. Current with edge roughness present, I_{rough}, is derived from I_{bal} and is modeled as follows:

$$I_{rough} = A \cdot I_{bal}(\varepsilon_{\alpha,eff}) \tag{21}$$

B. Full GNRFET Model, Vias, and Interconnects

Figure 7 shows the SPICE implementation of a GNRFET with four parallel GNRs equivalent to that in Figure 2. Each transistor highlighted in red corresponds to an individual GNR, which is modeled by the circuit in Figure 4. C_{GD} and C_{GS}, given by (22), are parasitics introduced by the fringing fields between the gate and the reservoirs. They are modeled empirically based on data from FastCap [27]. When two GNRFETs are connected, graphene-metal contact resistance exists externally between gates and drains/sources.

$$C_{GD} = C_{GS} = 1.26 \times 10^{-10} W_G (0.8 - 0.2T_{ox} + 0.015T_{ox}^2) \tag{22}$$

The local GNR interconnects (20 nm wide) between transistors are much shorter than the mean free path of graphene and have negligible resistance. For this reason, resistance of interconnects within logic gates is neglected in a first-order model, as in [11]. On the other hand, the impact of the graphene/metal contact resistance introduced by vias is significant. The contact resistance is modeled based on experimental results from [24].

IV. Experimental Results

The equivalent circuit model and all equations in III were implemented in HSPICE. In IV-A, the compact model is validated against numerical simulation in Nano TCAD ViDES [8,17] and compared with measurement data from fabricated GNRFETs. In IV-B, we implemented digital logic gates with our GNRFET SPICE model, performed delay and power analysis, and compared them with those implemented in Si-CMOS 16-nm High Performance technology library from PTM [28].

A. Transistor Model Validation

*1) **Default Device**:* First, we simulated a GNRFET with parameters $N = 12$, $L_{CH} = 15$ nm, $L_{RES} = 10$ nm, $T_{ox} = 1$ nm, $f_{dop} = 0.005$, and $V_{FB} = 0$, which is the default device setting in ViDES. The I-V curves of the GNRFET biased at $0 \leq V_{GS} \leq 0.8V$ and $0 \leq V_{DS} \leq 0.8$ V are plotted in Figure 8, in which *num* stands for ViDES and *ana* stands for our model. The voltage range is chosen by assuming a maximum supply voltage $V_{DD} = 0.8$ V, similar to that in the Si-CMOS 16-nm technology ($0.7 - 0.9$ V). It is shown that our model agrees very well with numerical simulations. By defining $I_{on} = I(V_{GS} = V_{DS} = V_{DD})$ and $I_{off} = I(V_{GS} = 0, V_{DS} = V_{DD})$, it can be observed that the I_{on}/I_{off} ratio is reduced at higher V_{DS}. This is caused by an increased Ψ_{CH} due to high V_{DS}. This also serves as a guideline of choosing V_{DD} as it cannot be raised too high in order to maintain a high I_{on}/I_{off} ratio suitable for digital applications. While a low V_{DD} gives a higher subthreshold swing, the I_{on}/I_{off} ratio reaches maximum around $V_{DD} = 0.5$V.

*2) **Variation in Design Parameters**:* Next, we validated that the model responds correctly to changes in design parameters, specifically, N, f_{dop}, T_{ox}, and p_r. I_{on} and I_{off} at $V_{DD} = 0.5$ V were computed at various settings in our model and in ViDES.

Figure 9 shows the effect of N. Our model tracks the periodic effect on band gaps discussed in [25]. For $N = 8, 11, 14$, and 17, the band gap is very small, resulting in a low I_{on}/I_{off} ratio. For $N = 6, 9, 12, 15$, and 18, there is a moderate band gap, which results in a high I_{on}/I_{off} ratio and a high I_{on}. For $N = 7, 10, 13$, and 16, the band gap is the largest, which results in the highest I_{on}/I_{off} ratio. However, I_{on} is still low because the channel is never fully enhanced. Also note that the I_{on}/I_{off} ratio tends to increase as N decreases.

Figure 10 shows the effect of f_{dop}. Doping affects the band bending between the channel and the drain $\Psi_{CH,D}$, and further controls Tr and I_{DS}. Figure 11 shows the effect of T_{ox}. T_{ox} is inversely correlated to $C_{G,CH}$; a smaller T_{ox} implies a larger $C_{G,CH}$,

which provides a better control of Ψ_{CH}. Thus, I_{on} is increased and I_{off} is reduced as T_{ox} decreases. Figure 12 shows the effect of edge roughness in terms of p_r. Edge roughness reduces I_{on}. It also reduces band gaps, which leads to an increase in I_{off}. Even though our model does not match the ViDES data perfectly, it captures the deterioration of the I_{on}/I_{off} ratio as edge roughness is increased.

*3) **Comparison with Measurement Data from Fabricated GNRFETs**:* Among all existing work on fabricated GNRFETs, the single-layer SB-type GNRFET in [20] with W $\sim$ 2 nm is closest to our target range of design parameters. Most of other works evaluated their GNRFETs under high V_{GS} range (e.g., up to 40V) [18,19,21,22]. In [9], a comparison between SB-type and MOSFET-type GNRFETs showed that SB-type FETs have up to 50% lower current than MOSFET-type ones. We conducted a similar comparison between the device in [20] and a $N = 16$ MOSFET-type GNRFET with $p_r = 0.1$ in order to account for the edge roughness (effective $W = 2.1$ nm). For I_{on} and I_{off} with $V_{DS} = 10$ mV, 0.1 V, and 0.5 V respectively across a 2-V range of V_{GS}, the error is within a range of 25% to 100%. The sources of error include the following: 1) The effect of Schottky barriers. 1) Fabricated GNRs do not have a well-defined N, making it difficult for a direct comparison. 3) Current fabricated GNRs have unpredictable width variation and edge roughness. 4) Our model assumes ballistic transport, while the fabricated GNRs in [20] have lengths $>$ 100 nm, greater than the mean free path.

B. Circuit-Level Evaluation

We used our SPICE model to perform DC and transient analysis on basic digital circuits defined in SPICE netlists. This gives insightful information on how GNR-based circuits would perform once fabrication techniques become mature. In our SPICE simulations, an input slew of 10 ps was used, and a 1 fF load was added to the outputs.

*1) **Impact of Supply Voltage**:* We evaluated the delay and power of a 7-stage, fanout-of-4 buffer chain under various supply voltages to understand the power-delay trade-off. The buffer chain was implemented in Si-CMOS, ideal GNRFETs (with no graphene-metal contact resistance), GNRFETs with graphene-metal contact resistance, and GNRFETs with graphene-metal contacts and edge roughness. We implemented Si-CMOS with the 16-nm High-Performance library from Predictive Technology Model (PTM) [28], and implemented GNRFETs with our SPICE model. The minimum-size GNRFET is set to have 6 ribbons in order to match the dimensions of Si-CMOS. Graphene-metal junctions are present in circuit layouts, as discussed in II-B, and they are modeled with a 20-kΩ resistor by assuming a 50-nm via width. Limitations on fabrication techniques contribute to edge roughness. We simulated the cases of $p_r = 5\%$ and 10%. Considering graphene-metal contacts and edge roughness makes our simulations closer to reality. The ideal GNRFET, although not practical, gives an upper bound on circuit performance.

Figure 13 shows the impact of supply voltage V_{DD} on the circuit performance. Graphene-metal contact resistance and edge roughness are nearly inevitable in practice, and they significantly increase delay and leakage power. The optimal operating V_{DD} is around 0.5 V, if delay, dynamic power, and leakage power are all considered.

*2) **Impact of Process Variation**:* Process variation on GNRFETs will result in fluctuations in W_{CH}, L_{CH}, T_{ox}, and f_{dop}. To evaluate the impacts on circuit performance due to these variations, we performed a series of SPICE simulations on the buffer chain in IV-B1 by varying these design parameters to find their respective impacts on the circuit level.

Figure 14 shows the effects of width (N and W_{CH}) variation, which are consistent with Figure 9. $N = 10$ gives high delay and low power due to its low I_{on} and low I_{off} currents. $N = 8$ and 14 have a I_{on}/I_{off} ratio close to 1, and I_{on} and I_{off} are both high. Thus, the delay is low while the leakage power is extremely high. With edge roughness, N_{eff} (corresponding to $\varepsilon_{\alpha,eff}$ in (19)) falls between N and $N - 1$, changing the behavior of the circuit accordingly. Moreover, GNRFETs with higher edge roughness tend to be affected less by the periodic behavior. This explains the dramatic difference between ideal GNRFETs and GNRFETs with $p_r = 0.1$ at $N = 8$ and 14. For the ideal case, $I_{on}/I_{off} \sim 1$; for the latter case, $N_{eff} = 7.57$ and 13.57, and the corresponding I_{on}/I_{off} ratios are high.

The effects of other parameters, L_{CH}, T_{ox}, and f_{dop}, are shown in Table 1, in which we reported the maximum and minimum of delay and power, obtained by varying one design parameter of interest at a time. Among L_{CH}, T_{ox}, and f_{dop}, L_{CH} has the least effect, T_{ox} has an impact on everything, and f_{dop} greatly changes the leakage power. Gate input capacitance is related to L_{CH} and T_{ox}. I_{on} is affected by T_{ox}. Doping mainly controls I_{off}. I_{on} and input capacitance affect delays. I_{off} contributes to leakage power. These observations are consistent with our model.

*3) **Performance Comparison Between GNR and Si-CMOS**:* We compared delay and power performance on a set of digital circuits, implemented with Si-CMOS and GNRFETs, respectively. We choose the Si-CMOS technology node to be the 16-nm HP library from PTM with a nominal $V_{DD} = 0.7$ V because it provides the minimal energy-delay product for voltages ranging between 0.3-1.0 V. According to the exploration in IV-B1 and IV-B2, GNRFETs with $N = 12$, $f_{dop} = 0.001$, and $V_{DD} = 0.5$ V is predicted to have the minimal energy-delay product, and hence we choose to adopt this setting to have a fair comparison with Si-CMOS. To match the default dimensions of Si-CMOS, our GNRFET is set to have 6 GNRs, $L_{CH} = 16$ nm, and $T_{ox} = 0.95$ nm. The set of circuits we simulated include INV, NAND2, NOR2, NAND3, NOR3, the buffer chain in IV-B1, and C17 from ISCAS85. Based on results in Table 2, edge roughness plays a significant role in degrading the current in GNRFETs. As a result, Si-CMOS performs better in delay unless GNR is ideal. In terms of dynamic power, GNRFET has lower consumption than Si-CMOS mostly due to lower V_{DD} and lower gate capacitance. In terms of leakage power for GNRFET, when a sufficiently high V_{DS} is applied, the confined states in the valence band of the channel align with the occupied states of the drain, resulting in band-to-band injection of holes in the channel [10]. This is captured in (13), which describes an exponential relation between V_{DD} and the tunneling probability. First of all, when $V_{DD} = 0.7$ V, GNRFET has a higher leakage power than Si-CMOS shown in Figure 13. However, when V_{DD} is smaller (e.g., 0.5 V), the tunneling is significantly reduced, consuming much lower leakage especially for the ideal case. In Figure 15, we compared the waveforms of two 11-stage ring oscillators, implemented with Si-CMOS and ideal GNR, respectively. Ideal GNR demonstrated a 5.5% higher frequency than Si-CMOS, consistent with our observation in other circuits.

V. Conclusion

We presented a parameterized, SPICE-compatible compact model of a MOSFET-type GNRFET. It captured the effects of $N(W_{CH})$, L_{CH}, T_{ox}, f_{dop}, and edge roughness on current and charge. In addition, we presented a GNR-based circuit architecture that integrates gates and interconnects. The model and the architecture allow circuit-level performance evaluations of GNRFETs under process variation. We observed that GNRFETs are promising compared to CMOS for low power applications, since they have similar delay with smaller leakage power. It is possible that GNRFETs would provide higher operating frequency if the threshold voltages were tuned to

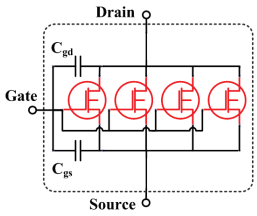

Fig. 7. SPICE model of the GNRFET in Figure 2.

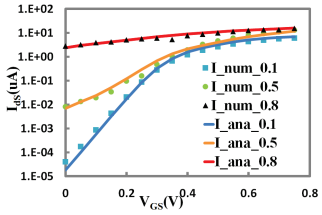

Fig. 8. I_{DS} vs V_{GS} with V_{DS} 0.1, 0.5, 0.8 V in an N-type GNRFET.

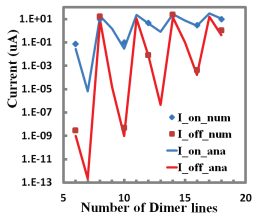

Fig. 9. (I_{on} and I_{off}) vs N. Note that ViDES only supports even N.

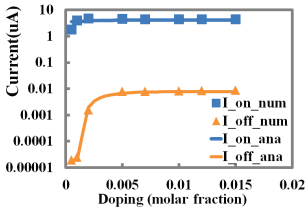

Fig. 10. (I_{on} and I_{off}) vs f_{dop}, doping fraction in reservoirs.

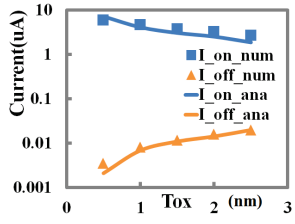

Fig. 11. (I_{on} and I_{off}) vs T_{ox}, oxide thickness.

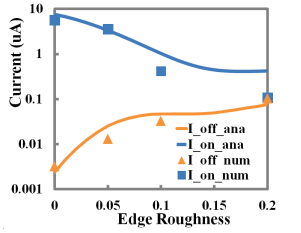

Fig. 12. (I_{on} and I_{off}) vs p_r, edge roughness probability.

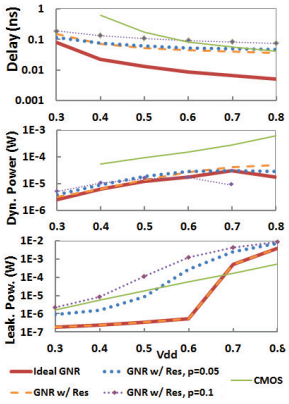

Fig. 13. Delay, dynamic power, and leakage power vs V_{DD}

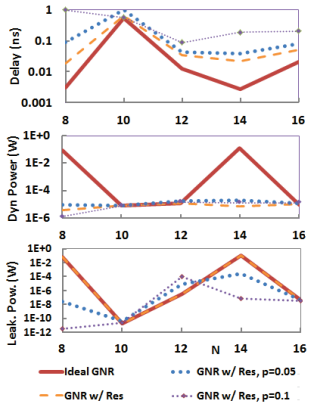

Fig. 14. Delay, dynamic power, and leakage power vs N

achieve the same leakage power as CMOS. We also showed that edge roughness can critically reduce the performance and leakage power advantages of GNRFETs.

Acknowledgment

This work is partially supported by NSF under the grant CCF 07-46608. Ying-Yu Chen and Artem Rogachev contributed equally in this work. Special thanks to Hai Wei and Max Shulaker from Stanford Univ. for their helpful discussions.

Table 1: Effect of variations on GNRFET circuits

		Parameters		Delay (ns)		Dynamic Power (W)		Leakage Power (W)	
		Max	Min	Max	Min	Max	Min	Max	Min
Ideal GNR	L_{CH}	25nm	10nm	0.016	0.011	1.23E-05	1.07E-05	1.17E-04	1.10E-04
	T_{ox}	2.5nm	0.5nm	0.020	0.009	1.39E-05	9.25E-06	2.74E-04	3.52E-05
	f_{dop}	0.015	0.001	0.014	0.012	2.43E-05	1.02E-05	1.25E-04	6.68E-07
GNR w/ Res	L_{CH}	25nm	10nm	0.041	0.030	1.52E-05	1.17E-05	1.09E-04	1.09E-04
	T_{ox}	2.5nm	0.5nm	0.040	0.034	1.55E-05	1.02E-05	2.59E-04	3.38E-05
	f_{dop}	0.015	0.001	0.037	0.031	1.22E-05	1.02E-05	1.23E-04	2.10E-07
GNR w/ Res, p=0.05	L_{CH}	25nm	10nm	0.055	0.038	2.29E-05	1.56E-05	1.27E-03	1.26E-03
	T_{ox}	2.5nm	0.5nm	0.051	0.043	2.64E-05	3.75E-06	2.61E-03	3.45E-04
	f_{dop}	0.015	0.001	0.045	0.041	1.78E-05	1.33E-05	1.28E-03	2.88E-05
GNR w/ Res, p=0.10	L_{CH}	25nm	10nm	0.113	0.077	1.95E-05	1.36E-05	1.62E-03	1.60E-03
	T_{ox}	2.5nm	0.5nm	0.113	0.084	3.21E-05	4.38E-06	2.24E-03	6.57E-04
	f_{dop}	0.015	0.001	0.091	0.090	1.99E-05	1.41E-05	1.61E-03	1.68E-04

Table 2: Comparison with CMOS

Circuit	Delay (ns)					Dynamic Power (W)					Leakage Power (W)				
	Ideal GNR	GNR w/ Res	GNR w/ R, p=0.05	GNR w/ R, p=0.1	CMOS	Ideal GNR	GNR w/ Res	GNR w/ R, p=0.05	GNR w/ R, p=0.1	CMOS	Ideal GNR	GNR w/ Res	GNR w/ R, p=0.05	GNR w/ R, p=0.1	CMOS
inv	1.28E-02	2.68E-02	3.77E-02	7.93E-02	1.45E-02	3.34E-08	3.34E-08	3.31E-08	3.03E-08	9.80E-08	6.10E-11	6.10E-11	1.50E-09	1.98E-08	3.60E-08
nand2	1.74E-02	3.50E-02	5.30E-02	1.22E-01	3.70E-02	4.89E-08	4.90E-08	4.43E-08	4.02E-08	1.78E-07	1.22E-10	1.22E-10	3.04E-09	3.64E-08	7.18E-08
nor2	1.91E-02	3.45E-02	5.31E-02	1.24E-01	2.00E-02	4.85E-08	4.86E-08	4.44E-08	4.02E-08	1.75E-07	1.22E-10	1.22E-10	3.00E-09	3.73E-08	7.51E-08
nand3	2.06E-02	4.40E-02	7.00E-02	1.65E-01	6.30E-02	3.20E-08	3.20E-08	3.01E-08	2.58E-08	9.30E-08	1.83E-10	1.83E-10	4.50E-09	4.90E-08	1.07E-07
nor3	1.28E-02	3.78E-02	5.90E-02	1.48E-01	3.60E-02	3.19E-08	3.19E-08	3.03E-08	2.57E-08	9.28E-08	1.83E-10	1.83E-10	4.55E-09	5.32E-08	2.21E-08
buf chain	1.16E-02	3.08E-02	4.19E-02	1.79E-01	5.50E-02	1.01E-05	1.01E-05	1.36E-05	1.37E-04	2.88E-04	3.34E-07	3.34E-07	8.32E-06	1.08E-04	1.66E-04
c17	1.83E-02	6.09E-02	7.16E-02	1.40E-01	6.00E-02	5.85E-08	5.85E-08	8.19E-08	1.16E-07	2.10E-07	7.30E-10	7.30E-10	1.64E-08	1.86E-07	2.56E-07

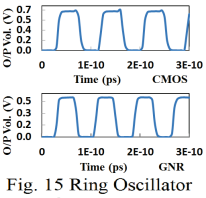

Fig. 15 Ring Oscillator Waveforms

References

[1] J. Zhang et al., "Robust Digital VLSI using Carbon Nanotubes," Keynote paper, *IEEE Trans. CAD*, 2012.

[2] S. Chilstedt, C. Dong, and D. Chen, "Carbon nanomaterials transistors and circuits, transistors: Types, materials and applications," *Nova Science Publishers*, 2010.

[3] W. de Heer et al., "Pionics: the emerging science and technology of graphene-based nanoelectronics," *IEDM*, 2007.

[4] M. P. Levendorf et al., "Transfer-Free Batch Fabrication of Single Layer Graphene Transistors," *Nano Letters*, 2009.

[5] P. Wessely et al., "Transfer-free fabrication of graphene transistors," *J. Vac. Sci. Technol.* B 30, 2012.

[6] H. Raza, "Zigzag graphene nanoribbons: bandgap and midgap state modulation," *J. Phys.: Condens. Matter* 23, 2011.

[7] S.S. Chauhan et al., "Band gap engineering in zigzag graphene nanoribbons - an ab initio approach," *J. Comput. Theor. Nanosci.* 9, 2012.

[8] G. Fiori and G. Iannaccone, "Simulation of graphene nanoribbon field effect transistors," *Elec. Dev. Lett.*, 2007.

[9] Y. Yoon et al., "Performance comparison of Graphene Nanoribbon FETs with Schottky contacts and doped reservoirs," *TED*, 2008.

[10] P. Michetti and G. Iannaccone, "Analytical model of one-dimensional carbon-based Schottky-barrier transistors," *TED*, 2010.

[11] M. Choudhury, Y. Yoon, J. Guo, and K. Mohanram, "Technology exploration for graphene nanoribbon FETs," *DAC*, 2008.

[12] S. Fregonese, C. Maneux, and T. Zimmer, "A versatile compact model for ballistic 1D transistor: Applications to GNRFET and CNTFET," *ISDRS*, 2009.

[13] J. Deng and H.-S. Wong, "A compact SPICE model for carbon-nanotube field-effect transistors including nonidealities and its application Part I: Model of the intrinsic channel region," *TED*, 2007.

[14] M. B. Henry and S. Das, "SPICE-compatible compact model for graphene field-effect transistors," *ISCAS*, 2012.

[15] K. Nakada and M. Fujita, "Edge state in graphene ribbons: Nanometer size effect and edge shape dependence," *Phys. Rev. B, (Condensed Matter)*, 1996.

[16] A. Betti, G. Fiori, and G. Iannaccone, "Strong mobility degradation in ideal graphene nanoribbons due to phonon scattering," *Appl. Phys Lett.*, 2011.

[17] Nano TCAD ViDES. [Online]. http://vides.nanotcad.com/vides/

[18] Z. Chen, Y.-M. Lin, M. J. Rooks, and P. Avouris, "Graphene nano-ribbon electronics," *Physica E: Low-dimensional Systems and Nanostructures*, 2007.

[19] X. Wang and H. Dai, "Etching and narrowing of graphene from the edges," *Nature Chemistry*, 2010.

[20] X. Wang et al., "Room-Temperature All-Semiconducting Sub-10-nm Graphene Nanoribbon Field-Effect Transistors," *Phys. Rev. Lett.*, 2008.

[21] X. Li, X. Wang, L. Zhang, S. Lee, and H. Dai, "Chemically derived, ultrasmooth graphene nanoribbon semiconductors," *Science*, 2008.

[22] L. Jiao et al., "Aligned graphene nanoribbons and crossbars from unzipped carbon nanotubes," *Nano Res.* 2010.

[23] S. Prezioso et al., "Large area extreme-UV lithography of graphene oxide via spatially resolved photoreduction," *Langmuir*, 2012.

[24] K. Nagashio et al., "Metal/graphene contact as a performance killer of ultra-high mobility graphene analysis of intrinsic mobility and contact resistance," *IEDM*, 2009.

[25] Y.-W. Son, M. L. Cohen, and S. G. Louie, "Energy gaps in graphene nanoribbons," *Phys. Rev. Lett.*, 2006.

[26] Y. Taur and T. H. Ning, Fundamentals of Modern VLSI Devices. Cambridge University Press, 2010.

[27] FastCap 2. [Online]. http://www.fastfieldsolvers.com

[28] [Online]. http://www.ptm.asu.edu