GRAPHENE NANO-RIBBON AND TRANSITION METAL DICHALCOGENIDE FIELD-EFFECT TRANSISTOR MODELING AND CIRCUIT SIMULATION

BY

YING-YU CHEN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Associate Professor Deming Chen, Chair
Associate Professor Xiuling Li
Professor Elyse Rosenbaum
Professor Martin D. F. Wong

ABSTRACT

This dissertation presents a modeling and simulation study of graphene nanoribbon and transition metal dichalcogenide field-effect transistors. Through compact modeling, SPICE implementation of the transistors is realized, and circuit-level simulation is enabled. Extensive simulation studies are performed to evaluate the performance of these two emerging devices.

To my parents and Scott, for their love and support.
To my feathered friends, for our invaluable shared memory.

ACKNOWLEDGMENTS

I would like to thank Artem Rogachev, Amit Sangai, Morteza Gholipour, and Zelei Sun for their collaboration in our various research projects that are included in this dissertation.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

In recent years, conventional silicon-based CMOS transistor scaling has become increasingly difficult due to increased wire resistivity, significant mobility degradation, and large dopant fluctuations. Researchers have begun to turn to various emerging materials and novel transistor designs in order to keep up with Moore's law. In particular, the graphene nano-ribbon field-effect transistor (GNRFET) has drawn a lot of attention due to the outstanding electrical and physical properties of graphene. Preliminary theoretical, simulation, and experimental work also shows that GNRFET has great potential in high-performance and low-power applications.

However, process variation in transistor dimensions, oxide thickness, doping level, and graphene-specific line edge roughness (shown in Figure 1.1) has various effects on the current-voltage (I-V) characteristics of the transistor, and hence it affects the delay and power performance on the circuit level significantly. The challenges in mass-producing high-quality nanoscale graphene nano-ribbons (GNRs) also bring concerns on the prospects of this emerging technology. In particular, the width of GNRs needs to scale below 10 nm for the material to open up a band gap and become semiconducting, and thus posing even more challenges in fabrication.

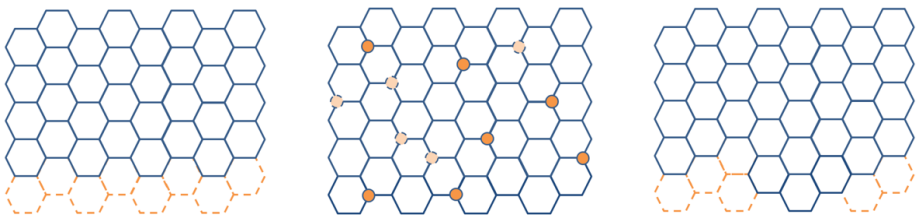

Figure 1.1: From left to right: Width variation on a GNR, doping variation on a GNR, and line edge roughness on a GNR. Note that line edge roughness changes the effective edge width and the edge chirality, which changes the energy band structure of the GNR and hence affects the I-V characteristics.

In response to the concerns to GNRs, researchers turn their attention to other emerging 2-D materials, the transition metal dichalcogenide (TMD) being a very promising one. Monolayer TMDs have a band gap of 1-2 eV by nature and are very suitable for making transistors without scaling down to the nanometer range, which greatly reduces the difficulty of production and also mitigates undesirable effects from process variation. As a result, TMD field-effect transistors (TMDFETs) have been regarded as a promising transistor design in the post-CMOS era. Moreover, TMDFETs are suitable for building flexible electronics due to their 2-D structure, and the effects of bending are of interest as well.

Fabrication technology of GNRFET and TMDFET is still in an early stage. Therefore, simulation is required to assess the performance of these emerging devices. Simulation Program with Integrated Circuit Emphasis (SPICE) is a software used by many circuit designers and computer-aided design engineers to simulate circuits. By developing SPICE models that describe GNRFETs and TMDFETs, we enable simulations of circuits based these novel devices. We also open up opportunities for GNRFET and TMDFET to be compared or integrated with other technologies.

In our work, we model and evaluate two varieties of GNRFETs, Metal-Oxide-Semiconducting-(MOS-) type and Schottky-Barrier-(SB-)type GNRFETs, and two varieties of TMDFETs, made of MoS_2 and WSe_2. We implement these transistor models in SPICE to enable circuit-level simulation. We thoroughly discuss and explore their respective strengths in terms of transistor-level properties and circuit-level delay and power performance. Especially, we evaluate their performance with and without different sources of process variation. For TMDFETs, we also evaluate the effects under bending (in terms of applied strain). We also compare the GNRFETs and TMDFETs with Si-CMOS devices based on the Predictive Technology Models (PTM).

To summarize, the main contributions of this dissertation are as follows:

- Bridging the gap between transistor device design, circuit design, and CAD for two emerging nanotechnologies.
- Developing the *first* parameterized SPICE-compatible models for MOS-GNRFET, SB-GNRFET, and flexible TMDFET.
- Modeling process variation in several design parameters as well as

GNR-specific line edge roughness.

- Modeling bending in terms of applied strain in TMDFET.
- Proposing a GNR-based digital circuit architecture that integrates transistors and interconnects.
- Exploring the design space of GNRFET and TMDFET for desirable transistor-level properties.
- Proposing a new SB-GNRFET design that mitigates the undesirable high leakage current.
- Analyzing transistor-, gate-, and circuit-level properties of GNRFET and TMDFET circuits.
- Simulating non-trivial GNRFET and TMDFET circuits other than *inv* or ring oscillators, providing a realistic view on how GNRFET and TMDFET circuits perform.
- Comparing GNRFET and TMDFET circuits with Si-CMOS circuits.
- Performing Monte Carlo simulations on GNRFET circuits to provide insights on the effect of process variation.

Some of the work was published in [1, 2, 3, 4, 5, 6, 7, 8, 9].

The rest of the dissertation is organized as follows: Chapter 2 and Chapter 3 present our SPICE-compatible MOS-GNRFET and SB-GNRFET models for the evaluation of GNRFET circuits, including model derivation, model validation, and circuit-level simulation results. Chapter 4 evaluates the potential of MOS- and SB-GNRFET as future low-power devices. Furthermore, an improved design of SB-GNRFET based on the model we developed is presented in Chapter 5. Chapter 6 presents our TMDFET SPICE model and circuit-level simulation results. Chapter 7 presents more extensive modeling and simulation of TMDFETs with scaling studies. Chapter 8 draws conclusions.

CHAPTER 2

MOS-GNRFET MODELING AND SIMULATION

2.1 Introduction

Field-effect transistors (FETs) built with carbon-based nano-materials have emerged as promising next-generation devices because of their outstanding electrical properties and integration capabilities via new fabrication techniques [10, 11, 12, 13, 14, 15]. The most studied are carbon nanotube FETs (CNFETs) and graphene nano-ribbon FETs (GNRFETs). Compared to cylindrical carbon nanotubes (CNTs), graphene nanoribbons (GNRs) can be grown through a silicon-compatible, transfer-free, and *in situ* process [13, 16, 17], thus having no alignment and transfer-related issues as encountered by CNT-based circuits [13]. However, graphene-based circuits face other types of challenges, including small band gap, degraded mobility, and unstable conductivity due to process variation [18, 19, 20, 21, 22]. 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 in evaluating futuristic graphene circuits. GNRFET simulations based on non-equilibrium Green's function (NEGF) formalism have been published [23, 24], which are the most accurate, but are also of the highest complexity. A semi-analytical model was developed in [25], but could not be straightforwardly used in circuit simulation since it still required non-closed-form numerical integrals. A lookup-table-based circuit-level simulator was implemented in [26, 27], and an accurate physics-based compact model was developed in [28] 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 param-

eter arises, a complete set of device simulations is 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 [29, 30] or Graphene FETs (GFETs[1][31]) in which such parameterized compact models are proposed, but we are the first to do so on GNRFETs. We have released this model on NanoHub [32] to aid designers in exploring graphene-based circuits and evaluating their potentials. For example, computer-aided design (CAD) algorithms targeting graphene-based circuits have been proposed [33], and they can definitely benefit from more accurate SPICE-level simulations.

In addition, most existing work regarding graphene circuits focuses either on logic gates [24, 26, 27, 28] or on interconnects [15] 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 this chapter are as follows:

- Developing the first parameterizable SPICE-compatible GNRFET model.
- Modeling process variation in several design parameters as well as graphene-specific line edge roughness.
- Proposing a GNR-based digital circuit architecture that integrates transistors and interconnects.

[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.

- Exploring the design space of GNRFET for desirable transistor-level properties.
- Analyzing transistor- and circuit-level properties of GNRFET circuits.
- Simulating non-trivial GNRFET circuits other than *inv* or ring oscillators, providing a realistic view on how GNRFET circuits perform.
- Comparing GNRFET circuits with Si-CMOS circuits.
- Performing Monte Carlo simulations on GNRFET circuits to provide insights on the effect of process variation.

The rest of the chapter is organized as follows: Section 2.2 provides additional background on GNRFETs and discusses their use in logic gates; Section 2.3 presents our SPICE-compatible GNRFET model for the evaluation of GNRFET circuits; Section 2.4 presents the experimental results; and Section 2.5 draws conclusions.

2.2 Building Circuits with GNRFETs

2.2.1 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 [13]. Graphene must be processed into narrow strips (GNRs) with widths below 10 nm in order to open a band gap and become semiconducting [13]. Theoretical work has shown that GNRs have band gaps inversely proportional to their widths [20]. Conductivity is also determined by the edge state [20]. GNRs with predominantly *armchair* edges are observed to be semiconducting, while GNRs with predominantly *zigzag* edges demonstrate metallic properties [13].[2] The width of a GNR (denoted W_{CH}) is commonly defined via the number of dimer lines N as illustrated in Figure 2.1, where $W_{CH} = (N-1) \cdot \sqrt{3} \times 0.144/2$ nm [35].

[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 [22, 34]. In this work, we will focus on armchair GNRs.

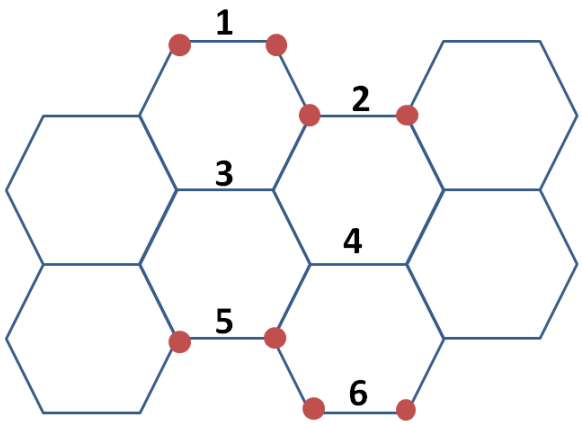

Figure 2.1: Lattice structure of a armchair-type GNR with $N = 6$. N is the number of dimer lines in the armchair orientation.

There are two varieties of GNRFETs: *SB-type* and *MOS-type* [13]. SB-type uses metal contacts and a graphene channel, which form Schottky barriers at junctions. In MOS-type GNRFETs, the reservoirs are doped with donors or acceptors. Doping with donors (acceptors) results in an N-type (P-type) GNRFET, in which current is dominated by electron (hole) conduction. MOS-type GNRFETs demonstrate a higher I_{on}/I_{off} ratio and outperform SB-type ones in digital circuit applications [24]. Therefore, we choose to model MOS-type GNRFETs here.

GNR fabrication techniques include lithography, chemical synthesis, and unzipping of carbon nanotubes [21, 36, 37, 38, 39, 40], etc. Due to limitation of resolution, lithography can only pattern GNRs down to 20 nm in width and tends to produce uneven edges [36]. In [37], a method to produce GNRs $\sim$ 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 [38]. Another bottom-up chemical synthesis approach can produce atomically precise GNRs in different chirality and patterns under 2 nm [39]. Extreme ultraviolet (EUV) lithography is also promising [41]. Further improvement in fabrication technology is necessary to realize mass production of GNR circuits.

Mobility of GNRFETs have been studied [18, 21]. In [21], mobility of a GNRFET with a 2.5 nm-wide GNR is reported to be 171-189 $cm^2/V \cdot s$, calculated based on partial measurements and electrostatic simulations. In [18], GNRFET's mobility is estimated using full-band electron and phonon

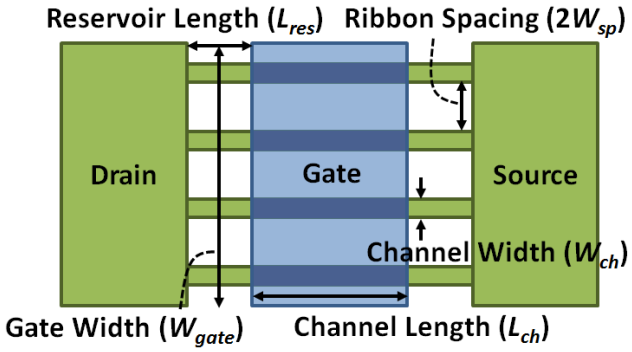

Figure 2.2: The structure of a four-ribbon MOS-type GNRFET. A common drain and a common source are shared by the ribbons.

dispersion relations, and is reported to be $\sim 500\ cm^2/V \cdot s$ 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 [18]. Moreover, the mean free path is almost equal to the channel length for such a feature size, and carriers exhibit ballistic transport [18]. Therefore, mobility is less of a concern in this work.

2.2.2 Device Structure and Circuit-Level Architecture

Figure 2.2 shows the structure of the MOS-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 [40, 37] and modeled in [26, 27]. 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 [42]. 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

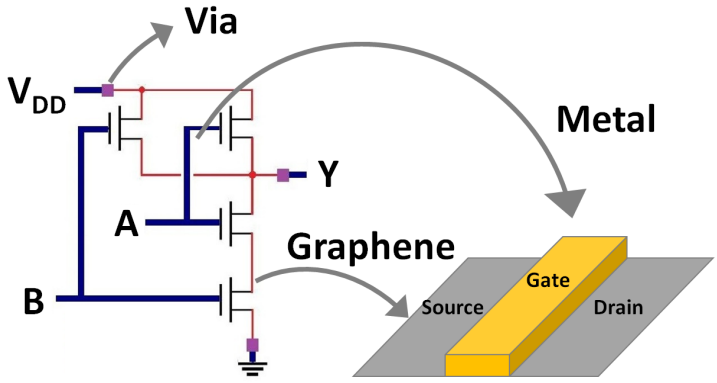

Figure 2.3: A *nand2* gate implemented in the proposed architecture of MOS-GNRFET circuits. Inputs A and B, output Y, and power rails V_{DD} and *gnd* are distributed on the metal layers (bold blue lines). Vias (purple squares) are needed to connect graphene and metal layers. Local interconnects between drains and sources are made of graphene (thin red lines), in order to avoid extra vias.

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 reduces its usage by connecting sources and drains on the graphene layer. Figure 2.3 demonstrates the proposed architecture by showcasing a *nand2* gate.

2.3 Modeling GNRFET and GNR Circuits

This section covers the modeling of GNRFET circuits. In Section 2.3.1, the model of a single GNR ribbon is developed. In Section 2.3.2, 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.

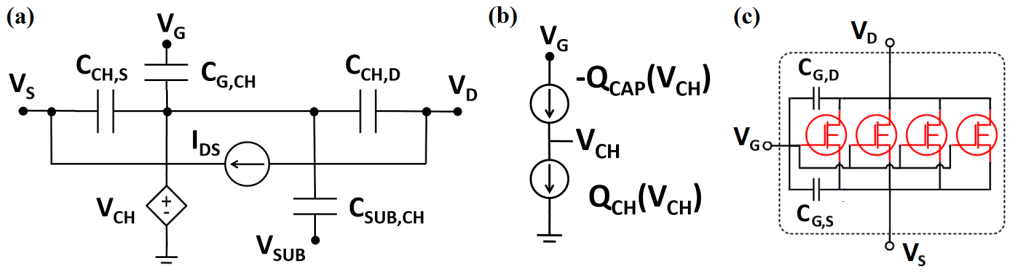

Figure 2.4: **(a)** SPICE model of a single GNR. **(b)** SPICE setup for solving Ψ_{CH}. V_{CH} is set to be equal to the channel potential Ψ_{CH}. **(c)** SPICE model of the GNRFET in Figure 2.2.

2.3.1 Single GNR Model

Figure 2.4 (a) 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.

2.3.1.1 **Computing the Subbands**

A positive subband edge ε_α is given by (2.1) [19, 25], where N is the number of dimer lines as defined in Section 2.2.1, $t = 2.7$ eV is the tight-binding hopping parameter, α is the subband index ($1 \leq \alpha \leq N$), and $\delta\varepsilon_\alpha$ is the edge correction factor, given by (2.2), in which $v = 0.12$ eV is the energy correction of the hopping parameter at the edges in the tight-binding Hamiltonian. A negative subband edge is computed similarly with a negative sign.

$$\varepsilon_\alpha = \left| t \cdot \left(1 + 2\cos\left(\frac{\pi\alpha}{N+1} \right) + \delta\varepsilon_\alpha \right) \right| \tag{2.1}$$

$$\delta\varepsilon_\alpha = \frac{4vt}{N+1} \cos^2\left(\frac{\pi\alpha}{N+1} \right) \tag{2.2}$$

The lowest lying subbands dominate the electrostatic and conduction properties [25]. 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 (2.3). Then, α_1 and α_2 correspond to the two integer values closest to α_0, as in (2.4). Plugging α_1 and α_2 into (2.1) gives the subbands.

$$\alpha_0 = \frac{(N+1)\cos^{-1}(-0.5)}{\pi} = \frac{2N+2}{3} \tag{2.3}$$

$$\alpha_1 = \lfloor \alpha_0 \rfloor; \alpha_2 = \alpha_1 + 1 \tag{2.4}$$

2.3.1.2 Finding Channel Potential Ψ_{CH}

Let Q_{CH} be the channel charge and Q_{CAP} be the charge across all the 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 2.4 (b)) is constructed in SPICE to solve for Ψ_{CH}. Note that a similar solver was used in the Stanford CNFET model [29, 30]. Next, we derive Q_{CH} and Q_{CAP}.

2.3.1.3 Finding Channel Charge Q_{CH}

Q_{CH} is derived from carrier density. Electron density n_α in subband α is given by (2.5). Here, $f(E)$ given by (2.6) is the Fermi-Dirac distribution function, and $D_\alpha(E)$ given by (2.7) is the density of states (DOS) in a GNR based on [25]. 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 (2.8) [25], k is Boltzmann's constant, T is temperature, and $a = 2.46 \times 10^{-10}$ meters is the lattice constant.

$$n_\alpha = \int_0^\infty f(E) \cdot D_\alpha(E) \mathrm{d}E \tag{2.5}$$

$$f(E) = \frac{1}{1 + e^{\frac{E-E_F}{kT}}} \tag{2.6}$$

$$D_\alpha(E) = \frac{2\sqrt{M_\alpha}}{\pi\hbar} \cdot \frac{\varepsilon_\alpha + E}{\sqrt{\varepsilon_\alpha E(E + 2\varepsilon_\alpha)}} \tag{2.7}$$

$$M_\alpha = \frac{2\hbar^2\varepsilon_\alpha}{3a^2t^2 \cdot \cos(\frac{\pi\alpha}{N+1})} \tag{2.8}$$

The integral in (2.5) has no closed-form solution. A closed-form approximation was derived in [25] by approximating $f(E)$ with 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 (2.5) cannot be solved directly, we approximate $f(E)$ with an exponential function when $E_F - E_C < 0$, a step function when $E_F - E_C > 2kT$, and a smoothing function when $E_F - E_C$ is in between.

2.3.1.3.1 Exponential Approximation Here, $f(E)$ is approximated by a decaying exponential function $f'(E)$ [43] as follows:

$$f(E) \sim f'(E) = f(0) \cdot e^{\frac{-E}{\beta kT}} \tag{2.9}$$

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

$$\beta(E_{FC}) = \frac{3}{\ln f(-E_{FC}) \cdot \left[1 + \exp(\frac{3kT - E_{FC}}{kT})\right]} \tag{2.10}$$

where $E_{FC} = E_F - E_C$. Since $E_C = 0$, $E_{FC} = E_F$. 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

$$n_{\alpha,exp}(E_{FC}) = \frac{\sqrt{M_\alpha(\beta kT)^3}\left(1 + \frac{2\varepsilon_\alpha}{\beta kT}\right)}{2\pi\hbar\varepsilon_\alpha} \cdot e^{\frac{E_{FC}}{\beta kT}} \tag{2.11}$$

2.3.1.3.2 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

$$
\begin{aligned}
n_{\alpha,step}(E_{FC}) &= \int_0^{E_F} 1 \cdot D_\alpha(E)\mathrm{d}E \\
&= \frac{2\sqrt{M_\alpha}}{\pi\hbar}\sqrt{\max\left(\frac{E_{FC}(E_{FC}+2\varepsilon_\alpha)}{\varepsilon_\alpha}, 0\right)}
\end{aligned} \tag{2.12}
$$

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

2.3.1.3.3 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 (2.13), where m is the relative weight defined in (2.14). To make the expressions more general, E_{FC} is introduced. 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 $E_{FC} > 2kT$.

$$
n_\alpha(E_{FC}) = m \cdot n_{\alpha,exp}(E_{FC}) + (1-m)n_{\alpha,step}(E_{FC}) \tag{2.13}
$$

$$
m = \frac{1}{1 + e^{\frac{3(E_{FC}-kT)}{kT}}} \tag{2.14}
$$

The effectiveness of (2.13) 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 2.5, where *Numerical* was obtained by evaluating the integral in (2.5), *Boltzmann* was obtained from expressions in [25], *Exponential* was obtained from (2.11), and *Combined* was obtained from (2.13). 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.

2.3.1.3.4 Computing Channel Charge Q_{CH} Total channel charge Q_{CH} is derived by analyzing the band diagram. Figure 2.6 (a) 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 $V_{DS} > 0$,

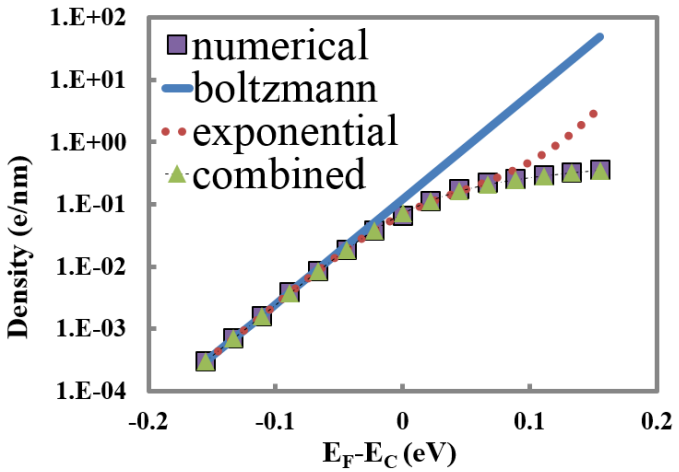

Figure 2.5: Charge density n_α vs. $E_F - E_C$ in the case of $\varepsilon_\alpha = 0.3$ eV ($N = 12$).

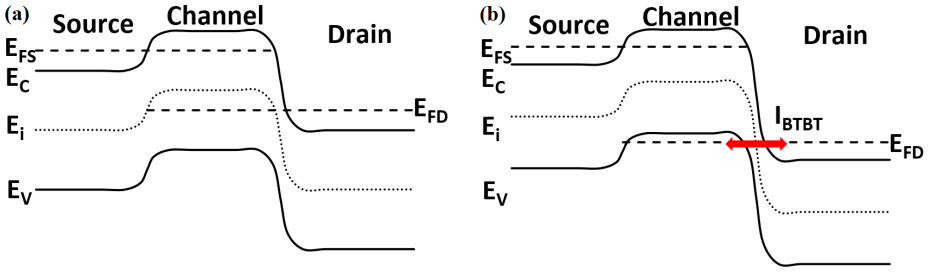

Figure 2.6: Typical band diagrams of a GNRFET **(a)** under low V_{DS} and **(b)** under high V_{DS}. In (b), the band bending is high enough to induce band-to-band tunneling current (I_{BTBT}).

$E_{FD} < E_{FS}$. 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 in the channel ($E_{V,CH}$), and holes tunnel from the drain into the channel. The tunneling probability $Tr(\Psi_{CH,D})$ is given by (2.15), 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. The equation takes the form of a sigmoid function that smoothly transitions from 0 to 1 at the onset where band-to-band tunneling starts taking place. 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 $\gamma = 1/6$. The equation and the parameters are obtained by experimenting with channel charge data extracted from NanoTCAD ViDES [23] that could not be accounted for by the majority carrier contribution alone.

$$Tr(\Psi_{CH,D}) = \left(1 + e^{\frac{(2+\eta_{0.5})\varepsilon_\alpha - q\Psi_{CH,D}}{\gamma\varepsilon_\alpha}}\right)^{-1} \tag{2.15}$$

The final expression of Q_{CH} (2.16) 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 $E_C = \varepsilon_\alpha - q\Psi_{CH}$, and the valence band $E_V = -\varepsilon_\alpha - q\Psi_{CH}$. Also, the Fermi level at source/drain equals to the applied voltage. Thus, $E_{FS} - E_C = -qV_S - (\varepsilon_\alpha - q\Psi_{CH})$.

$$\begin{aligned} Q_{CH}(\Psi_{CH}, V_D, V_S) = \frac{qL_{CH}}{2} \sum_\alpha [& -n_\alpha(q\Psi_{CH} - \varepsilon_\alpha - qV_S) \\ & -n_\alpha(q\Psi_{CH} - \varepsilon_\alpha - qV_D) \\ & + Tr(\Psi_{CH,D}) \cdot p_\alpha(qV_D - q\Psi_{CH} - \varepsilon_\alpha)] \end{aligned} \tag{2.16}$$

2.3.1.4 **Finding** Q_{CAP}

Q_{CAP} (2.17) 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 (2.18). $C_{DIBL,D}$ and $C_{DIBL,S}$ are effective capacitors that model the drain-induced barrier-lowering (DIBL) effect. They were empirically modeled as $0.15C_{G,CH} \cdot Tr$ 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 gate oxide material.

$$\begin{aligned} Q_{CAP} &= C_{G,CH}(V_G - V_{FB} - \Psi_{CH}) \\ &+ C_{SUB,CH}(V_{SUB} - V_{FB} - \Psi_{CH}) \\ &+ C_{DIBL,D}(V_D - \Psi_{CH}) + C_{DIBL,S}(V_S - \Psi_{CH}) \end{aligned} \tag{2.17}$$

$$C_{G(SUB),CH} = \frac{5.55 \times 10^{-11} \epsilon_r L_{CH}}{\left(1 + \frac{1.5T_{ox}}{W_G}\right) \ln\left(\frac{5.98W_{CH}}{0.8T_{ox}}\right)} \tag{2.18}$$

2.3.1.5 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 voltage-controlled capacitors by defining the charge equation.

2.3.1.6 Current Modeling

Given Ψ_{CH}, the electron current I_e is computed from (2.19) based on the Landauer-Büttiker formalism [25, 28]. Here, h is Plank's constant, and $f(\cdot)$ is the Fermi-Dirac distribution. $T(E)$ is the tunneling probability. In the case of thermionic conduction, $T(E) = 1$. $E_{FD,C}$ ($E_{FS,C}$) is the difference between the E_C in the channel and E_F on the drain (source) side, as in Figure 2.6. 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 [43], (2.19) can be evaluated analytically, which yields (2.20). In an N-type GNRFET, $I_{DS} = I_e$, while in a P-type GNRFET, $I_{DS} = I_h$, which is obtained similarly.

$$I_e = \frac{2q}{h} \sum_{\alpha} \int_0^{\infty} T(E) \left[f(E - E_{FS,C}) - f(E - E_{FD,C}) \right] \mathrm{d}E \tag{2.19}$$

$$\begin{aligned} I_e(\Psi_{CH}, V_D, V_S) = \frac{2qkT}{h} \sum_{\alpha} \Bigg[& \ln\left(1 + e^{\frac{q(\Psi_{CH} - V_S) - \varepsilon_\alpha}{kT}}\right) \\ & - \ln\left(1 + e^{\frac{q(\Psi_{CH} - V_D) - \varepsilon_\alpha}{kT}}\right) \Bigg] \end{aligned} \tag{2.20}$$

2.3.1.7 Considering Band-to-Band Tunneling

When V_{DS} is high enough to incur significant band bending, the band-to-band tunneling (BTBT) effect starts to occur in the channel, contributing

to additional current. Figure 2.6 (b) shows a band diagram with significant band bending such that BTBT occurs. It sometimes contributes to leakage current [44]. We took a similar approach to the work of [29] and [45] of CNFETs to model the BTBT current in GNRFETs, which is also based on the Landauer-Büttiker equation (2.19), with $T(E)$ representing the BTBT probability. The resulting BTBT current becomes (2.21). T_{BTBT} is the tunneling probability, expressed as (2.21). $\Psi_{bi} = 0.4$ is the built-in potential. l_{relax} is the distance for the potential drop across the drain-channel junction to relax.

$$I_{BTBT} = \frac{4qkT}{h} \sum_{\alpha} T_{BTBT} \left[\ln \frac{1 + e^{\frac{q(V_D - V_S) - \varepsilon_\alpha - q\Psi_{bi}}{kT}}}{1 + e^{\frac{-\varepsilon_\alpha - q\Psi_{bi}}{kT}}} \right]$$

$$T_{BTBT} = \frac{-\pi^3 \cdot \sqrt{M_\alpha} (0.5\varepsilon_\alpha)^{1.5}}{9\hbar\sqrt{q} \left(\frac{q(V_D - \Psi_{CH} + V_S + \Psi_{bi})}{l_{relax}} \right)}$$

2.3.1.8 **Considering Line Edge Roughness**

To date, fabrication technology cannot produce GNRs with perfectly smooth edges. The uneven edges result in a phenomenon called *line edge roughness*, which affects the properties of GNRs. Line edge roughness is characterized by p_r, the probability that any atom on the edges of a GNR is removed, as in [23]. The removal of atoms has two effects: (1) Subbands (2.1) varies throughout the channel as N is no longer constant. (2) Ballistic transport is disrupted. These effects strongly depend on which atoms are removed [23]; 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 line edge roughness on the circuit level.

To model the varying width, we introduce the concept of an effective subband edge $\varepsilon_{\alpha,eff}$ given by (2.21), where $\varepsilon_{\alpha,N}$ is the ε_α for a given N. In a unit segment of GNR, there are eight atoms (shown as red dots in Figure 2.1) 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 (2.21). The scattering coefficient A is introduced to account for the current reduction due to disrupted ballistic transport. It is empirically modeled as (2.22).

$$\varepsilon_{\alpha,eff} = (1-p_r)^8 \varepsilon_{\alpha,N} + 1(1-(1-p_r)^8)\varepsilon_{\alpha,N-1} \tag{2.21}$$

$$A = 0.98(1-4p_r)^6 + 0.02 \tag{2.22}$$

The current equations derived in Sections 2.3.1.6 and 2.3.1.7 assume ballistic transport and are denoted as $I_{bal} = I_e + I_{BTBT}$ combined. Current with line edge roughness present, I_{rough}, is derived from I_{bal} and is modeled as follows:

$$I_{rough} = A \cdot I_{bal}(\varepsilon_{\alpha,eff}) \tag{2.23}$$

2.3.2 Full GNRFET Model, Vias, and Interconnects

Figure 2.4 (c) shows the SPICE implementation of a GNRFET with four parallel GNRs equivalent to that in Figure 2.2. Each transistor highlighted in red corresponds to an individual GNR, which is modeled by the circuit in Figure 2.4. C_{GD} and C_{GS}, given by (2.24), are parasitics introduced by the fringing fields between the gate and the reservoirs. They are modeled empirically based on data from FastCap [46]. When two GNRFETs are connected, graphene-metal contact resistance exists externally between gates and drains/sources.

$$C_{GD} = C_{GS} = 1.26 \times 10^{-10} W_G (0.8 - 0.2T_{ox} + 0.015T_{ox}^2) \tag{2.24}$$

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 [26, 27]. 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 [42].

2.3.3 Discussion on Model Empiricism

In summary, empirical parameters occur in our model in the following situations:

1. Smoothing functions for the transitions in (13), (15).

2. Capacitance equations (18) and (26). As GNRFETs do not have a simple parallel plate capacitor structure, we resorted to empirical modeling based on data collected from FastCap.

3. Capacitance for modeling DIBL.

4. Line edge roughness scattering coefficient.

It is possible to replace the empirical equations and parameters if more accurate description is found. In particular, should the device geometry differ from our default design by a great deal, the users may use FastCap to obtain new capacitance values that suit their design better.

2.3.4 Note on Gummel Symmetry Test

Gummel Symmetry Test (GST) is a common test on compact models. The transistor is biased with a fixed V_{GS} and V_{BS}, and V_D and V_S are set to V_x and $-V_x$, respectively. Then, the drain-source current I is measured under a sweep of V_x. For a model to pass GST, it needs to satisfy two criteria: (1) the symmetry on the drain and the source of the transistor, i.e. $I(V_x)$ is an odd function and (2) no singularity occurs at $V_x = 0$ (by checking the continuity of higher-order derivatives). GST is particularly important for a compact model to be used in the distortion analysis in analog circuits, but it is not a strict requirement for other applications [47].

Figure 2.7 shows the current and derivatives vs. V_x relationship near $V_x = 0$. It shows that despite the continuity, the current is not symmetric, i.e., the drain and source in our model are not interchangeable. This is due to the fact that some of the equations we developed are not identical for drain and source. In particular, our modeling of DIBL and BTBT is based on the assumption that $V_D > V_S$ and that these effects are much more significant on the drain side. A compact model of CNFET similar to ours also does not feature interchangeable drain and source [29, 30]. Despite this limitation, the model is perfectly fine for digital circuits designed under the prevalent complementary-symmetry metal-oxide-semiconductor (CMOS) style.

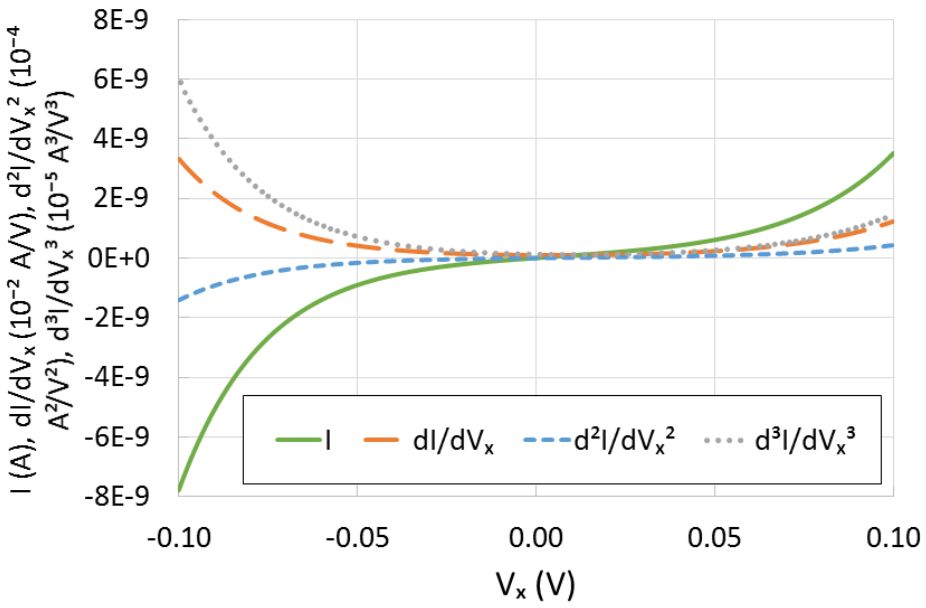

Figure 2.7: Gummel symmetry test of our GNRFET model, showing I, $\frac{dI}{dV_x}$, $\frac{d^2I}{dV_x^2}$, and $\frac{d^3I}{dV_x^3}$ vs. V_x from V_x = -0.1 to 0.1 V.

2.4 Experimental Results

The equivalent circuit model and all equations in Section 2.3 were implemented in HSPICE as a *subckt*. The work [48] analyzes how a generic CNT/graphene transistor *subckt* is processed in SPICE. In Section 2.4.1, the compact model is validated against numerical simulation in NanoTCAD ViDES [23, 35] and compared with measurement data from fabricated GNRFETs. With the accuracy of our SPICE model thoroughly validated, we can proceed with SPICE simulations of GNRFET and GNR-based circuits. This gives insightful information on how GNR-based circuits would perform once fabrication techniques become mature. In Sections 2.4.2, 2.4.3 and 2.4.4, we implemented digital logic gates with our GNRFET SPICE model, performed transistor- and circuit-level analyses, and compared them with those implemented in Si-CMOS 16-nm high-performance (HP) and low-power (LP) libraries from Predictive Technology Models (PTM) [49]. In particular, Section 2.4.4 focuses on the effect of process variation based on Monte Carlo simulations.

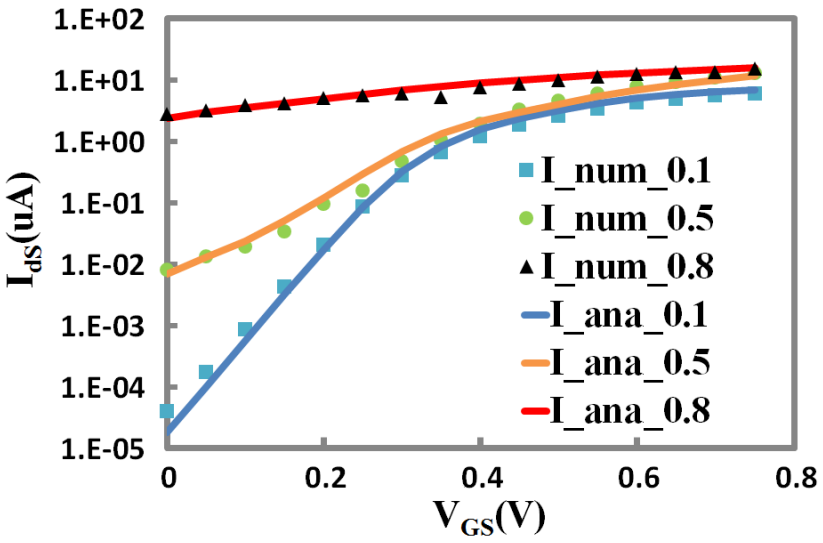

Figure 2.8: I_{DS} vs. V_{GS} with V_{DS} $0.1, 0.5, 0.8$ V in an N-type GNRFET.

2.4.1 Transistor Model Validation

2.4.1.1 **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 2.8, 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.

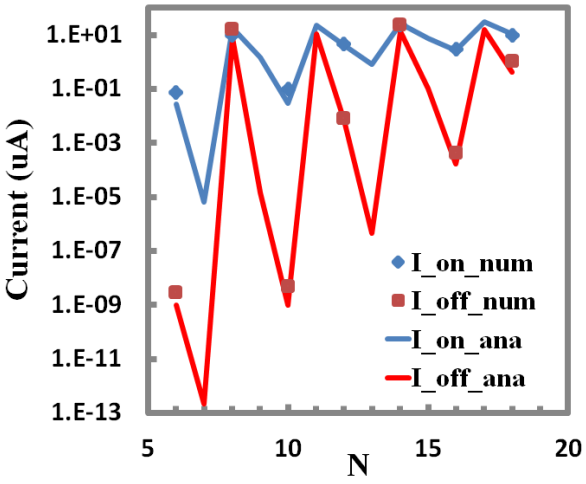

Figure 2.9: (I_{on} and I_{off}) vs. N. Note that ViDES only supports even N.

2.4.1.2 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 2.9 shows the effect of N. Our model tracks the periodic effect on band gaps discussed in [19]. 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 2.10 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 2.11 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 2.12 shows the effect of line 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 line edge roughness is increased.

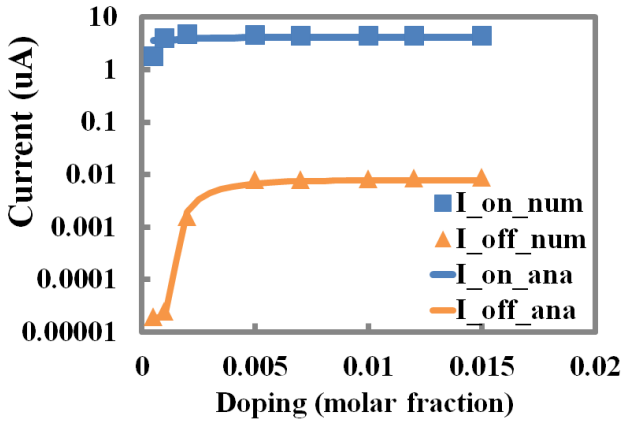

Figure 2.10: (I_{on} and I_{off}) vs. f_{dop}, doping fraction in reservoirs.

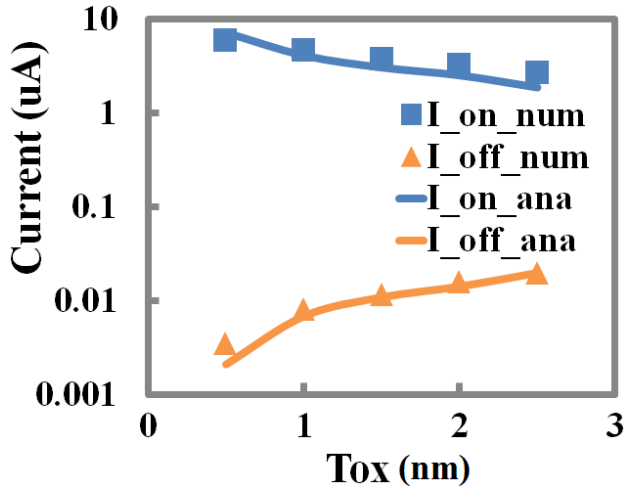

Figure 2.11: (I_{on} and I_{off}) vs. T_{ox}, oxide thickness.

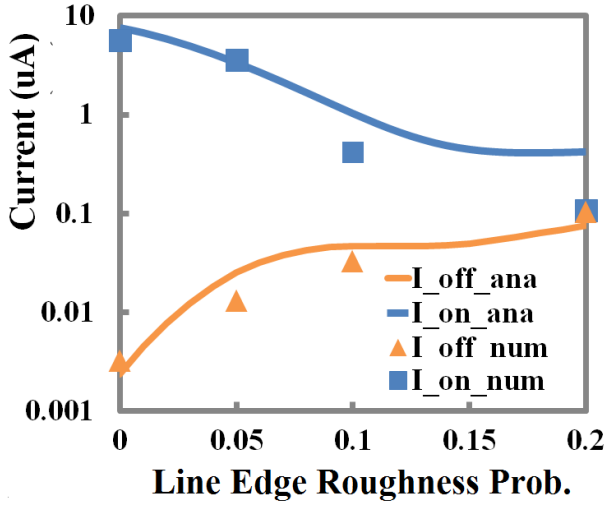

Figure 2.12: (I_{on} and I_{off}) vs. p_r, line edge roughness probability.

2.4.1.3 Comparison with Measurement Data from Fabricated GNRFETs

Among all existing work on fabricated GNRFETs, the single-layer SB-type GNRFET in [21] with W $\simeq$ 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 40 V) [36, 37, 38, 40]. In [24], 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 fabricated device in [21] and an $N = 16$ MOS-type GNRFET with $p_r = 0.1$ in order to account for the line 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) Schottky barriers exist in the fabricated device but not in MOS-type GNRFETs. (2) 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 line edge roughness. (4) Our model assumes ballistic transport, while the fabricated GNRs in [21] have lengths > 100 nm, greater than the mean free path. (5) Other experimental settings and nonidealities that are unclear to us.

Note that in [21], an *edge scattering probability* 20% was calculated, which is the probability of back-scattering that depends on the edge quality, while we used the probability of an atom on the edge being missing as the *line edge roughness probability*. This terminology was defined in the open-source NanoTCAD ViDES [23, 35], which we used to produce validation data points, and therefore we adopted this probability in our model. We did not find a straightforward relation between the two probabilities, but we were able to compare our model with the data reported in [21] by using a line edge roughness of 10%. Furthermore, as indicated in Figure 2.12, a line edge roughness of 20% leads to a poor I_{on}/I_{off} ratio of less than 10, which is not the case in the work of [21] where the I_{on}/I_{off} is greater than 10^6. Therefore, we believe simulating a 10% line edge roughness is sufficient in the experiments in all the following subsections.

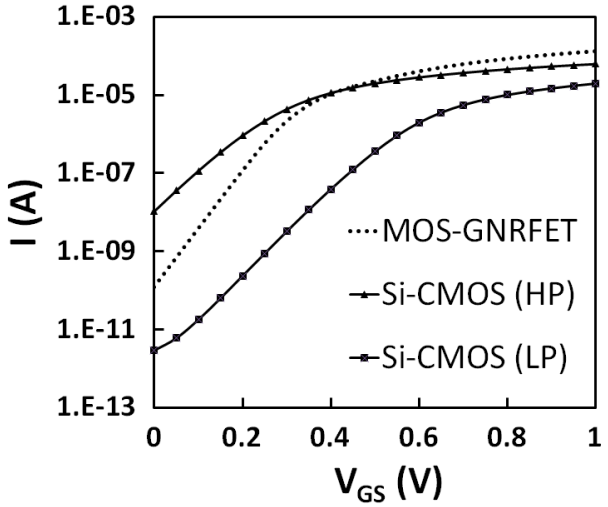

Figure 2.13: I_{DS} vs. V_{GS} for MOS-GNRFET, 16-nm high-performance (HP) Si-CMOS, 16-nm low-power (LP) Si-CMOS, respectively.

2.4.2 Transistor-Level Properties

In this section, we review the transistor-level characteristics of MOS-GNRFETs. Based on the explorations in [26, 27] and Section 2.4.1, MOS-GNRFETs work well under a low V_{DD} around 0.5 V. Therefore, we choose a nominal $V_{DD} = 0.5$ V in all the following experiments unless otherwise stated. Also based on the explorations in Section 2.4.1, we choose the design parameters as follows: $N = 12$, $f_{dop} = 0.001$, $T_{ox} = 0.95$ nm, and $L_{CH} = 16$ nm.

Figure 2.13 shows the I-V curves of MOS-GNRFET as well as the 16-nm Si-CMOS (HP) and 16-nm Si-CMOS (LP) transistors from PTM for comparison. The transistor dimensions of the GNRFETs are scaled to match the PTM libraries. Overall, Si-CMOS (HP) has the highest current, and the Si-CMOS (LP) has the lowest. MOS-GNRFET and Si-CMOS (LP) have better I_{on}/I_{off} ratios than Si-CMOS (HP).

Table 2.1 shows the subthreshold swing S and I_{on}/I_{off} ratio of each device under respectively chosen V_{DD}. It is shown that ideal MOS-GNRFETs have the lowest subthreshold swing (66.67 mV/dec)and the highest I_{on}/I_{off} ratio ($1.81\ ^{5}$). However, as line edge roughness comes into play, the subthreshold swing increases to 140.85 mV/dec, and the I_{on}/I_{off} ratio drops to 98.5. In other words, the transistor characteristics become comparable or even worse than Si-CMOS.

Table 2.1: Transistor Properties

Device	p_r	S (mV/dec)	I_{on}/I_{off}	V_{DD} (V)
Si-CMOS (HP)	–	93.46	3.49E+03	0.7
Si-CMOS (LP)	–	86.96	5.12E+06	0.9
MOS-GNRFET	0	66.67	1.81E+05	0.5
	0.05	83.33	3.69E+03	0.5
	0.1	140.85	9.85E+01	0.5

Subthreshold swing and I_{on}/I_{off} ratio of each device. For GNRFETs, devices of different line edge roughness (p_r) are listed as well.

2.4.3 Circuit-Level Evaluation

We performed HSPICE DC and transient analyses on digital circuits defined in SPICE netlists. We used an input slew of 10 ps and an output load of 1 fF. We first evaluated the noise margin of an inverter. Then, we evaluated delay and power of a buffer chain under various supply voltages to understand the power-delay trade-off. Next, the buffer chain is simulated with various design parameters such as N, f_{dop}, T_{ox}, and L_{CH} to evaluate the impact of process variation. Following is a thorough comparison performed on a set of digital benchmark circuits implemented with MOS-GNRFET and Si-CMOS under their respective optimal settings. Finally, we performed Monte Carlo simulations to investigate the impact of process variation on GNRFET circuits.

2.4.3.1 Noise Margin Analysis

Figure 2.14 shows the voltage transfer curves of inverters built with MOS-GNRFETs with different settings, namely, ideal MOS-GNRFETs (with no graphene-metal contact resistance), MOS-GNRFETs with graphene-metal contact resistance, and MOS-GNRFETs with graphene-metal contacts and line edge roughness ($p_r = 5\%$ and $p_r = 10\%$, respectively). V_{in} and V_{out} are the input and output voltages of the inverter, respectively. All inverters have full voltage swings. The ranges of V_{in} that result in correct operations are indicated by V_{IL} and V_{IH}, the maximum voltage for a valid low input and the minimum voltage for a valid high input, respectively. V_{IL} and V_{IH} are specifically measured as the points with slopes equal to 1. Table 2.2 shows

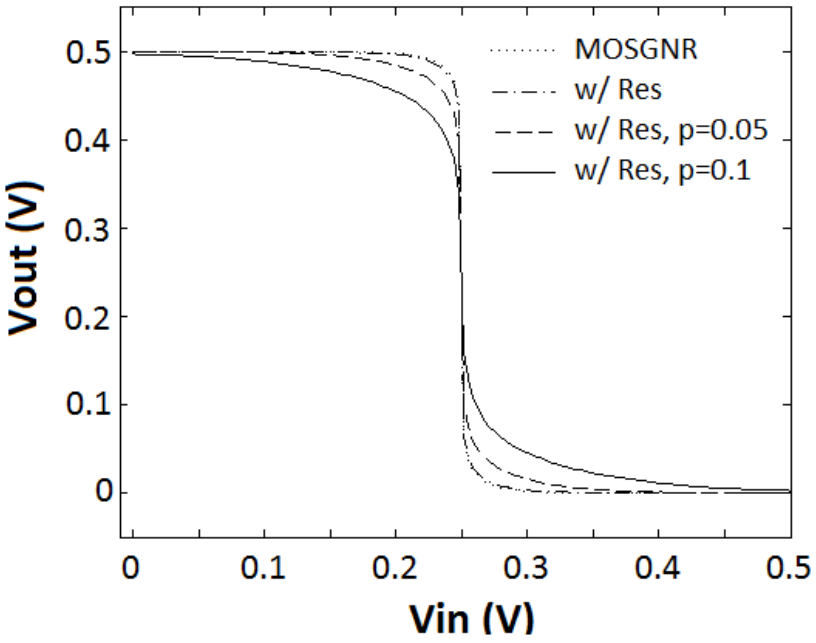

Figure 2.14: Voltage transfer curves of MOS-GNRFETs with different settings. Both inverters have full voltage swings. For the MOS-GNRFET inverter, $V_{IL} = 0.23$ V, $V_{OL} = 0.27$ V, and the noise margin is 92% of V_{DD}. Note that the curves of MOS-GNRFET with no line edge roughness, whether with contact resistance or not, almost overlap with each other.

the V_{IL}, V_{IH}, and the normalized voltage range of correct inverter operation. The ideal MOS-GNRFET inverter has a sharp voltage transfer curve, which makes it more robust, as V_{out} almost stays the same as V_{in} approaches V_{IL} or V_{IH}. Contact resistance on MOS-GNRFETs with $p_r = 0\%$ does not have much impact on the voltage transfer curve. On the other hand, line edge roughness significantly reduces the region of correct operation.

2.4.3.2 Impact of Supply Voltage

We evaluated the delay and power of a seven-stage, fanout-of-four buffer chain under various supply voltages to understand the power-delay trade-off. The buffer chain was implemented in Si-CMOS (LP), Si-CMOS (HP), ideal MOS-GNRFETs (with no graphene-metal contact resistance), MOS-GNRFETs with graphene-metal contact resistance, and MOS-GNRFETs with graphene-metal contacts and line edge roughness. We implemented Si-CMOS with the 16-nm HP and LP libraries from PTM, and implemented

Table 2.2: Noise Margin of Inverters

Device	p_r	V_{IL} (V)	V_{IH} (V)	NM/V_{DD}
MOS-GNRFET	0	0.2336	0.2664	0.9344
MOS-GNRFET w/ Res	0	0.2328	0.2672	0.9312
	0.05	0.2255	0.2746	0.9018
	0.1	0.2151	0.2849	0.8604

V_{IL}, V_{IH}, and the normalized voltage range of correct operations of inverters (indicated by noise margin divided by V_{DD}) measured on MOS-GNRFET inverters with different settings.

MOS-GNRFETs with our SPICE model. The minimum-size MOS-GNRFET is set to have six ribbons in order to match the dimensions of Si-CMOS. Graphene-metal junctions are present in circuit layouts, as discussed in Section 2.2.2, and they are modeled with a 20-kΩ resistor by assuming a 50-nm via width [42]. Limitations on fabrication techniques contribute to line edge roughness. We simulated the cases of $p_r = 5\%$ and 10%. Considering graphene-metal contacts and line edge roughness makes our simulations closer to reality. The ideal MOS-GNRFET, although not realistic, gives an upper bound on circuit performance.

Figure 2.15 shows the impact of supply voltage V_{DD} on the circuit performance. The metrics reported are delay, dynamic power, leakage power, total power, and energy-delay product (EDP). Delay is measured as the maximum propagation delay from a series of random input vectors. Dynamic power is measured based on the assumption that the circuits operates at a frequency based on the maximum propagation delay. Graphene-metal contact resistance and line 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.

2.4.3.3 Impact of Design Parameters

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 Section 2.4.3.2 by varying these design parameters to find their respective

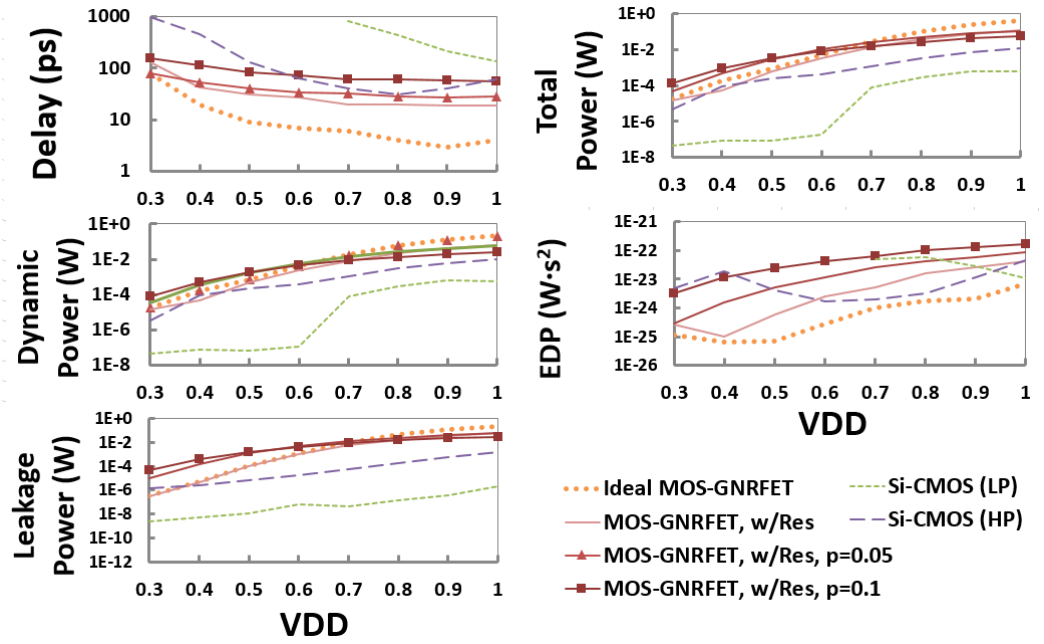

Figure 2.15: Delay, dynamic power, leakage power, total power, and EDP vs. V_{DD}.

impacts on the circuit level.

Figure 2.16 shows the impact of ribbon width N (W_{CH}). The results are consistent with the periodic band gaps in terms of N as reported in [19]. For examples, $N = 3p + 2$ (8, 11, 14) gives a small band gap, resulting in almost equally high I_{on} and I_{off}, corresponding to low delay and high power. $N = 3p+1$ (10, 13, 16) gives the largest band gap with low I_{on} and very low I_{off}, resulting in the highest I_{on}/I_{off} ratio. Therefore, the power, especially the leakage power, is the lowest. $N = 3p$ (9, 12, 15) gives a moderate band gap, and the delay and power performance is between the other two cases, with EDP being the lowest. Under the influence of line edge roughness, the effective band gaps fall between the band gaps corresponding to an effective width N_{eff} between N and $N-2$, making the periodic effect not so significant. Also, the scattering effect causes the current to drop. As a result, delay is generally higher and power is generally lower compared to the ideal cases. It is noteworthy that the I_{on}/I_{off} ratio in the case of $N = 3p + 2$ is extremely small and results in poor transistor operation. For example, Figure 2.17 shows an inverter with MOS-GNRFET of $N = 14$, with output voltage ranging from 0.074 V to 0.426 V, not reaching full swing. V_{IL} and V_{IH} in this case are 0.2801 V and 0.2199 V, respectively, making the noise margin

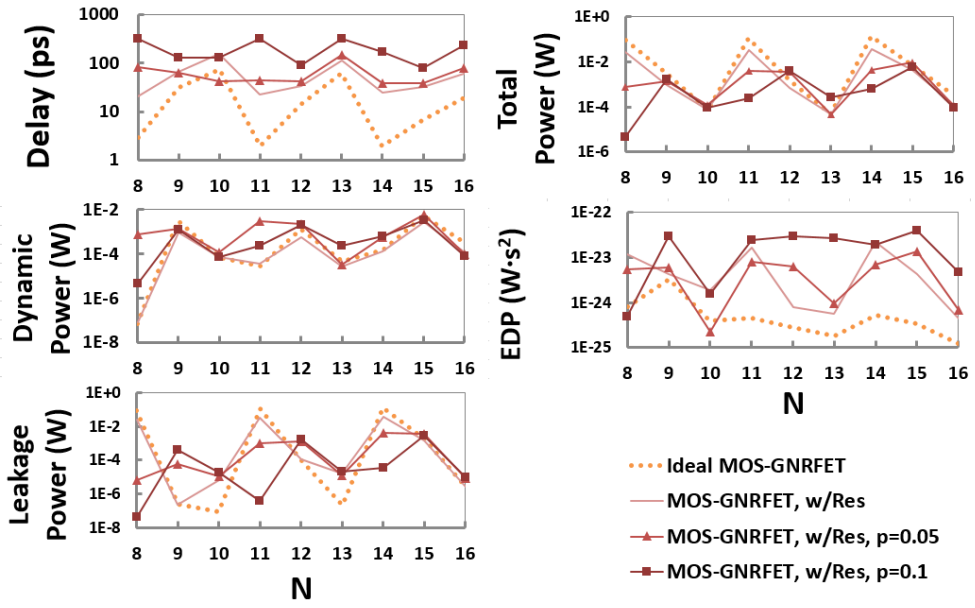

Figure 2.16: Delay, dynamic power, leakage power, total power, and EDP vs. N.

88% of V_{DD}. This observation is consistent with the transistor-level NEGF simulation results reported in [23]. In short, variation in ribbon width can cause significant performance degradation and is a possible major drawback for GNRFETs.

The effects of other parameters, L_{CH}, T_{ox}, and f_{dop}, are shown in Figure 2.18. 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. In general, changes in T_{ox} or L_{CH} affect delay, power, or EDP only within one order of magnitude. On the other hand, line edge roughness has a very high impact on delay and power. These observations are consistent with our model.

2.4.3.4 Performance Comparison between GNRFET and Si-CMOS

We compared delay and power performance on a set of digital circuits, implemented with Si-CMOS and MOS-GNRFETs, respectively.

We first evaluated the delay and power of basic logic gates such as *inv*,

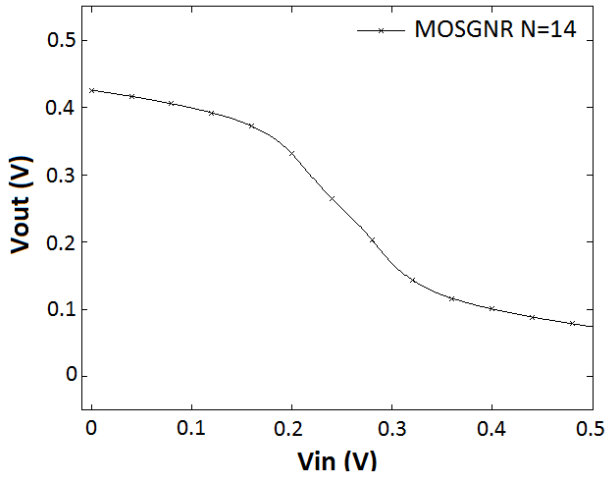

Figure 2.17: Voltage transfer curve of an inverter with MOS-GNRFET of $N = 14$, showing poor performance.

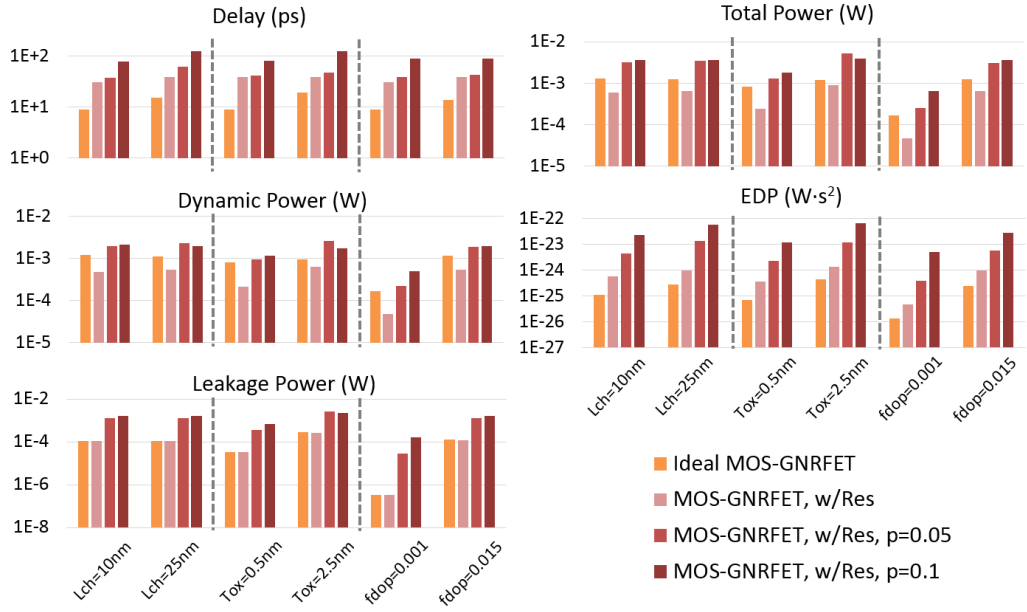

Figure 2.18: Delay, dynamic power, leakage power, total power, and EDP vs. L_{CH}, T_{ox}, and f_{dop}.

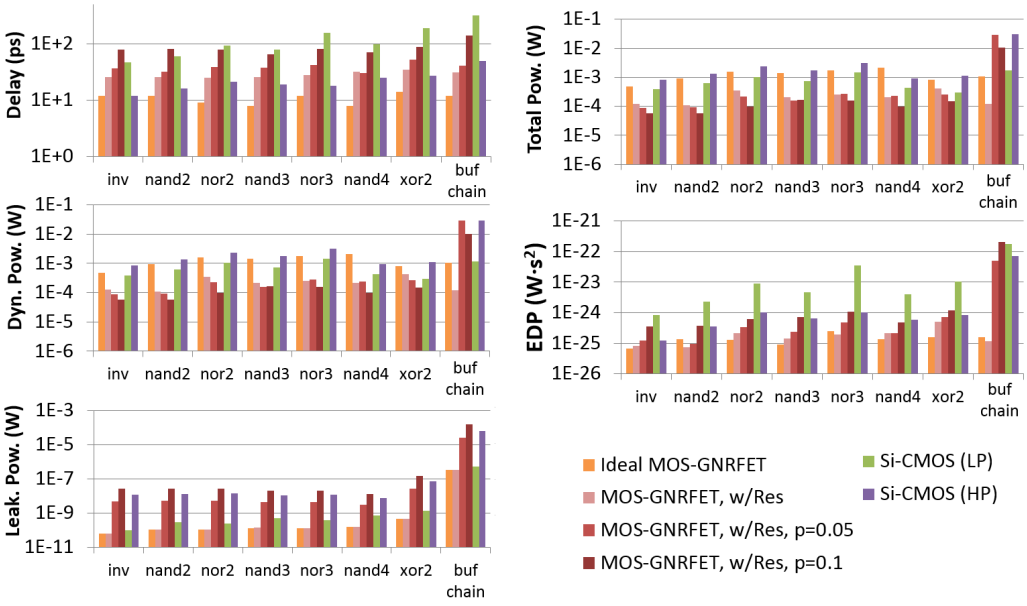

Figure 2.19: Simulation of basic logic gates, reporting delay, dynamic power, leakage power, total power, and EDP.

nand2, *nor2*, *nand3*, *nor3*, *nand4*, *xor2*, and a seven-stage, fanout-of-four buffer chain, which were implemented in Si-CMOS, ideal MOS-GNRFETs (with no graphene-metal contact resistance), MOS-GNRFETs with graphene-metal contact resistance, and MOS-GNRFETs with graphene-metal contacts and line edge roughness. We implemented Si-CMOS with the 16-nm LP and HP libraries from PTM, and implemented MOS-GNRFETs with our SPICE model. The minimum-size MOS-GNRFET was set to have six ribbons in order to match the dimensions of Si-CMOS. Gate sizing was done to balance the pull-up and pull-down networks in the logic gates. Graphene-metal junctions are present in circuit layouts, as discussed in Section 2.2.2, and they are modeled with a 20-kΩ resistor by assuming a 50-nm via width. We simulated the cases of $p_r = 5\%$ and 10%. For Si-CMOS, V_{DD} was chosen as the nominal V_{DD} recommended by PTM, which is 0.9 V for LP and 0.7 V for HP. The V_{DD} of MOS-GNRFET was chosen to be 0.5 V, according to the exploration in Section 2.4.3.2. The doping fraction of MOS-GNRFET was chosen to be 0.001, according to the exploration in Section 2.4.3.3. Figure 2.19 shows the delay and power evaluation results of the basic logic gates.

Next, we simulated a set of benchmark circuits under four settings: MOS-GNRFET with graphene-metal contacts and $p_r = 0\%$, MOS-GNRFET with graphene-metal contacts and $p_r = 10\%$, Si-CMOS (LP), and Si-CMOS(HP).

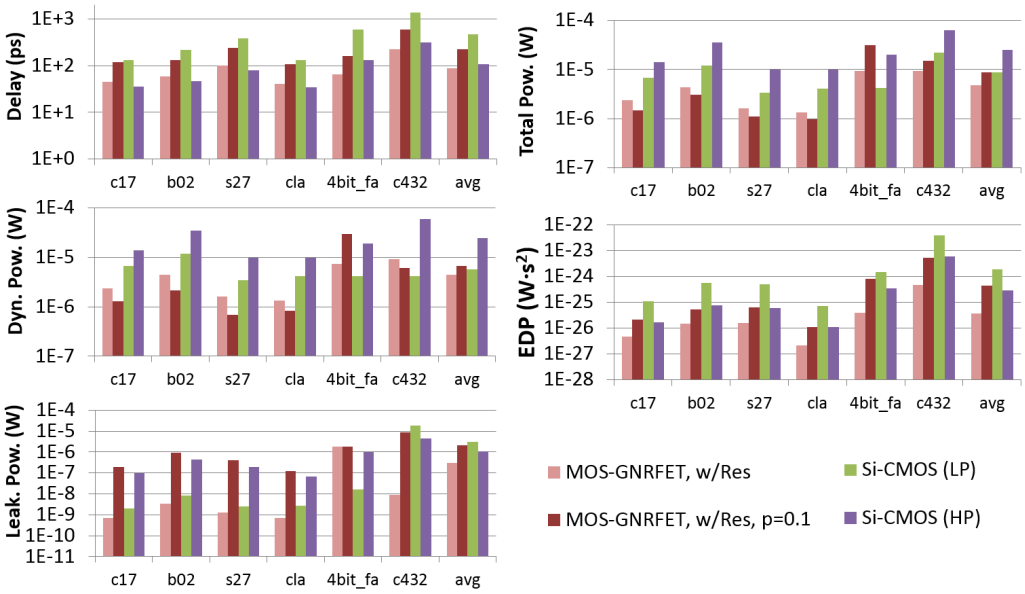

Figure 2.20: Simulation of benchmark circuits, reporting delay, dynamic power, leakage power, total power, and EDP.

Considering the fact that graphene-metal contact resistance cannot be avoided in the circuit architecture discussed in Section 2.2.2, we did not simulate ideal MOS-GNRFET without contact resistance here. Also, the performance of MOS-GNRFET with $p_r = 5\%$ normally lies between that of MOS-GNRFET with $p_r = 0\%$ and MOS-GNRFET with $p_r = 10\%$, so we did not simulate this case. The benchmark circuits we simulated include *c17* and *c432* from ISCAS '85, *b02* from ITC '99, *s27* from ISCAS '89, carry generator for the third bit of a carry look-ahead adder (*cla*), and a 4-bit full adder (*4bit_fa*). Sequential circuits *b02* and *s27* are converted into combinational circuits by the pseudo prime input method in order to have a consistent datapath delay definition compatible with other circuits. The summary of these circuits are presented in Table 2.3. We report delay, dynamic power, leakage power, total power, and EDP from circuits implemented in the four setups in Figure 2.20.

Based on results in Figures 2.19 and 2.20, line edge roughness plays a significant role in degrading the current in MOS-GNRFETs. As a result, Si-CMOS (HP) performs better in delay unless the MOS-GNRFET is ideal. In terms of dynamic power, MOS-GNRFET has lower consumption than Si-CMOS (HP) mostly due to lower V_{DD} and lower gate capacitance, and has comparable consumption to Si-CMOS (LP). In terms of leakage power for MOS-GNRFET, when a sufficiently high V_{DS} is applied, the confined states

Table 2.3: Benchmark Circuits

Circuit	# of Gates	# of PI	# of PO
c17	6	5	2
b02	25	3	1
s27	10	4	1
cla	5	7	1
4bit_fa	20	9	5
c432	153	36	7

Summary of numbers of gates, primary inputs (PI), and primary outputs (PO) of the benchmark circuits used in our experiments.

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 [25]. This is captured in equation (2.15), which describes an exponential relation between V_{DD} and the tunneling probability. First of all, when $V_{DD} = 0.7$ V, MOS-GNRFET has a higher leakage power than Si-CMOS (HP) shown in Figure 2.15. 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. Overall, ideal MOS-GNRFET has lower power consumption and comparable delay compared to Si-CMOS (HP), and it has lower delay and comparable power consumption compared to Si-CMOS (LP). In other words, ideal MOS-GNRFETs have advantages over both types of Si-CMOS transistors. However, MOS-GNRFET with nonidealities loses these benefits. In terms of EDP, ideal MOS-GNRFET performs the best, while MOS-GNRFET with $p_r = 10\%$ still has comparable EDP with Si-CMOS (HP).

In Figure 2.21, we compared the waveforms of two 11-stage ring oscillators, implemented with Si-CMOS (HP) and ideal MOS-GNRFET, respectively. Ideal MOS-GNRFET demonstrated a 5.5% higher frequency than Si-CMOS (HP), consistent with our observation in other circuits.

2.4.3.5 Discussion on Running Time

We performed the SPICE simulations on a machine with a 1.4 GHz AMD Opteron CPU. A typical transient analysis of a MOS-GNRFET inverter took 2.56 seconds, *c17* took 27,613.54 seconds ($\sim$ 7.67 hours), and *c432* took

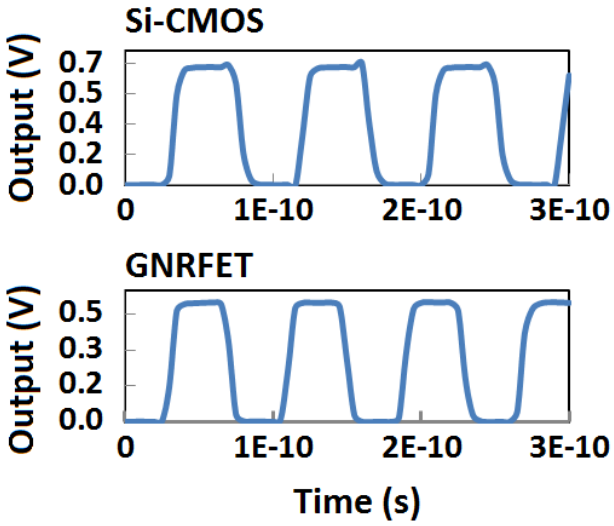

Figure 2.21: Simulation of ring oscillators in Si-CMOS and ideal MOS-GNRFET.

30,706.04 seconds ($\sim$ 8.53 hours). In comparison, *c432* implemented with the PTM Si-CMOS model took only 22.98 seconds. The slowdown mainly comes from the subcircuit implementation of our MOS-GNRFET model, especially with the solver side-circuit construct (Figure 2.4 (b)) to solve for the non-closed-form V_{CH}, as compared to the PTM models' utilization of SPICE's default transistor implementation. Still, the NEGF transistor simulation done by ViDES or similar tools takes hours to simulate one DC analysis of a single transistor, and our SPICE model greatly improves the running time such that one DC analysis finishes within a second, so we can scale up to circuit-level simulations with some loss of accuracy as indicated by the mismatch between our model and the ViDES simulations, shown in Figures 2.8-2.12.

2.4.4 Monte Carlo Simulation of Process Variation

To evaluate the impact of process variation on the circuit level, Monte Carlo (MC) simulations are necessary. When we evaluated the effects of varied design parameters in the previous sections, only one parameter was varied at a given time, which does not reflect the reality where more than one parameter may vary from the nominal value. Also, the sensitivity of each parameter to the resulting delay and power can be studied using MC simulations.

Table 2.4: Monte Carlo Distribution Parameters for Si-CMOS

Parameter	Distribution	Mean	Standard Deviation
Doping Level	Gaussian	2×10^{20}	2×10^{19}
Oxide Thickness T_{ox}	Gaussian	0.95 nm	0.1 nm
Channel Width W_{CH}	Gaussian	32 nm	3.2 nm

Table 2.5: Monte Carlo Distribution Parameters for MOS-GNRFET

Parameter	Distribution	Mean	Standard Deviation
Doping Level	Gaussian	2×10^{20}	2×10^{19}
Oxide Thickness T_{ox}	Gaussian	0.95 nm	0.1 nm
GNR Width N	Gaussian	12	1.2

HSPICE-based Monte Carlo simulations were run on the *c17* benchmark of ISCAS'85 for ideal MOS-GNRFET with contact resistance, MOS-GNRFET with contact resistance and $p_r = 0.1$, and 16-nm Si-CMOS (HP) from PTM.

A global distribution was defined for modeling the systematic gate-to-gate variation in parameters and a local distribution was used to model the random intra-gate variation among transistors [50]. The values and the distribution for each parameter are shown in Table 2.4 for Si-CMOS (HP) and in Table 2.5 for MOS-GNRFET, which are based on the assumptions made in [10, 51]. Note that for N, the numbers are rounded to integers.

2.4.4.1 Experiment Setup

Two random variables X and Y are independent if $P(X \wedge Y) = P(X)P(Y)$. The covariance between random variables X and Y is $COV(X, Y) = E[X - EX]E[Y - EY] = E[XY] - EXEY$. If $COV(X, Y) = 0$, X and Y are *uncorrelated*. If X and Y are independent then they are uncorrelated, but the converse is not true.

We utilize SPICE's built-in Monte Carlo feature in this set of experiments. All the design parameters used in the Monte Carlo simulations are generated as follows. Take one parameter, T_{ox} as an example. We first determine the average T_{ox} of each logic gate from a normal distribution (global distribution). Then, each transistor's actual T_{ox} is computed from the average T_{ox} of the gate added by a small amount of intra-gate variation, also from a normal distribution (local distribution). In our experiments and the example be-

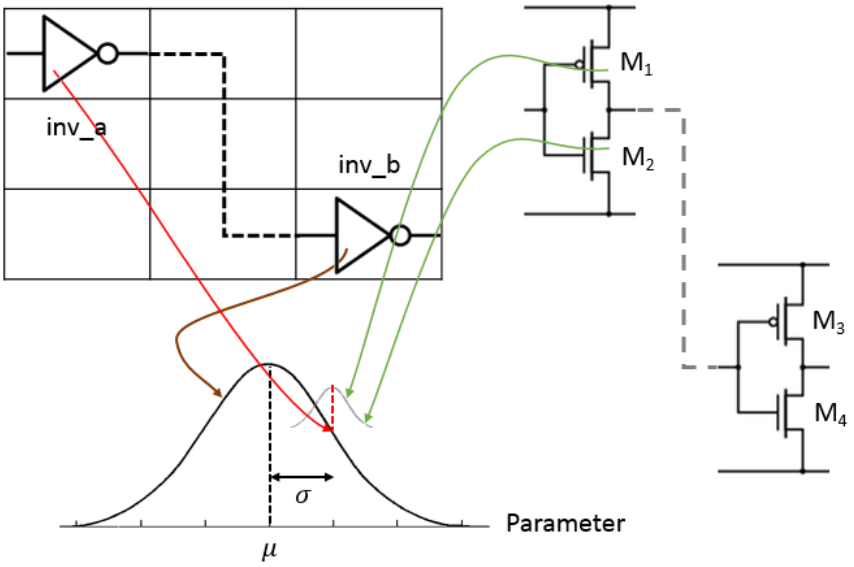

Figure 2.22: Example illustrating the setup of Monte Carlo simulation. The transistor parameters in *inv_a* and *inv_b* come from the same global distribution (indicated by the large normal distribution). Within a gate, e.g. *inv_a*, each transistor has an additional local variation (indicated by the small normal distribution) superimposed on the global distribution.

low, the local distribution's standard deviation is 0.1 of the global one. This is based on the fact that transistors within a logic gate are usually placed close to each other. Therefore, their design parameters should be highly correlated. This is to model intra-gate correlation in circuits despite the lack of actual placement information in HSPICE circuit implementation. Meanwhile, transistors not in the same logic gate have uncorrelated parameters. Despite not having placement information, a measurement study shows that within-die spatial correlation is almost nonexistent [52], and another study shows that a model with spatial correlation only has marginal effect on circuit optimization results over the model without [53].

We use a circuit that contain two *inv* gates, *inv_a* and *inv_b*, as shown in Figure 2.22, to demonstrate the computation of randomized design parameters. The inverters *inv_a* and *inv_b* are located far from each other in the circuit. Let X_A and X_B be random variables representing the average values of the design parameter in concern within *inv_a* and *inv_b*, respectively. Assume that the design parameter of each gate independently follows a normal distribution with mean μ and standard deviation σ under process

variation. We have $X_A \sim N(\mu, \sigma)$ and $X_B \sim N(\mu, \sigma)$. Since X_A and X_B are independent and identically distributed (i.i.d.), $COV(X_A, X_B) = 0$, and their covariance matrix is simply

$$\begin{bmatrix} VAR(X_A) & COV(X_A, X_B) \\ COV(X_B, X_A) & VAR(X_B) \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

Next, we add a small amount of intra-gate variation to each transistor. Let X_i be a random variable representing the design parameter value of transistor M_i. Let $\delta_i \sim N(0, 0.1\sigma)$ be a random variable representing the amount of intra-gate variation. We have

$$\begin{aligned} X_1 &= X_A + \delta_1 \\ X_2 &= X_A + \delta_2 \\ X_3 &= X_B + \delta_3 \\ X_4 &= X_B + \delta_4 \end{aligned}$$

The covariance between X_1 and X_2 is

$$\begin{aligned} COV(X_1, X_2) &= E[X_1 - EX_1]E[X_2 - EX_2] \\ &= E[X_1 X_2] - EX_1 EX_2 \end{aligned}$$

where

$$\begin{aligned} E[X_1 X_2] &= E[(X_A + \delta_1)(X_A + \delta_2)] \\ &= E[X_A^2] + E[X_A \delta_1] + E[X_A \delta_2] + E[\delta_1 \delta_2] \\ &= (Var(X_A) + E^2[X_A]) + 0 + 0 + 0 \\ &= \sigma^2 + \mu^2 \end{aligned}$$

and thus

$$\begin{aligned} COV(X_1, X_2) &= (\sigma^2 + \mu^2) - \mu^2 \\ &= \sigma^2 \end{aligned}$$

meaning they are correlated. Meanwhile, the covariance between X_1 and X_3

is

$$\begin{aligned} COV(X_1, X_3) &= E[X_1 - EX_1]E[X_3 - EX_3] \\ &= E[X_1X_3] - EX_1EX_3 \end{aligned}$$

Since $E[X_AX_B] = E[X_A]E[X_B]$ as X_A and X_B are independent, we have

$$\begin{aligned} E[X_1X_3] &= E[(X_A + \delta_1)(X_B + \delta_3)] \\ &= E[X_AX_B] + E[X_A\delta_3] + E[X_B\delta_1] + E[\delta_1\delta_3] \\ &= E[X_A]E[X_B] + 0 + 0 + 0 \\ &= \mu^2 \end{aligned}$$

and thus

$$\begin{aligned} COV(X_1, X_3) &= \mu^2 - \mu^2 \\ &= 0 \end{aligned}$$

meaning they are uncorrelated. Therefore, the covariance matrix of X_1, X_2, X_3, and X_4 is

$$\begin{bmatrix} VAR(X_1) & COV(X_1, X_2) & COV(X_1, X_3) & COV(X_1, X_4) \\ COV(X_2, X_1) & VAR(X_2) & COV(X_2, X_3) & COV(X_2, X_4) \\ COV(X_3, X_1) & COV(X_3, X_2) & VAR(X_3) & COV(X_3, X_4) \\ COV(X_4, X_1) & COV(X_4, X_2) & COV(X_4, X_3) & VAR(X_4) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & \sigma^2 & 0 & 0 \\ \sigma^2 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & \sigma^2 \\ 0 & 0 & \sigma^2 & \sigma^2 \end{bmatrix}$$

In this example, we show that our randomization method results in positively correlated transistors in the same logic gate, while the transistors in different logic gates are uncorrelated.

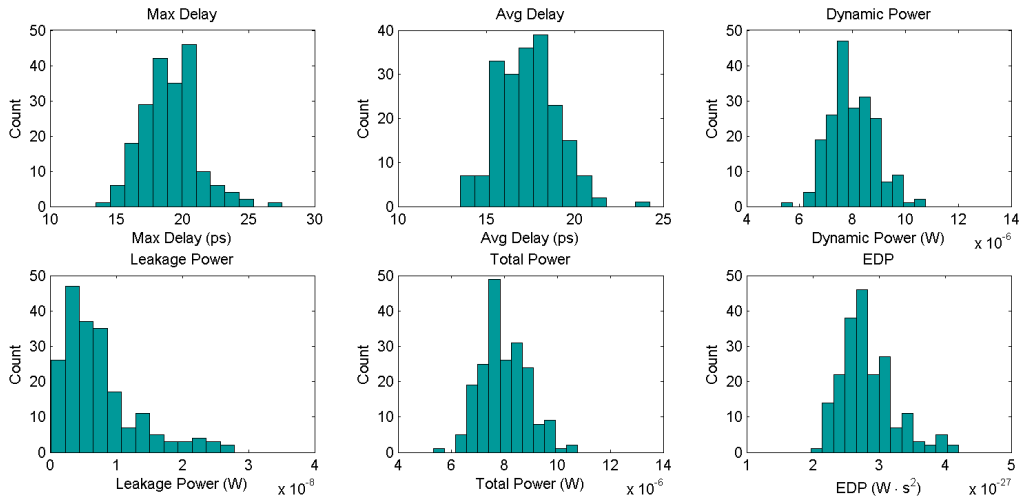

Figure 2.23: Monte Carlo simulation of a Si-CMOS inverter with all three parameters varied.

2.4.4.2 Simulation Results

The High-Performance (HP) 16-nm Si-CMOS serves as a baseline for the subsequent simulations of MOS-GNRFET circuits. The *inv* and *c17* simulation results for Si-CMOS when all parameters are varied are given in Figures 2.23 and 2.24. The results for ideal MOS-GNRFET are given in Figures 2.25 and 2.26. The results for MOS-GNRFET with contact resistance and $p_r = 0.1$ are given in Figures 2.27 and 2.28.

2.4.4.3 Discussion

We analyze the Monte Carlo simulation results based on the exploration in Section 2.4.3.3. For N variation, I_{on} and I_{off} both change drastically, as shown in Figure 2.9. Figure 2.18 shows that f_{dop} changes the I_{on} and I_{off} more than T_{ox} and L_{CH}, and thus it is the dominating factor when these parameters are varied simultaneously. For f_{dop}, I_{on} only changes when the f_{dop} varies for a few orders of magnitude, but I_{off} decreases exponentially with f_{dop} when doping is low, as shown in Figure 2.10. For T_{ox} variation, I_{on} increases exponentially with T_{ox} and I_{off} decreases exponentially with T_{ox}, but not as drastically as with f_{dop} variation. L_{CH} only has a second-order effect on I_{on} and I_{off}.

As I_{on} and I_{off} is dominated by f_{dop}, the leakage power follows a log-normal

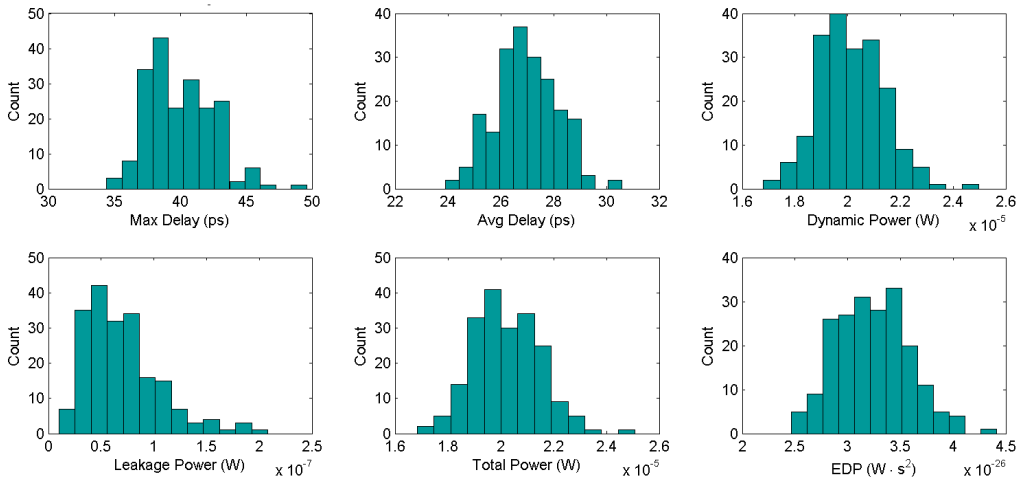

Figure 2.24: Monte Carlo simulation of a Si-CMOS *c17* circuit with all three parameters varied.

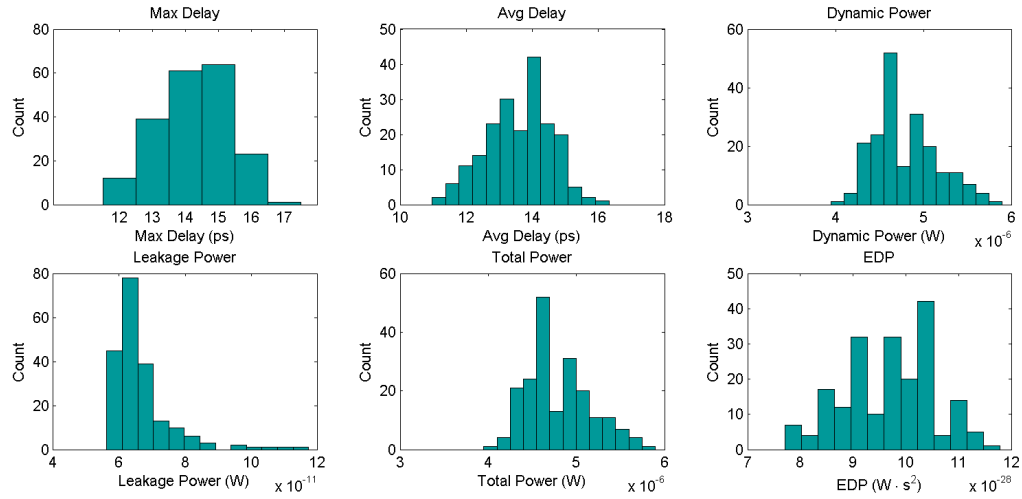

Figure 2.25: Monte Carlo simulation of an inverter of ideal MOS-GNRFET with contact resistance with all three parameters varied.

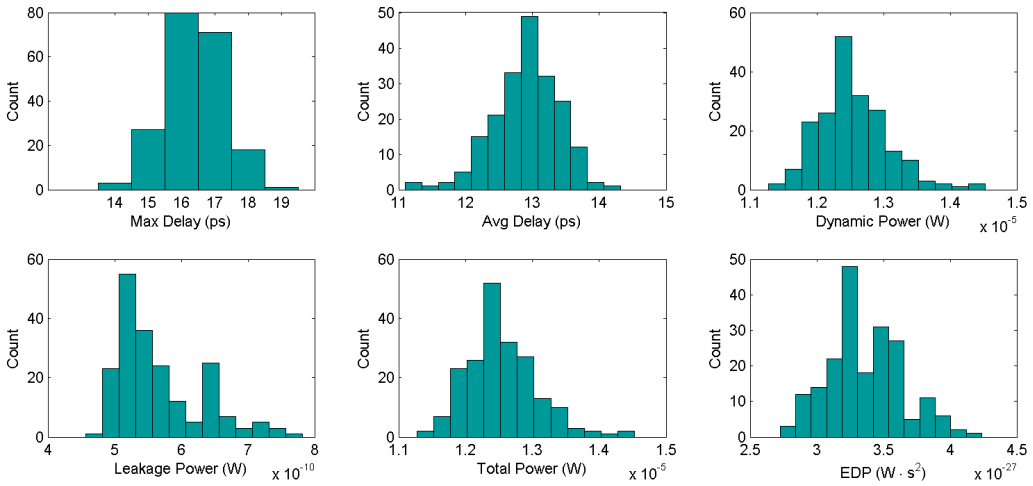

Figure 2.26: Monte Carlo simulation of a *c17* circuit of ideal MOS-GNRFET with contact resistance with all three parameters varied.

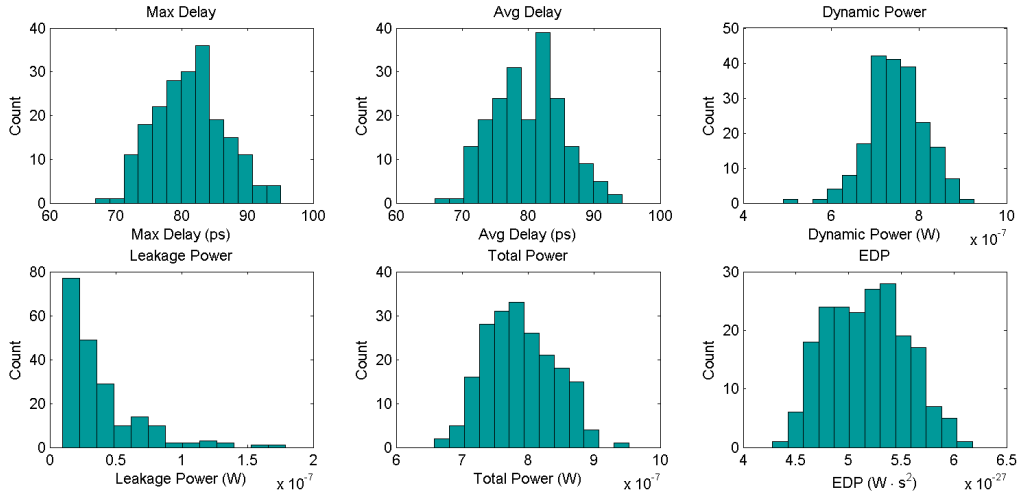

Figure 2.27: Monte Carlo simulation of an inverter of MOS-GNRFET with contact resistance and $p_r = 0.1$ with all three parameters varied.

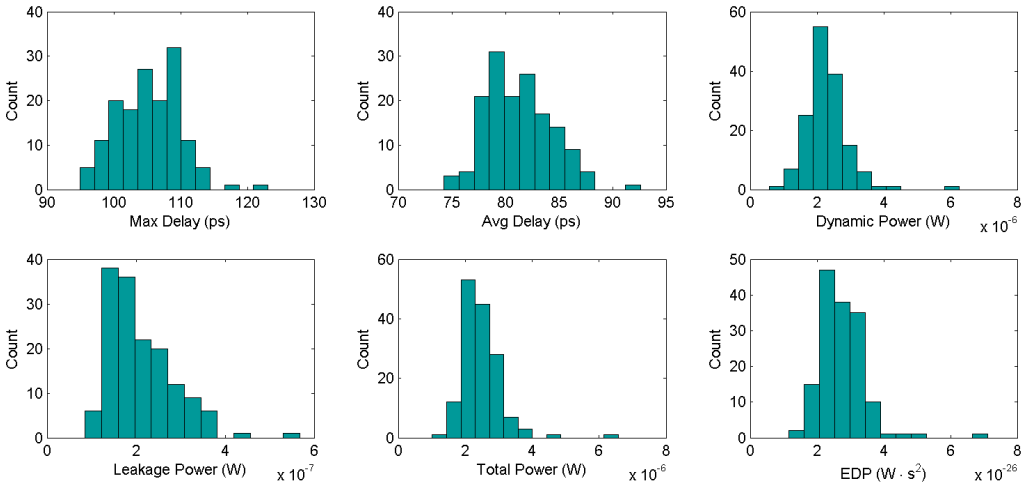

Figure 2.28: Monte Carlo simulation of a *c17* circuit of MOS-GNRFET with contact resistance and $p_r = 0.1$ with all three parameters varied.

distribution, as I_{off} has an exponential relationship with doping. Delay, dynamic power, and total power follow a bell-shaped distribution skewed to the left as I_{on} is linear in f_{dop} but is exponential in T_{ox}, thus making the distribution a mixture of normal and log-normal. EDP is computed from delay and total power and follows a distribution that is close to normal. For channel width variation, the effect on Si-CMOS is continuous, whereas for MOS-GNRFET, the GNR width is discrete, and therefore the effects are more discrete. Note that for MOS-GNRFET circuits, there are a few outliers that fall outside of the normal or log-normal distribution. This is likely because a different value of N was chosen for a transistor and caused it to have a very different I-V curve.

In the above simulations, only a relatively simple circuit was simulated due to the time-consuming nature of Monte Carlo simulations. As discussed in Section 2.4.3.5, each data point of one transient analysis of *c17* takes up to hours to complete. We also did not consider spatial correlation due to the lack of layout information in SPICE-level netlists, only intra-gate correlation. Evaluation on large-scale circuits with more elaborate variation modeling such as spatial correlation would require a different framework, such as integration with higher-level CAD tools.

2.5 Conclusion

We presented a parameterized, SPICE-compatible compact model of a MOS-type GNRFET. It captured the effects of V_{DD}, $N(W_{CH})$, L_{CH}, T_{ox}, f_{dop}, and line 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 Si-CMOS since they have either lower delay or lower power. We also showed that ribbon width variation and line edge roughness can critically reduce the performance and leakage power advantages of GNRFETs, which is a major shortcoming of this emerging technology.

In terms of process variation evaluation, we performed a series of deterministic and Monte Carlo simulations to show the effects from each design parameter and provided various insights on the different aspects of GNRFET's performances. Such data may provide early guidance for future experimental studies of GNRFETs.

CHAPTER 3

SB-GNRFET MODELING AND SIMULATION

3.1 Introduction

Graphene has recently received a lot of attention as a material for nanoelectronic devices due to its outstanding physical and electrical properties [54, 36, 11]. The thin, planar, and robust lattice makes graphene potentially compatible with the existing Si-CMOS manufacturing technology [17] and suitable for making flexible electronics [55]. Meanwhile, successfully fabricated devices have been demonstrated [17, 55, 21, 40, 41], where the fabricated graphene nanoribbons (GNRs) can have widths < 10 nm and of high quality with fairly smooth edges. In particular, a Si compatible, transfer-free, and *in situ* GNR field-effect transistor (GNRFET) fabrication method is presented in [17], demonstrating compatibility and integrability of GNRFETs with the existing Si technology available in the industry. One advantage in the Schottky-barrier (SB)-type SB-GNRFET is that it requires no doping in the terminals or the channel. Therefore, it reduces the technical difficulty in fabrication and eliminates doping variation. As a result, most fabricated GNRFETs reported in literature are SB-type [17, 55, 21, 40, 41].

Modeling and computer simulation are very useful in providing physical insights of GNRFETs and evaluating the performance of futuristic graphene-based circuits. Numerical simulations based on nonequilibrium Green's function formalism have been implemented in the 3-D device simulator NanoTCAD ViDES in [23, 56, 35]. Nonclosed-form analytical models that describe SB-GNRFETs are presented in [25] and [57]. In terms of high-level simulations, a circuit simulation framework of SB-GNRFETs based on lookup tables is presented in [26, 27]. Due to the complicated tunneling effects occurring at the Schottky barriers, no physics-based closed-form model of SB-GNRFETs has been developed yet.

In this chapter, we develop a physics-based analytical model for the current-voltage (I-V) characteristics of SB-GNRFETs based on two approaches: (1) cutoff model and (2) Fermi-transmission product approximation (FTPA) model. The cutoff model of the Schottky barrier tunneling probability, which we proposed in [2], has a low computational complexity, but is oversimplified and thus inaccurate in the region where the Schottky barrier tunneling effect is prominent. Therefore, this model results in an overestimation of the OFF current I_{off}, giving a pessimistic view of the I_{on}/I_{off} ratio. The FTPA model is based on elaborate approximations of Schottky barrier tunneling, channel charge, and channel current, which provides improved accuracy over the cut-off model while maintaining compactness. With the proposed model, which is released on nanoHUB [32], we enable an accurate and realistic simulations of SB-GNRFET circuits.

For a fair comparison and for the increasing trend in the use of multi-gate (MG) transistors, we compare with MG Si-CMOS designs (e.g., FinFETs) [58] in our circuit simulation experiments. Note that double-gate (DG) graphene-based transistors are fabricated in a planar fashion due to graphene's thin-film structure [59], while MG Si-CMOS transistors are usually of a 3-D FinFET-like structure. Nevertheless, they both demonstrate better gate control ability than single-gate (SG) designs and are likely to be adopted in the upcoming technology nodes.

With the proposed model, we perform a comparative study on SG and DG SB-GNRFETs with MG Si-CMOS on their respective circuit-level delay and power performance. Because GNRFET is regarded as a next-generation device, we are interested in its scalability in future technology nodes. Therefore, we simulate benchmark circuits on the 16-, 14-, 10-, and 7-nm technology nodes to provide insights on scalability. We show that SB-GNRFET circuits have a consistently decreasing trend in delay, power, and energy-delay product (EDP) with respect to the transistor size, indicating that SB-GNRFET is a promising device in future technology nodes.

To summarize, the main contributions of this chapter are as follows:

- Proposing an effective and detailed closed-form approximation of the Schottky barrier tunneling effect, the channel charge, and the channel current.

- Developing a highly accurate compact SB-GNRFET model, supporting

both SG and DG transistor designs.

- Evaluating the effect of design parameters and process variations on the performance of SB-GNRFETs.
- Comparing circuit-level performance among MG Si-CMOS, SG SB-GNRFET, and DG SB-GNRFET.
- Providing insights on technology scaling with the above technology nodes.

The rest of this chapter is organized as follows. Section 3.2 covers more background knowledge on graphene and GNRs. Section 3.3 discusses the modeling of tunneling in SB-GNRFETs and presents our SB-GNRFET compact model. Section 3.4 presents the experimental results, including model validation, transistor level evaluation, and circuit simulations. Finally, the conclusion is drawn in Section 3.5.

3.2 Graphene Energy Dispersion

This section along with Section 3.3 describes the SB-GNRFET modeling approach. This is started with the calculation of the energy dispersion, band-edge energy, density of states (DOSs), and effective mass (EM) for the GNR. Then, the Schottky barrier model is presented and is followed by the computation of the tunneling probability using Wentzel-Kramers-Brillouin (WKB) approximation. Afterward, an analytical model is proposed to obtain the channel charge and current.

Graphene is a single atomic layer of graphite with 2-D honeycomb crystal lattice. It is a zero-bandgap material that makes it metallic and unable to be turned ON or OFF [13]. Energy gap can, however, be induced by means of lateral confinement [60]. In order to open the band gap and make graphene into a good semiconductor, it is patterned into 1-D GNRs with widths less than 10 nm [13]. The band gap of a GNR is mainly inversely proportional to its width [19]. The width of a GNR (denoted by W_{CH}) is commonly defined via the number of dimer lines N as shown in Figure 3.1, as $W_{CH} = \sqrt{3}d_{cc}(N-1)/2$, where $d_{cc} = 0.142$ nm refers to the carbon-carbon

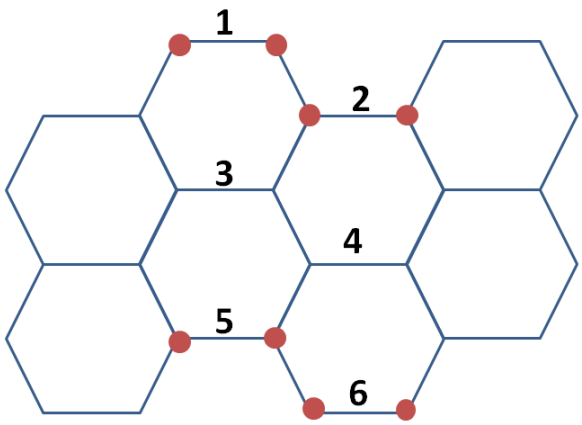

Figure 3.1: Lattice structure of an AGNR with $N = 6$. N is the number of dimer lines in the armchair orientation.

bond distance [61]. As the width of the GNR increases, the band structure of GNRs gradually returns to that of a 2-D graphene sheet.

Based on the edge geometry, GNRs are categorized into two types: armchair-GNRs (AGNR) and zigzag-GNRs (ZGNR) [20]. In this chapter, we focus on AGNRs due to its semiconducting property. The energy dispersion relation of an AGNR for subband $\alpha = 1, 2, ..., N$ is given in [62, 63] as

$$E_\alpha(k) = \pm t\sqrt{(1 + 4A_\alpha \cos\frac{\sqrt{3}a_1 k}{2}) + 4A_\alpha^2} \tag{3.1}$$

where k is the wavevector, $A_\alpha = \cos(\pi\alpha/(N+1))$, $a_l = \sqrt{3}d_{cc}$, and $t = 2.7$ eV is the nearest neighbor overlap energy. The latter parameter is different for the carbon atoms at the edge of the ribbon. This can be accounted by the edge-corrected energy dispersion $E_\alpha^c(k) = E_\alpha(k) + E_\alpha^\delta(k)$, in which the correction energy is obtained using the approach given in [19] as

$$E_\alpha^\delta(k) = s_\alpha \frac{4\nu t}{N+1} \sin^2(\frac{\alpha\pi}{N+1}) \cos(ka_1) \tag{3.2}$$

where $\nu = 0.12$ eV, and

$$s_\alpha = \begin{cases} 1 & A_\alpha \geq \frac{1}{2} \\ -1 & \text{otherwise} \end{cases} \tag{3.3}$$

The density of states (DOS) can be obtained from the effective-mass (EM) approximation as

$$D_\alpha(E) = \frac{2}{\pi\hbar}\sqrt{\frac{M_\alpha}{2E}} \tag{3.4}$$

where $\hbar$ is the reduced Planck's constant, $E = E_\alpha^c(k) - \varepsilon_\alpha$ is the energy with respect to the band edge energy $\varepsilon_\alpha = E_\alpha^c(0)$, and M_α is the effective mass given by

$$M_\alpha = -\frac{2\hbar^2\varepsilon_\alpha}{3a_l^2t^2A_\alpha} \tag{3.5}$$

3.3 Modeling SB-GNRFET

This section covers background knowledge on the Schottky barrier and transmission coefficient calculation. It follows with the circuit modeling of the SB-GNRFET. Then, we present the approximations for the channel charge and current that are required to make the compact analytical model.

An SB-GNRFET consists of a GNR-based channel and metal electrodes: gate, drain, and source. An example of SB-GNRFET is given in Figure 3.2. The interface between the metal drain/source and the GNR channel results in a Schottky barrier at the graphene-metal junction. SB-GNRFETs have an ambipolar I-V curve with minimum current at $V_{GS} = 1/2 V_{DS}$ [25, 57].

Multiple GNRs can be connected in parallel to increase driving strength, as in Figure 3.3. The GNRFET of interest has the following design parameters: (1) L_{CH} is channel length; (2) W_{CH} is the ribbon width; (3) W_G is the gate width; (4) $2W_{sp}$ is the ribbon spacing; and (5) T_{ox} is the oxide thickness.

3.3.1 Schottky Barriers and Tunneling

Schottky barriers are introduced on the interface of metal and graphene. With Schottky barriers present, the charge transport in the device is dominated by Schottky barrier tunneling. The Schottky barrier width is modulated by the gate voltage, changing the tunneling probability for carriers. The band diagram of SB-GNRFET has three distinctive regions: two injecting regions at the ends of the GNR and a central region where ballistic

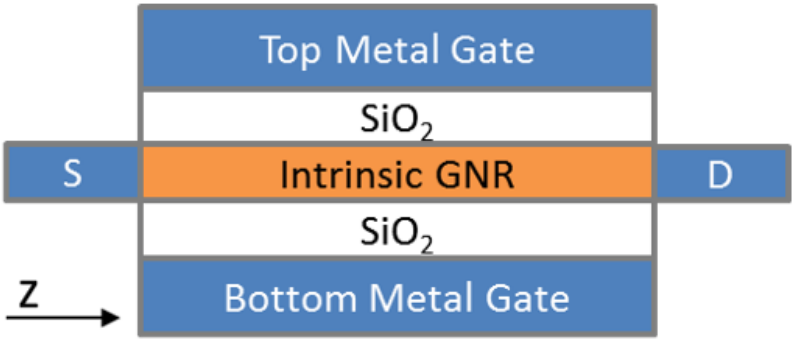

Figure 3.2: Cross section of a double-gate SB-GNRFET device.

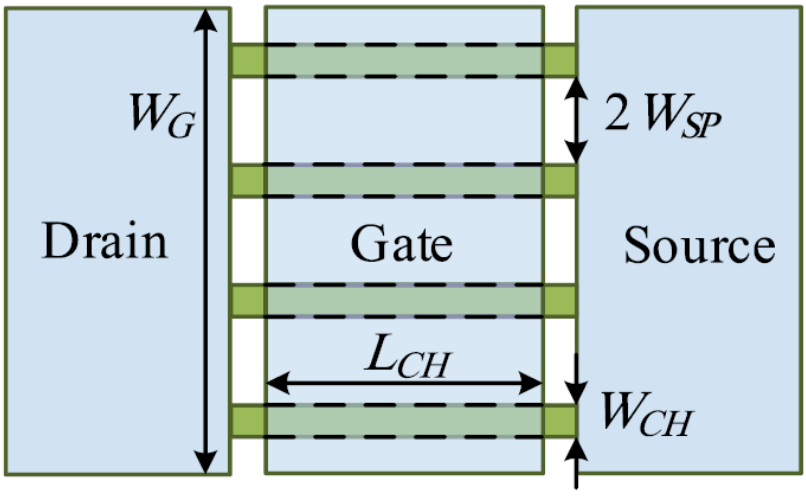

Figure 3.3: Structure of a four-ribbon SB-GNRFET. A common drain and a common source are shared by the ribbons.

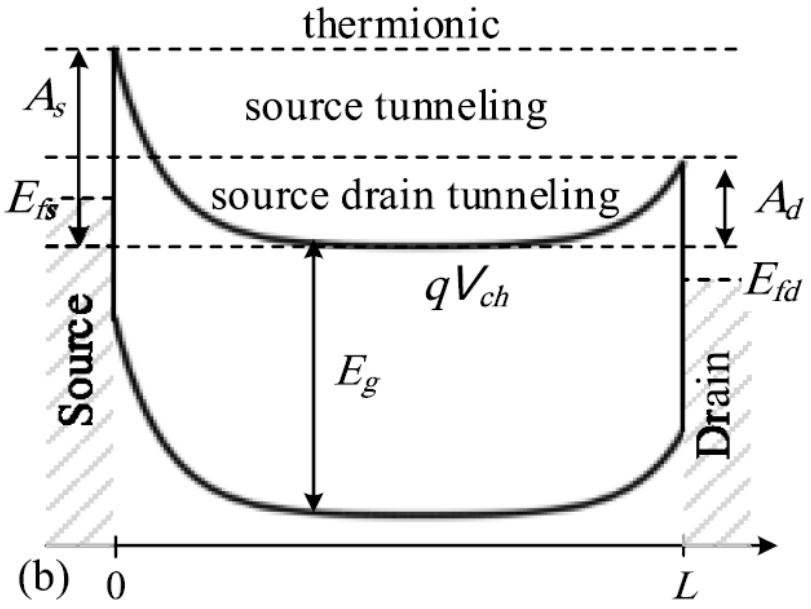

Figure 3.4: Cross section of a double-gate SB-GNRFET device.

transport occurs, as shown in Figure 3.4. In a sufficiently-long-channel device, the central region is of flat-band type with the electrostatic potential of φ_{ch}.

The Schottky barrier profile near the metal/GNR interface, which can be analytically solved from the 1-D Laplace equation [25], takes the following form along the channel direction z

$$E_{SB}(z) \propto \frac{2}{\pi}\cos(e^{-\frac{z\pi}{2T_{ox}}}) \tag{3.6}$$

which can be simplified using $\cos^{-1}$ expansion as

$$E_{SB}(z) = A_s e^{-\frac{z}{\lambda}} \tag{3.7}$$

where $\lambda = 2T_{ox}/\pi$ is the scale length, and $A_s = q\varphi_{ch}$ for the lowest subband at the source. The valence band has the same profile as the conduction band but is downshifted by an amount equal to the GNR energy gap $E_g = 2\varepsilon_\alpha$. In the case of $|\varphi_{ch}| > E_g$, the spatial band diagram curvature becomes high enough to trigger band-to-band tunneling (BTBT), as shown in Figure 3.5 (a). In this case, a carrier with energy $0 < E < A_s - 2\varepsilon_\alpha$ experiences a Schottky barrier of a height $A_s = E + 2\varepsilon_\alpha$.

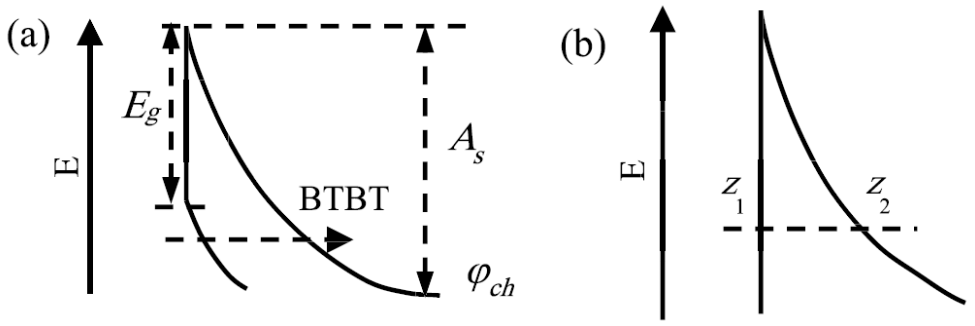

Figure 3.5: Schottky barrier, **(a)** band-to-band tunneling, **(b)** classical turning points.

The tunneling phenomenon is characterized by the transmission coefficient $T(E)$ of a carrier. In the case of thermionic conduction, in which the carrier has higher energy than the Schottky barrier, $T(E)$ is equal to unity. For the tunneling conduction, the transmission through a single Schottky barrier is computed based on the Wentzel-Kramers-Brillouin (WKB) approximation [64]

$$T(E) = \exp\left\{-2\int_{z_1}^{z_2} \mathrm{Im}[k_z(E)]dz\right\} \tag{3.8}$$

where $z_1 = 0$ and $z_2 = -\lambda \ln(E/A_s)$ are the classical turning points, as shown in Figure 3.5, and $\mathrm{Im}[k_z(E)]$ is the imaginary part of the wavevector. The momentum $k_z(E)$ is related to the energy through the GNR E - k dispersion relationship of (3.1), which can be obtained by expanding $\cos(x) \simeq (1-x^2/2)$ as

$$k_z(E) \simeq \sqrt{\frac{M_\alpha}{\hbar^2 \varepsilon_\alpha}} \sqrt{E_\alpha^2 - \varepsilon_\alpha^2} \tag{3.9}$$

Inserting barrier profile of $E_{SB}(z)$ into this equation results in

$$\mathrm{Im}[k_z(E)] \simeq \sqrt{\frac{M_\alpha}{\hbar^2 \varepsilon_\alpha}} \sqrt{\varepsilon_\alpha^2 - (E + \varepsilon_\alpha - A_s e^{-\frac{z}{\lambda}})} \tag{3.10}$$

Integrating for $E < A_s$ leads to the transmission coefficient of

$$T(E) = \exp\left\{-2\lambda\sqrt{\frac{M_\alpha}{\hbar^2\varepsilon_\alpha}}\left[(E+\varepsilon_\alpha)(\frac{\pi}{2} - \arctan\frac{\varphi_\alpha}{\gamma_1}) \right.\right. \tag{3.11}$$

$$\left.\left. +\gamma_1 + \gamma_2\left(\arctan\left(\frac{\gamma_1\gamma_2}{A_s(E+\varepsilon_\alpha) - E(E+2\varepsilon_\alpha)}\right)\right)\right]\right\} \tag{3.12}$$

where

$$\gamma_1 = \sqrt{\varepsilon_\alpha^2 - \varphi_\alpha^2} \tag{3.13}$$

$$\gamma_2 = \sqrt{(\varphi_\alpha + A_s)^2 - \varepsilon_\alpha^2} \tag{3.14}$$

$$\varphi_\alpha = \epsilon_\alpha + E - A_s \tag{3.15}$$

$$\theta_0 = \begin{cases} \pi & E(2\varepsilon_\alpha + E) < A_s(\varepsilon + E) \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

3.3.2 Full SB-GNRFET Model

The equivalent circuit of the full GNRFET as in Figure 3.3, is shown in Figure 3.6 (c). It consists of multiple parallel GNRs and parasitic capacitors C_{gd} and C_{gs}. Each transistor symbol marked in red represents a single GNR and is modeled by the circuit in Figure 3.6 (a). In a single GNR, I_{ds} models the current flowing through the channel, the capacitors $C_{ch,d}$, $C_{ch,s}$, $C_{g,ch}$, and $C_{sub,ch}$ model the parasitics, and the voltage-controlled voltage source V_{ch} represents the channel voltage that corresponds to the channel potential φ_{ch}, expressed as $V_{ch} = q\varphi_{ch}$. The capacitors C_{gd} and C_{gs} are modeled based on FastCap [46], which are functions of W_G and T_{ox}, as

$$C_{gd} = C_{gs} = 1.26 \times 10^{-10} W_G(0.8 - 0.2T_{ox} + 0.015T_{ox}^2) \tag{3.17}$$

Intrinsic capacitors $C_{ch,d} = \partial Q_{ch}/\partial V_D$ and $C_{ch,S} = \partial Q_{ch}/\partial V_S$ are implemented in SPICE as voltage-controlled capacitors by defining the charge equations. The total channel charge Q_{ch} is derived from the electron and hole density of each subband coming from the drain and the source. Depending on the magnitude of the applied bias, multiple reflections can arise between the series combination of source and drain Schottky barriers. The total charge of carriers subband α can be expressed as

$$Q^i_\alpha(\varphi_{ch}) = q \int D_\alpha(E) G^i_Q(E) dE \tag{3.18}$$

where $G^i_Q(E)$ is defined as

$$G^i_Q(E) = T^i_{TS}(E) \cdot f(E - E^i_{(\alpha,s)}) + T^i_{TD}(E) \cdot f(E - E^i_{(\alpha,d)}) \tag{3.19}$$

where i can be either e or h, representing total electron or hole charge in the subband α, q is the electron charge, and $f(\cdot)$ is the Fermi-Dirac distribution function given by

$$f(E - E_F) = \left[1 + \exp\left(\frac{E - E_F}{k_B T}\right)\right]^{-1} \tag{3.20}$$

$$E^e_{\alpha,j} = -(\varepsilon_\alpha - q\varphi_{ch} + V_j) \; j = s, d \tag{3.21}$$

$$E^h_{\alpha,j} = -(-\varepsilon_\alpha + q\varphi_{ch} - V_j) \; j = s, d \tag{3.22}$$

where k_B is the Boltzmann constant, T is the temperature in Kelvin, and $j = s, d$ denotes the source and drain terminals, respectively. The tunneling coefficients within this formalism play a critical role as

$$T_{TS}(E) = \frac{T_s(2 - T_d)}{T_s + T_d - T_s T_d} \tag{3.23}$$

$$T_{TD}(E) = \frac{T_d(2 - T_s)}{T_s + T_d - T_s T_d} \tag{3.24}$$

where T_s (T_d) is the transmission coefficient of a carrier going through the Schottky barrier on the source (drain) side, given by (3.11). Note that T_s (T_d) should be computed for both electrons and holes. The total mobile charge $Q_{ch} = \sum_\alpha (Q^h_\alpha - Q^e_\alpha)$ must be equal to the charge Q_{cap} across the gate, source and drain capacitors that couple into the channel.

$$Q_{cap}(\varphi_{ch}) = -\sum_{i=g,s,d} C_{i,ch}(V_i, V_{FB,i} - q\varphi_{ch}) \tag{3.25}$$

$$C_{g,ch} = \frac{5.55 \times 10^{-11} \epsilon_\gamma L_{CH}}{\left(1 + \frac{1.5T_{ox}}{W_G}\right) \ln\left(\frac{5.98W_{CH}}{0.8T_{ox}}\right)} \tag{3.26}$$

$$C_{s,ch} = C_{d,ch} = 0.1C_{g,ch} \tag{3.27}$$

where $V_{FB,i}$ is the flatband voltage and ϵ_r is the relative permittivity of the material. Equation (3.26) is also modeled empirically from data extracted from FastCap. Both Q_{ch} and Q_{cap} are functions of φ_{ch} and have to be equal in magnitude due to charge conservation. As a result, equating Q_{ch} and Q_{cap} yields a solution of φ_{ch}, which can be obtained using the equation solver circuit of Figure 3.6 (b) [29, 30] in SPICE simulations. In the solver, two voltage-controlled current sources are connected in series, forcing V_{ch} to take a value such that the currents from the two sources are equal.

3.3.3 Compact Modeling of a Single GNR with Schottky Barriers

As $T(E)$ in (3.11) takes a complicated form, it often results in nonclosed-form solutions when integrating with other quantities, making the model non-compact. This section describes the two approaches which we employed to reduce the computational burden.

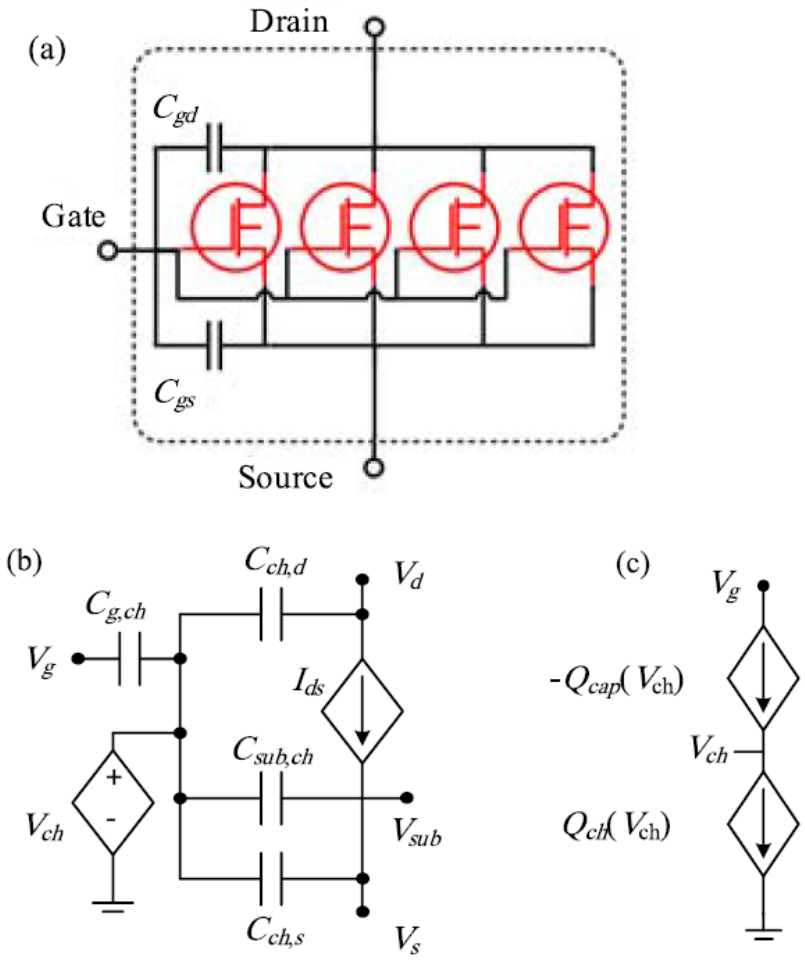

Figure 3.6: (a) SPICE model of an SB-GNRFET with four parallel GNRs. (b) SPICE model of a single GNR. (c) SPICE setup for solving V_{ch}, V_g, V_d, V_s, and V_{sub} are voltages of each terminal.

3.3.3.1 Cutoff Model

In this section, we present a simple cutoff approximation for $T(E)$. A cutoff approximation is proposed in [45, 65] and was shown to work with SB-type carbon-nanotube (CNT) FET, which is a 1-D carbon-based device like GNRFET. In the cutoff model, E_{SB} is simplified into a rectangular shape of an effective height $\Phi_{(SB,eff)}$ and thickness d_{SB}, as illustrated in Figure 3.7 (left). With this barrier, a carrier can never tunnel through when its energy is lower than $\Phi_{(SB,eff)}$, and can always tunnel through when its energy is higher than $\Phi_{(SB,eff)}$. The resulting $T(E)$ becomes a step function:

$$T(E) = \begin{cases} 0 & \text{if } E \leq \Phi_{SB,eff} \\ 1 & \text{if } E > \Phi_{SB,eff} \end{cases} \tag{3.28}$$

as illustrated in Figure 3.7 (right), where the effective barrier height $\Phi_{(SB,eff)}$ is determined by

$$\Phi_{(SB,eff)} = (A_s - \varphi_{ch})e^{-\frac{d_{SB}}{\lambda}} + \varphi_{ch} \tag{3.29}$$

which minimizes the mismatch of integrations under exponential and rectangular profiles. This is an oversimplification, but with a careful selection of d_{SB}, it provides a fair approximation, as shown in [65]. In our work, d_{SB} is chosen as

$$d_{SB} = d_{SB0}\frac{\sqrt{0.042M_e}}{\sqrt{M_\alpha}} \tag{3.30}$$

where d_{SB0} is the reference Schottky barrier thickness based on a chosen reference effective mass, and M_e is the electron mass. The resulting current through the channel is computed by using the Landauer-Büttiker formalism [66] and recognizing the Fermi-Dirac integral of order 0, as

$$I_e = \frac{2qkT}{h}\sum_{\alpha}\left[\ln\left(1+e^{\frac{q\Phi_{SB,eff}-V_S-q\varepsilon_\alpha}{kT}}\right) - \ln\left(1+e^{\frac{-q\Phi_{SB,eff}+V_D-q\varepsilon_\alpha}{kT}}\right)\right] \tag{3.31}$$

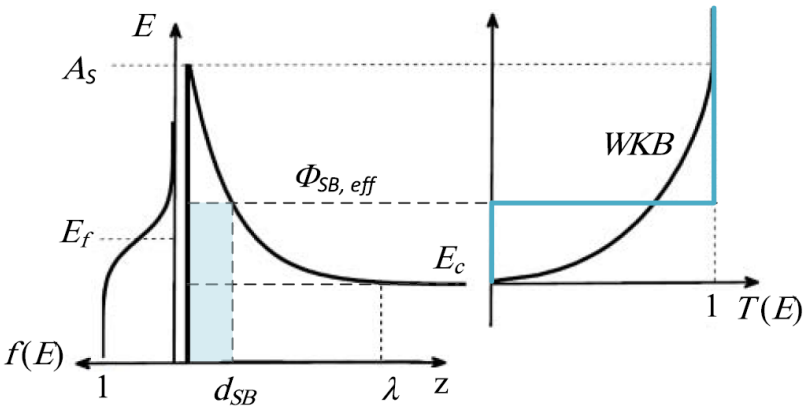

Figure 3.7: Schottky barrier profile (left) and transmission probability (right), with their simplification. z is the channel direction. E is the energy. $T(E)$ is the tunneling probability. E_F is the Fermi level on the source side. E_C is the conduction band edge. Φ_{SB} is the Schottky barrier height, while $\Phi_{SB,eff}$ is the effective Schottky barrier height after approximation, with the blue-shaded area being the Schottky barrier profile after approximation. λ_{SB} is the scale length of the exponential Schottky barrier profile. d_{SB} is the effective Schottky barrier thickness.

3.3.3.2 Fermi-Transmission Product Area (FTPA) Model

In this section, we propose a more elaborate approximation to make the integration in (3.18) closed-form. Following is the explanation of the approximation for the total electron charge. The same method can be employed to compute the total hole charge.

The term G_Q^e of electrons for different cases is shown in Figure 3.8. For all the cases, G_Q^e can be approximated by using two lines defined by four values E_{c1}, $E_{c2,Q}$, $G_{Q,0}$ and $G_{Q,1}$. However, it introduces some error for high gate voltages (see Figure 3.8 (c) and (d)). These approximations are shown as dashed lines in Figure 3.8. Here, we explain the attainment of the above-mentioned four values. The term $G_{Q,0}$ is the value of G_Q^e in which the energy E is equal to zero. The term E_{c1} is predominantly determined by the Fermi distribution of source or drain terminals, whichever has a lower voltage, and the corresponding transmission coefficient which is in its linear region, as shown in Figure 3.9. The linear part of the transmission coefficient is defined as $T_{lin} = c \cdot E$, where the constant value c does not need to be computed since it will disappear in the subsequent equations. Therefore, the approximation of G_Q^e defined in 3.32 as the product of T_{lin}

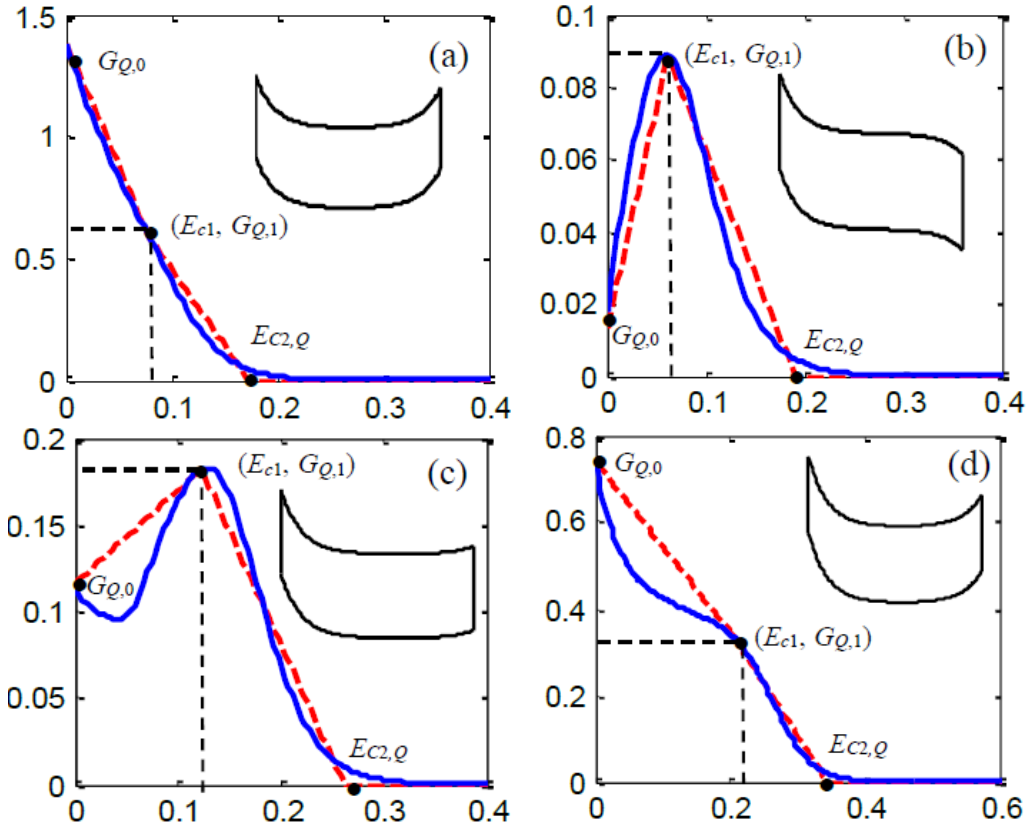

Figure 3.8: G_Q^e of electrons for different bias voltages. **(a)** $V_d = 0.1$, $V_g = 0.8$, **(b)** $V_d = 0.5$, $V_g = 0.4$, **(c)** $V_d = 0.4$, $V_g = 0.55$, **(d)** $V_d = 0.3$, $V_g = 0.9$. Solid lines are numerical and the red dashed lines are closed-form analytical. Inset in each subfigure is the energy band diagram in such condition.

$$G_Q^e \simeq T_{lin} \cdot f(E - E_f) \tag{3.32}$$

$$E_f = -(\varepsilon_\alpha - qV_{ch} + \min(V_s, V_d)) \tag{3.33}$$

E_{c1} is the maximum of G_Q^e in Figures 3.8 (b) and (c), and is the local maximum in its small vicinity in Figures 3.8 (a) and (d). We approximate G_Q^e as two different functions in the two regions separated by E_{c1}. As a maximum, E_{c1} is obtained by computing the differentiation $\partial G_Q^e / \partial E$, equating to zero, and solving, which is expressed as

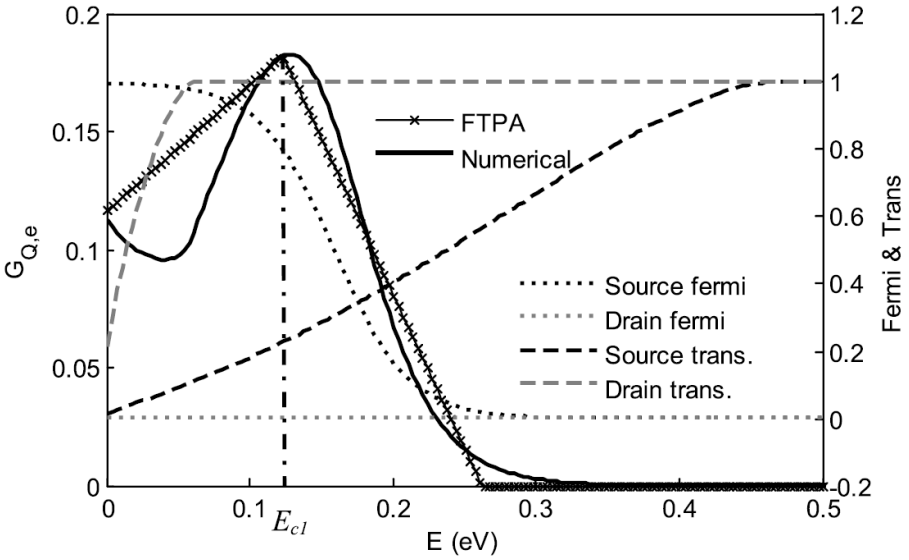

Figure 3.9: Fermi distribution, transmission coefficient and the product G_Q^e of electrons for a GNRFET with $N = 12$ and $T_{ox} = 2$ nm at $V_d = 0.4$ and $V_g = 0.55$.

$$E_{c1}(E_f, T) = k_B T \left[W\left(\exp\left(\frac{E_f}{k_B T} - 1 \right) + 1 \right) \right] \tag{3.34}$$

where $W(\cdot)$ is the Lambert W product logarithm function. Figure 3.10 shows E_{c1} as a function of E_f for different temperatures. E_{c1} can be approximated as constant, parabolic, or linear functions of E_f in three different regions defined by $E_{f1} = -0.05$, and $E_{f2} = 0.145$. The E_{f1} and E_{f1} are obtained in order to achieve the best fitting in different regions in the temperature range of 200 K $\leq$ T leq 400 K.

$$E_{cq}(E_f, T) = \begin{cases} k_B T & E_f < E_{f1} \\ p_1 E_f^2 + p_2 E_f + p_3 & E_{f1} < E_f < E_{f2} \\ p_4 E_f + p_5 & E_f > E_{f2} \end{cases} \tag{3.35}$$

where $p_i, i = 1, 2, ..., 5$, are temperature-dependent coefficient as $p_i = \eta_{i,1} \cdot T + \eta_{i,2}$. Values of $\eta_{i,1}$ and $\eta_{i,2}$ obtained by curve fitting are given in Table 3.1.

The term $G_{Q,1}$ is the value of G_Q^e in which the energy E is equal to E_{c1} and can be computed using (3.19), and the E_{c1} obtained by (3.34). The term

Table 3.1: Values of Temperature-Dependent Coefficients

i	1	2	3	4	5
$\eta_{i,1}$	-0.0041	-1.33e-4	1.016e-4	-4.47e-4	6.29e-5
$\eta_{i,2}$	3.4092	0.2827	0.0035	1.0082	-0.0315

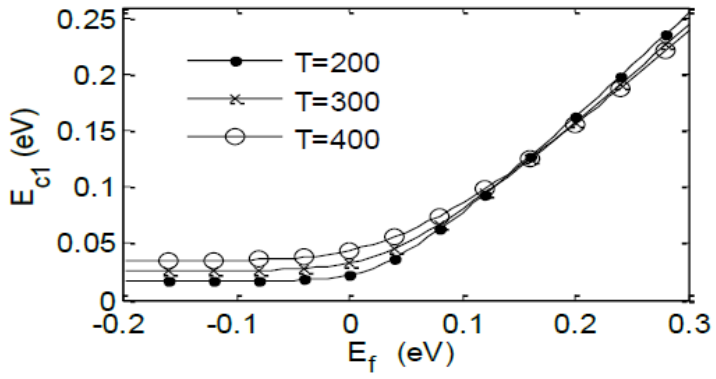

Figure 3.10: E_{c1} as a function of E_f for different temperatures T.

$E_{c2,Q}$ is the point that G^e_Q reaches zero, which is predominantly determined by the exponentially decreasing Fermi distribution compared with the linear variation of transmission coefficient. By substituting the transmission probability T_{lin} in (3.32) with unity (its upper bound), and approximating $E_{c2,Q}$ as the value that makes $f(E_{(c2,Q)} - E_f) = f(E_{c1} - E_f)/10$, one can obtain an approximation of (3.32 at E = $E_{c2,Q}$ as $f(E_{c2,Q} - E_f) \simeq G_{(Q,1)}/10$. This equation can be used to obtain the value of $E_{c2,Q}$. The integration of (3.18) can be, therefore, analytically computed as

$$Q^e_\alpha = q\int_0^{E_{c1}} D_\alpha(E)(a_1E + b_1)dE + q\int_{E_{c1}}^{E_{c2,Q}} D_\alpha(E)(a_2E + b_2)dE \quad (3.36)$$

$$= \frac{4}{3}\frac{q\sqrt{2M_\alpha}}{\pi\hbar}\left[\sqrt{E_{cq}}G_{Q,0} + \frac{G_{Q,1}E_{c2,Q}}{\sqrt{E_{c1}} + \sqrt{E_{c2}}}\right] \quad (3.37)$$

The same method can be used to compute the hole charge by using $E_f = -(-\varepsilon_\alpha + q\varphi_{ch} - \max(V_s, V_d))$.

Given φ_{ch}, the current through the channel is computed by using the Landauer-Büttiker formalism [66] as

$$I_{\alpha}^{i} = \frac{q}{\pi\hbar}\int G_{I}^{i}(E)dE \tag{3.38}$$

$$G_{I}^{i}(E) = T_{I}^{i}[f(E - E_{\alpha,S}^{i}) - f(E - E_{\alpha,d}^{i})] \tag{3.39}$$

$$T_{I} = \frac{T_{s}T_{d}}{T_{s} + T_{d} - T_{s}T_{d}} \tag{3.40}$$

Note that T_I should be computed for both electrons and holes. Here, we are again facing the complexity of integrating G_I^i into closed-form expressions. The term G_I^i can be approximated in the same way as G_Q^i. The term E_{c1} takes the same value obtained for charge approximation given in (3.34). The term $G_{I,1}$ is computed from (3.39) for $E = E_{c1}$, and $E_{c2,I}$ is calculated by using $f(E_{c2,I}) - E_f = G_{I,1}/10$. The current of electrons then can be analytically computed by using the integration of (3.38) as

$$I_{\alpha}^{e} = \frac{q}{\pi\hbar}\left[\int_{0}^{E_{c1}}(a_1'E + b_1')dE + \int_{E_{c1}}^{E_{c2,I}}(a_2'E + b_2')dE\right] \tag{3.41}$$

$$= \frac{q}{\pi\hbar}[G_{I,0}E_{c1} + G_{I,1}E_{c2,I}] \tag{3.42}$$

The same method can be used to compute the current of holes. The total current is given by $I_{DS} = I^e - I^h$.

3.3.3.3 Nonidealities

3.3.3.3.1 Line Edge Roughness Existing fabrication technology tends to produce GNRs with imperfect edges, which affects the quantum effects occurring inside GNRs. It is called *line edge roughness*, and is characterized by p_r, the probability that any atom on the edges of a GNR is missing due to imperfect fabrication [23]. There are two main effects from line edge roughness: (1) Change in effective subbands, and (2) disruption in ballistic transport.

To model the varying width, we introduce the concept of an effective subband $\varepsilon_{\alpha,eff}$ given by (3.43), where $\varepsilon_{\alpha,N}$ is the ε_α for a given N. In a unit segment of GNR, there are eight atoms (shown as red dots in Figure 3.1) 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 (3.43). The scattering coefficient A is introduced to account for the current reduction due to disrupted ballistic transport [1]. It is empirically modeled as (3.44). Because the work of [1] uses the same GNRs as those in this chapter, we believe it is valid to adopt the line edge roughness model here.

$$\varepsilon_{\alpha,eff} = (1-p_r)^8 \varepsilon_{\alpha,N} + (1-(1-p_r)^8)\varepsilon_{\alpha,N-1} \tag{3.43}$$

$$A = 0.98(1-4p_r)^6 + 0.02 \tag{3.44}$$

Current with line edge roughness present, I_{rough}, is derived as follows

$$I_{rough} = AI_{DS}(\varepsilon_{\alpha,eff}) \tag{3.45}$$

3.3.3.3.2 Mobility Degradation GNRFETs mobility is estimated using full-band electron and phonon dispersion relations in [18], and is reported to be $\sim$ 500 cm^2/V·s for 1-nm wide suspended GNR at room temperature. In this chapter, channel length and width are $\sim$ 15 and 1.5 nm, respectively. GNRs with this width have a mobility comparable with that of Si-CMOS [18, 67]. The mean free path is almost equal to the channel length for such a feature size, and carriers exhibit ballistic transport [18]. Moreover, the SB-GNRFET channel current is predominantly determined by the tunneling effect rather than the thermionic emission-diffusion, so we did not use the mobility as the metric to compare the performance of the device. In Section 3.4, we evaluate the device performance by comparing with the predictive technology models (PTMs) [49] in their default settings, which is a common baseline in the research community. In this way, we will be using the same baseline as other papers that are evaluating different emerging or enhanced

technologies.

3.3.3.3.3 Interface Charge The GNR to substrate or gate oxide (for DG structures) interface is another source of non-ideality. Interface charge can affect the charge transport in modern FET devices. A numerical study of the influence of substrate type and quality on GNR mobility and carrier transport is performed in [68]. HfO_2 and h-BN are optimal substrates for high and low impurity densities, respectively. Optimal performance can be achieved by employing a suitable substrate [68].

3.3.3.3.4 Contact Resistance The graphene-metal contact resistance at source/drain regions can also degrade the transistor performance. In SB-GNRFETs, the Schottky contact forms at the GNR to metal interface at the source and drain regions and the channel current is mainly due to the tunneling effect rather than the thermionic emission. Moreover, SB-GNRFETs have all metal drains, sources, and gate terminals. Therefore, there are no additional contact resistances in the interconnections [2], which may potentially be an advantage of SB-GNRFETs compared with circuits built with MOS-GNRFETs. However, the contact resistance can be considered by adding a series resistors at the source and drain terminals [1].

3.3.3.4 Single-gate SB-GNRFET

Single-gate (SG) GNRFETs are potentially easier to produce in practice due to the simpler structure. Therefore, simulation of these devices are of high interest. However, the asymmetry in the device structure makes the modeling more complicated, especially for the Schottky barrier profile and tunneling. As a result, we conduct an empirical approach that considers the geometrical and electrical parameters of the device. We performed extensive numerical ViDES simulations for the DG and SG SB-GNRFET devices with different geometrical and electrical parameters. The results, shown in Figure 3.11, show that the current of SG device, $I_{DS,SG}$, follows the current of DG device, $I_{DS,DG}$, in a predictable behavior affected by geometrical parameters, $_{ox}$ and N, and bias conditions, V_g and V_d. We derive an empirical fitted equation that relates the current of SG device to its DG counterpart as a function of device parameters as follows:

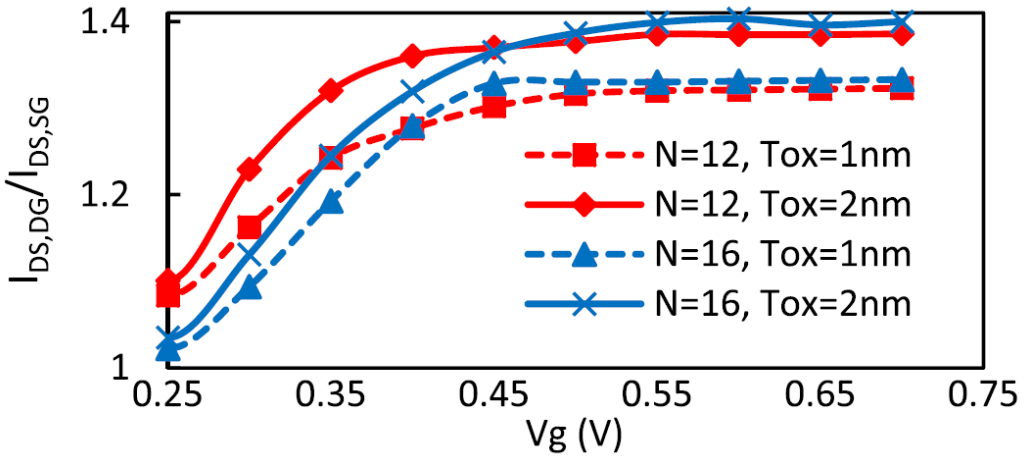

Figure 3.11: Ratio of the current of the DG device to its SG counterpart at $V_d = 0.5$ V.

$$r = \frac{I_{DS,DG}}{I_{DS,SG}} = 1 + \frac{f_1(T_{ox}, N)}{1 + e^{\frac{f_2(Bias)+f_3(N)}{c}}} \tag{3.46}$$

We derive an empirical equation which relates the current of SG device to its DG counterpart as a function of device parameters

$$r = 1 + \frac{0.065T_{ox} + 0.0042N + 0.2}{1 + e^{\frac{-(|V_{GS}-V_{DS}/2|+V_{DS}/2)+0.01N+0.17}{0.036}}} \tag{3.47}$$

3.4 Experimental Results

The proposed compact model is implemented in MATLAB using all the equations in Section 3.3. In Section 3.4.1, the implemented model is validated against numerical simulation in NanoTCAD ViDES. Then, the equivalent circuit model and all the model equations were implemented in HSPICE as a SB-GNRFET model library. With the accuracy of our SPICE model thoroughly validated, we can proceed with SPICE simulations of GNRFET and GNR-based circuits. This provides insightful information on how GNR-based circuits would perform once fabrication techniques become mature. In Sections 3.4.2 and 3.4.3, we implemented digital logic gates with our GNRFET

SPICE model; performed transistor-, gate-, and circuit-level analyses; and compared them with those implemented in MG Si-CMOS high-performance (HP) libraries from PTM.

3.4.1 Model Validation

In order to validate the proposed model, we implemented the model in MATLAB. The pseudocode of the model implementation is shown in Figure 3.12. It starts with the computation of energy dispersion parameters. Then, the channel voltage and charge are computed such that it satisfies the charge conservation equation. This includes the calculation of transmission coefficients, Fermi distributions, and FTPA model computations. Finally, the transistor current is computed using the obtained channel voltage. We compare the I-V curves obtained from the proposed analytical model with results from NanoTCAD ViDES. The I-V curves for design parameter $N = 12$ and $0 < V_g < 0.75$ are plotted in Figure 3.13, with $T_{ox} = 2$ nm in (a) and $T_{ox} = 1.5$ nm in (b). Figure 3.13 (c) shows the effect of N on the current. It is shown that the FTPA model agrees very well with ViDES results and has improved accuracy as compared to the cutoff model. In particular, the FTPA model gives a more realistic I_{on}/I_{off} ratio than the cutoff model (which results in, for example, 52% error in I_{on}/I_{off} of Figure 3.13(a) at $V_d = 0.6$ V), making the subsequent circuit simulations in the next sections more representative. Nevertheless, the cutoff model has the advantage of faster simulation time compared with the FTPA model, as shown in Figure 3.14. Furthermore, our models are much faster than the ViDES simulator that is based on time-consuming numerical computations. For example, a 5-ns transient simulation of a *nand2* gate lasts ~ 1 s to run, while ~ 10 min are needed to find a single bias point of a SB-GNRFET in ViDES on the same machine.

3.4.2 Transistor-Level Evaluation

The minimum current in SB-GNRFETs occurs at $V_{GS} = \frac{1}{2}V_{DS}$, making the transistor ambipolar in the operating region. This minimum point, which is called ambipolar conduction point, occurs at $V_{GS} = \frac{1}{2}V_{DS}$ for a midgap

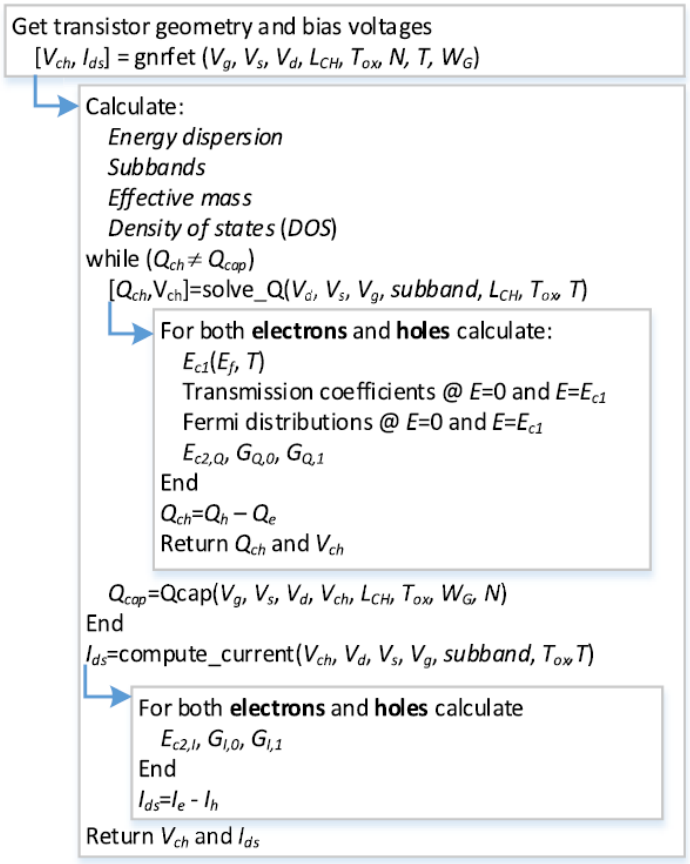

Figure 3.12: Flow of the SB-GNRFET compact model.

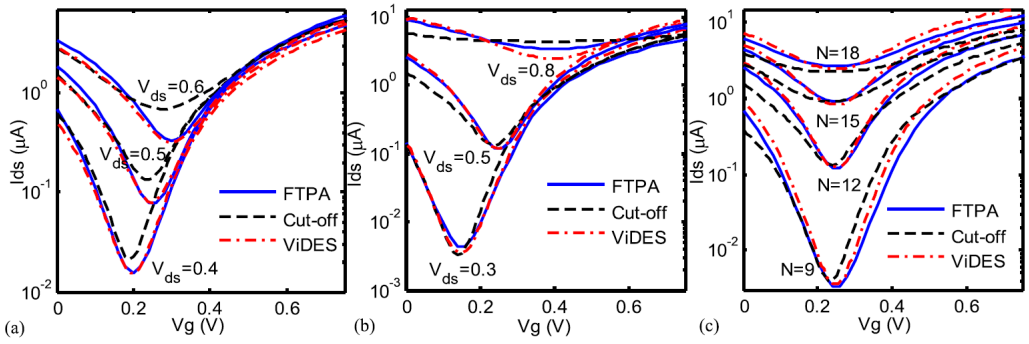

Figure 3.13: I-V curves of double-gate (DG) SB-GNRFET, **(a)** N=12, and T_{ox}=2 nm, **(b)** N=12, and T_{ox} =1.5 nm, **(c)** V_d=0.5V, and T_{ox}=1.5 nm.

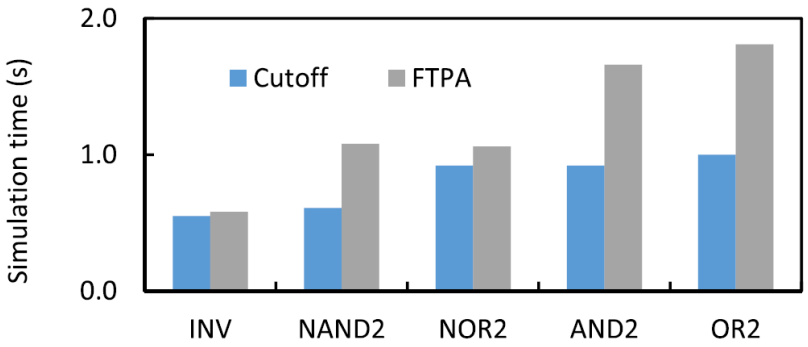

Figure 3.14: Simulation running time comparison between the cutoff model and the FTPA model.

Schottky barrier height. Ideally, the minimum current should occur in the OFF state when $V_{GS} = 0$. The minimum current point can be shifted to a different V_{GS} by tuning the gate work function using different metals [26, 27, 69]. An example of SB-GNRFETs working under $V_{DD} = 0.5$ V with ideal amount of shifting that produces desired p-type and n-type characteristics is shown in Figure 3.15 (a). Fabricated ambipolar transistors have adopted the shifting technique in [69] and [70]. In [69], the best shifting achieved was ~ 0.25 V for pMOS and ~ 1 V for nMOS, using Pd and Al as gates, respectively. Note that [69] and [70] focus on CNTs, but the techniques are expected to work for GNRs as they share many physical properties. In this work, we shift 0.25 V for p-type and 0.25 V for n-type transistors for ideally balanced device strength and performance.

Figure 3.15 (b) shows the I-V curves of MOS-GNRFET and SB-GNRFET, as well as the 16-nm HP Si-CMOS and 16-nm low-power (LP) Si-CMOS transistors from PTMs for comparison. We are using silicon as the channel material for MG PTM devices in all of our simulations. The transistor dimensions of the GNRFETs are scaled to match the PTM libraries. SB-GNRFET has a shifted I-V curve in order to obtain minimum current at $V_{GS} = 0$ V. We investigated both the ideal cases ($p_r = 0$) and the non-ideal cases of SB-GNRFETs to account for the line edge roughness due to imperfection from process variation that is unlikely to be avoided in practice [71, 72]. The fabrication of precise GNRs $\sim$ 2-nm wide is reported in [21] and [40], which is close to the 1.5-nm GNRFET we are modeling. The approach in [40] shows the possibility of fabricating such narrow ribbons with perfect edges. Therefore, for the non-ideal cases, we set the line edge roughness probability p_r

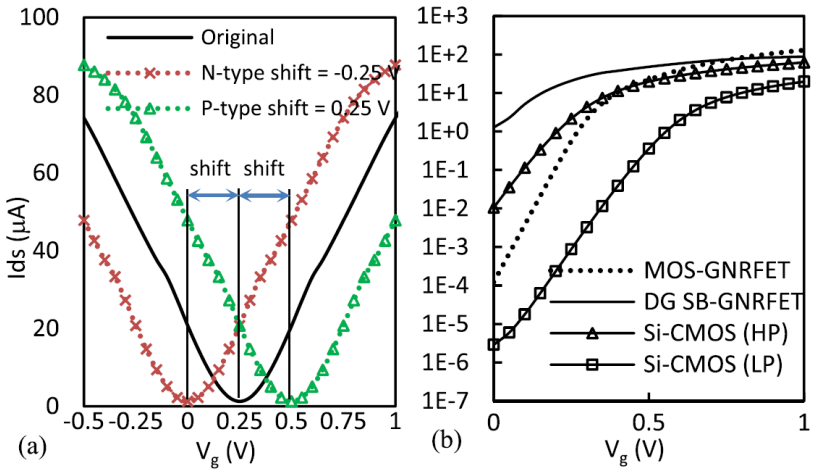

Figure 3.15: (a) Shifting ambipolar I-V curves of SB-GNRFET. (b) I_{DS} vs. V_{GS} for MOS-GNRFET, SB-GNRFET, 16-nm HP Si-CMOS, and 16-nm LP Si-CMOS, respectively. For GNRFET devices, $N = 12$, $nRib = 6$, and $T_{ox} = 1$ nm.

$= 0.1$. Higher p_r makes the I_{on}/I_{off} too small to be of practical use, and it is not the case of some fabricated GNRFETs. On the other hand, the ideal cases give an optimistic insight on how well GNRFET circuits may perform once fabrication technology becomes more mature. Overall, SB-GNRFET has the highest current, while the LP Si-CMOS has the lowest. Table 3.2 shows the subthreshold swing S and I_{on}/I_{off} ratio of each device under, respectively, chosen V_{DD}. It is shown that ideal SB-GNRFETs have the highest subthreshold swing and lowest I_{on}/I_{off} ratio.

Table 3.2: Subthreshold Swing and I_{on}/I_{off} Ratio of Each Device

Device	p_r	S (mV/dec)	I_{on}/I_{off}	V_{DD} (V)
Si-CMOS (HP)	-	93.46	3.49E+3	0.7
MOS-GNRFET w/ Res	0	66.67	1.81E+5	0.5
	0.1	140.85	9.85E+1	0.5
SB-GNRFET	0	133.51	3.76E+1	0.5
	0.1	735.29	2.49E+0	0.5

Here, we define $I_{on} = \text{I}\ (V_{GS} = 0.75, V_{DS} = 0.5)$ and $I_{off} = \text{I}\ (V_{GS} = 0.25, V_{DS} = 0.5)$ by assuming $V_{DD} = 0.5$ and an I-V curve shifting of ± 0.25 V. Figure 3.16 shows the effect of N on transistor current. In [19], a periodic effect on band gaps with respect to N is reported, and our model tracks the

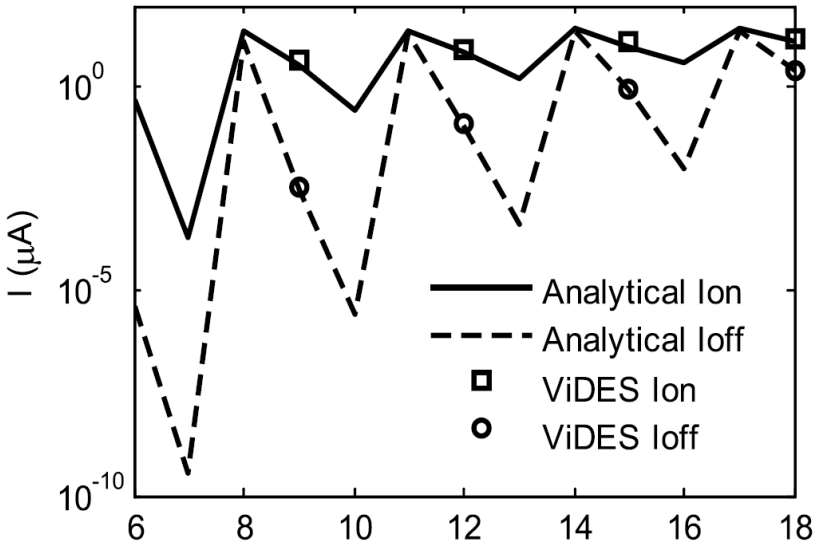

Figure 3.16: I_{on} and I_{off} vs. N. Comparing with ViDES results.

periodicity very well. 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, resulting 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, resulting in the highest I_{on}/I_{off} ratio. However, I_{on} is lower because the large band gaps prevent carriers from occupying the channel. In addition, the I_{on}/I_{off} ratio tends to decrease as N increases.

In order to determine a suitable supply voltage V_{DD} for SB-GNRFET, we simulated a buffer chain circuit under different V_{DD} and measured the EDP. Figure 3.17 shows the impact of supply voltage V_{DD} on the EDP. In general, lower V_{DD} results in lower EDP that indicates better overall performance, especially for the cases with line edge roughness. However, considering factors, such as noise margins and limitation on voltage shifting of SB-GNRFETs, $V_{DD} = 0.5$ V is chosen as the operating V_{DD} of GNRFETs. Note that ideal GNRFETs outperform non-ideal ones significantly in terms of EDP. SB-GNRFETs have also better I_{on}/I_{off} ratio in low V_{DD} values, as shown in Figure 3.18. Figure 3.19 shows the effect of T_{ox}. A higher T_{ox} implies a smaller tunneling probability through the Schottky barrier that results in the lower current. However, I_{on}/I_{off} ratio increases with higher T_{ox} values. Figure 3.20 shows the effect of line edge roughness on transistor current. Line edge roughness reduces I_{on}. It also reduces band gaps, which leads to

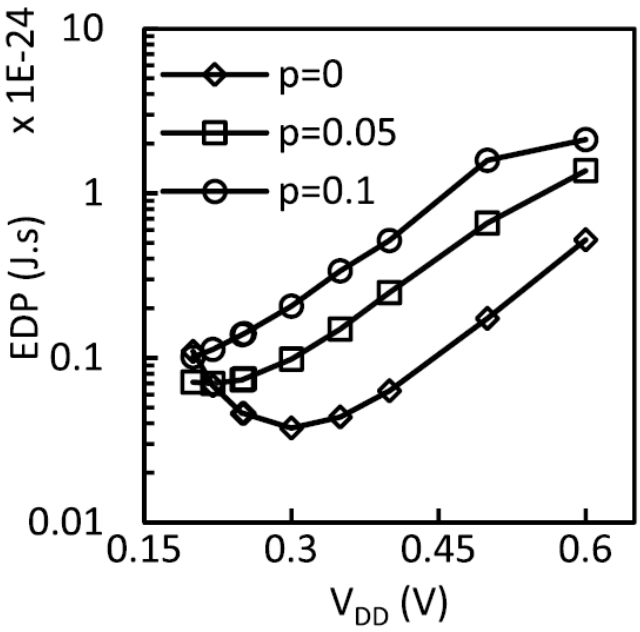

Figure 3.17: EDP vs. V_{DD} of buffer chain circuit.

an increase in I_{off}.

3.4.3 Circuit-Level Evaluation

We implemented the proposed model in HSPICE as a library for both SG and DG SB-GNRFET. To evaluate SB-GNRFET circuits, we implemented digital logic circuits with the aforementioned library. We also implemented circuits with the MG Si-CMOS HP library from PTM as a comparison with Si-based technology. We choose the HP Si-CMOS library over the LP one because the SB-GNRFET is known to be a HP rather than a LP device, according to the study in [2]. We implemented basic logic gates, such as *inv*, *nand2*, *nand3*, *nor2*, *xor2*, *nor3*, *nand4*, and *buf_chain*, as well as benchmark circuits *c17* and *c432* from ISCAS '85, *b02* from ITC '99, *s27* from ISCAS '89, carry generator for the third bit of a carry look-ahead adder (*cla*), one-bit full adder *1bit_fa*, and four-bit full adder *4bit_fa* in the following five technologies: (1) HP MG Si-CMOS from PTM, (2) ideal DG SB-GNRFET, (3) non-ideal DG SB-GNRFET, (4) ideal SG SB-GNRFET, and (5) non-ideal SG SB-GNRFET. We performed delay and power analysis on these technology nodes, as reported in the following subsections.

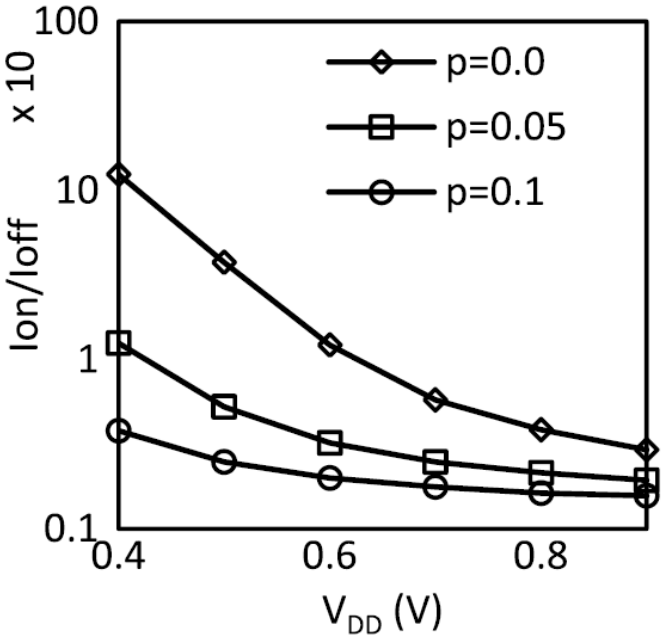

Figure 3.18: I_{on}/I_{off} vs. V_{DD} of SB-GNRFET.

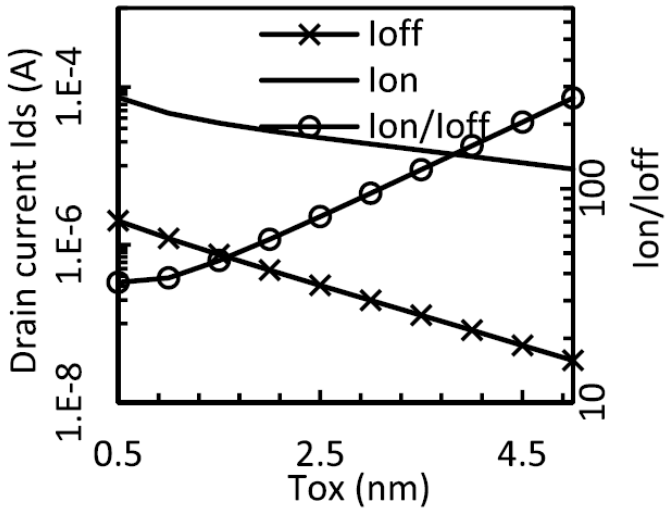

Figure 3.19: I_{on}, I_{off}, and I_{on}/I_{off} vs. T_{ox}, oxide thickness.

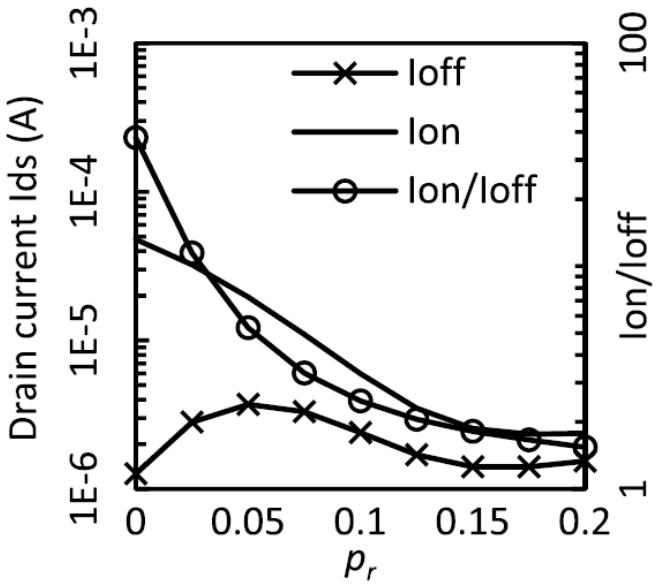

Figure 3.20: I_{on}, I_{off} and I_{on}/I_{off} vs. p_r, line edge roughness probability.

3.4.3.1 Properties of an Inverter

In this section, we analyze properties of an inverter built with SB-GNRFETs under $V_{DD} = 0.5$ V. We used our SPICE model to perform DC and transient analysis of the inverter. In Figure 3.21, V_{in} and V_{out} are the input and output voltages of the inverter, respectively. High line edge roughness probability p_r results in a higher propagation delay, as shown in Figure 3.22, due to the smaller transistor current. V_L and V_H are the inverter's low and high output voltages, respectively. Both V_L and V_H degrade with line edge roughness that results in the lower maximum swing, as shown in Figure 3.23.

Figure 3.24 shows the voltage transfer curves of an inverter built with SB-GNRFETs with different line edge roughness settings ($p_r = 0.05$ and $p_r = 0.1$). V_{in} and V_{out} are the input and output voltages of the inverter, respectively. High line edge roughness probability p_r results in a lower voltage swing. Figure 3.25 shows the normalized noise margin of different inverters. Si-CMOS inverter has nearly the same noise margin for different V_{DD} values, while the SB-GNRFET has better noise margin at low V_{DD} values. The line edge roughness significantly reduces the noise margin and the region of correct operation.

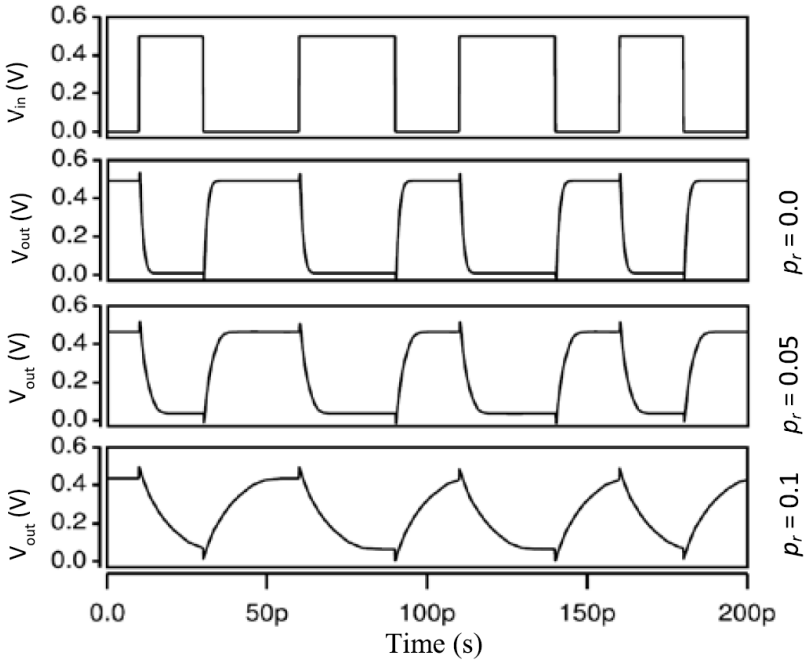

Figure 3.21: Effect of line edge roughness on an inverter's output waveform.

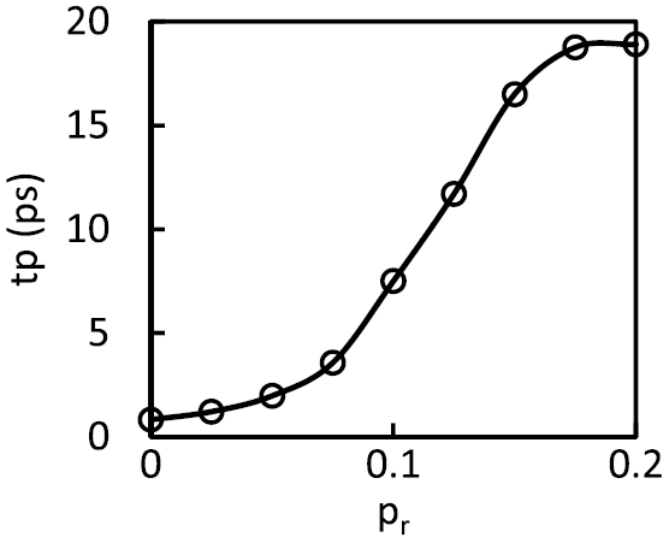

Figure 3.22: Effect of line edge roughness on an inverter's propagation delay.

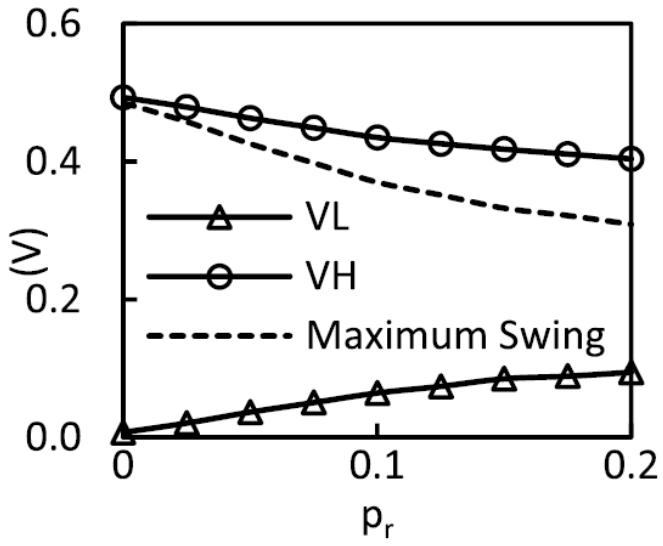

Figure 3.23: Effect of line edge roughness on an inverter's output voltage levels and maximum swing.

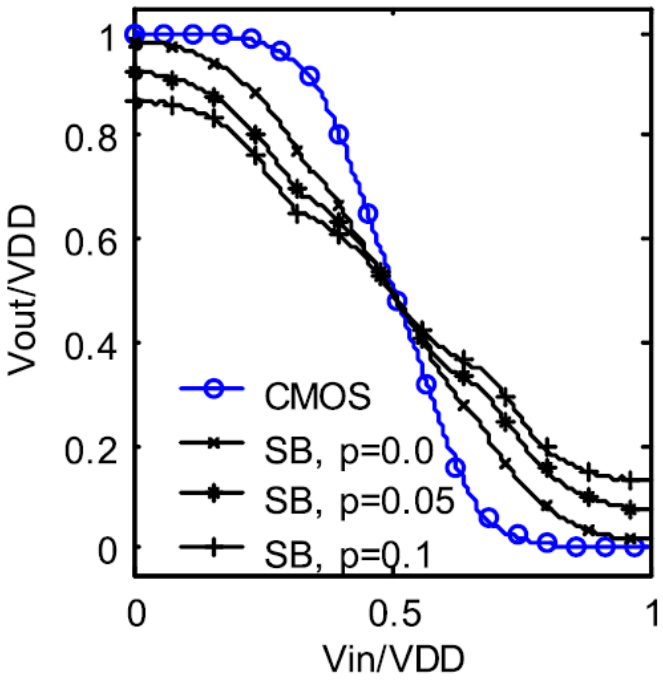

Figure 3.24: Transfer characteristic of SB-GNRFET inverter with different settings compared with Si-CMOS. Voltages are normalized to V_{DD} in each technology.

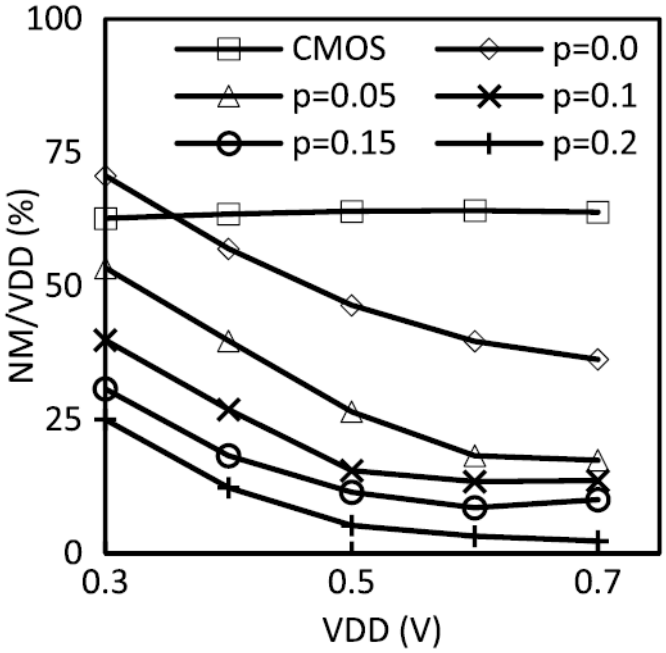

Figure 3.25: Normalized noise margin of inverters.

3.4.3.2 Supply Voltage Scaling

We investigate the effects of supply voltage (V_{DD}) scaling on SB-GNRFETs. Average delay, power, and EDP of benchmark circuits are reported in Figure 3.26. We show that the delay is nearly constant across different supply voltages, but power scales down as V_{DD} decreases. As a result, the EDP also gets better as V_{DD} decreases. This indicates that the SB-GNRFET has good potential in terms of low V_{DD} computing.

We also show that non-ideal SB-GNRFETs with process variation result in a large increase in the delay and EDP. In addition, we show that DG and SG SB-GNRFETs do not have significant difference in terms of circuit-level performance. This is because the I-V curves of SG and DG do not differ by a lot. The I_{on} is $\sim$ 30% different, while the I_{off} is only $\sim$ 5% different. Given the higher cost of manufacturing DG designs and the limited performance advantage, it may not be always desirable to prefer DG designs over SG.

3.4.3.3 Cross-Technology Comparison

Simulation results of basic logic gates and benchmark circuits are presented in Figures 3.27 and 3.28, respectively. We report maximum delay, dynamic

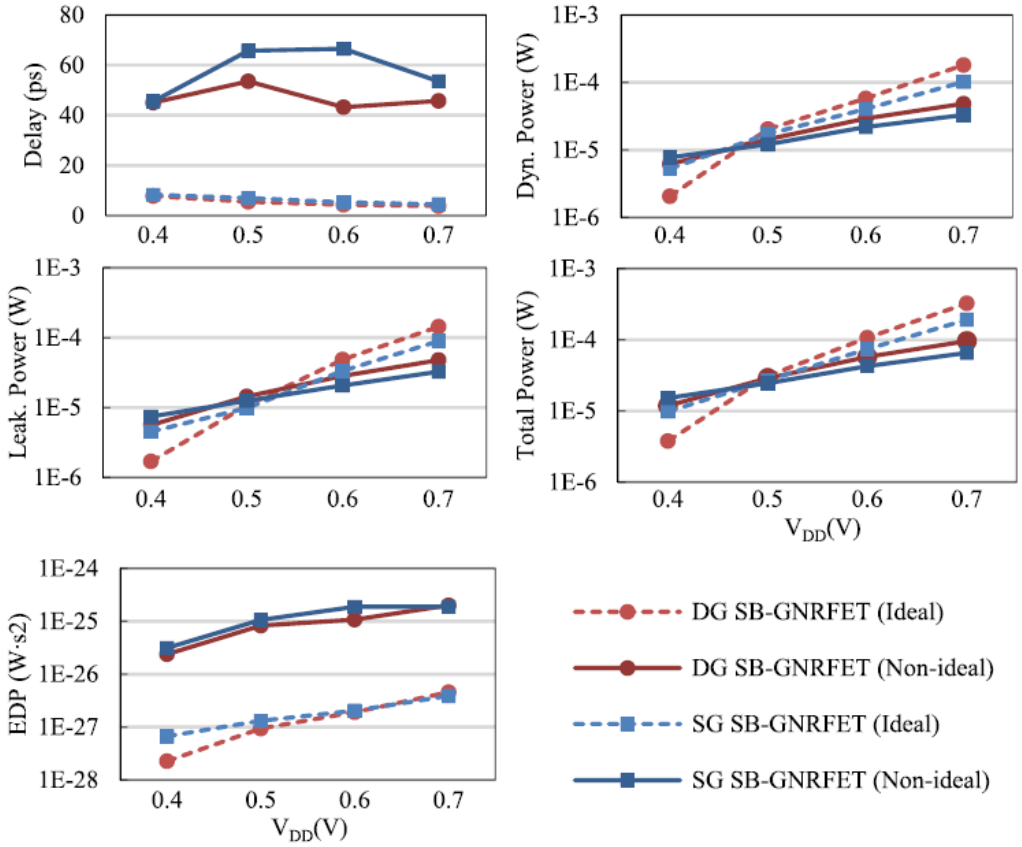

Figure 3.26: Average delay, power, and EDP of benchmarks with V_{DD} scaling.

power, leakage power, total power, and EDP for all the circuits. These parameters are measured using various SPICE simulation runs with careful setup. The delay and dynamic power are measured using SPICE transient analysis with randomly generated input signal pattern for different circuits. For fair comparison, we used similar input patterns for each circuit in different technologies. The leakage power is reported as the average value of the leakage powers of the circuit at all possible input logic level combinations that are obtained using separate transient SPICE simulation runs. The trends of delay, power, and EDP are mostly consistent across different circuits. Average values of five figures of merits: (1) delay, (2) dynamic power, (3) leakage power, (4) total power, and (5) EDP for different technologies are given in Figure 3.29. We show that ideal SB-GNRFET, either SG or DG, has lower delay than Si-CMOS (27% or 22% of Si-CMOS). SG SB-GNRFET has slightly lower dynamic power than DG SB-GNRFET due to its lower I_{on}. They have comparable leakage power due to very similar I_{off}. However, since SB-GNRFETs have higher I_{off} than Si-CMOS, as shown in Figure 3.15 (b), their leakage power is 4× and 10× higher than that of Si-CMOS for ideal and non-ideal SB-GNRFETs, respectively.

In terms of total power dissipation and EDP, SB-GNRFET outperforms Si-CMOS significantly. Ideal DG (SG) SB-GNRFET consumes only 57% (39.7%) total power, while non-ideal DG (SG) SB-GNRFET consumes 55.2% (44.7%) total power as compared with Si-CMOS. The EDP of the ideal DG (SG) SB-GNRFET is only 2.2% (2.9%) of the MG Si-CMOS, while non-ideal DG (SG) SB-GNRFET has 55.4% (81.5%) EDP as compared with Si-CMOS. Non-ideal SB-GNRFET consumes less dynamic power than the ideal one because its I_{on} is decreased by the presence of line edge roughness. This decrease also reduces the I_{on}/I_{off} ratio, as shown in Figure 3.20, making the transistor less efficient, as can be seen in the degradation in delay and EDP.

3.4.3.4 Technology Scaling

We investigate the trend of delay and power when the transistor size scales down. The technology nodes available in the MG Si-CMOS PTM library are 16, 14, 10, and 7 nm, with the supply voltage of 0.85, 0.8, 0.75, and 0.7 V, respectively. We scale the SB-GNRFET accordingly by putting in

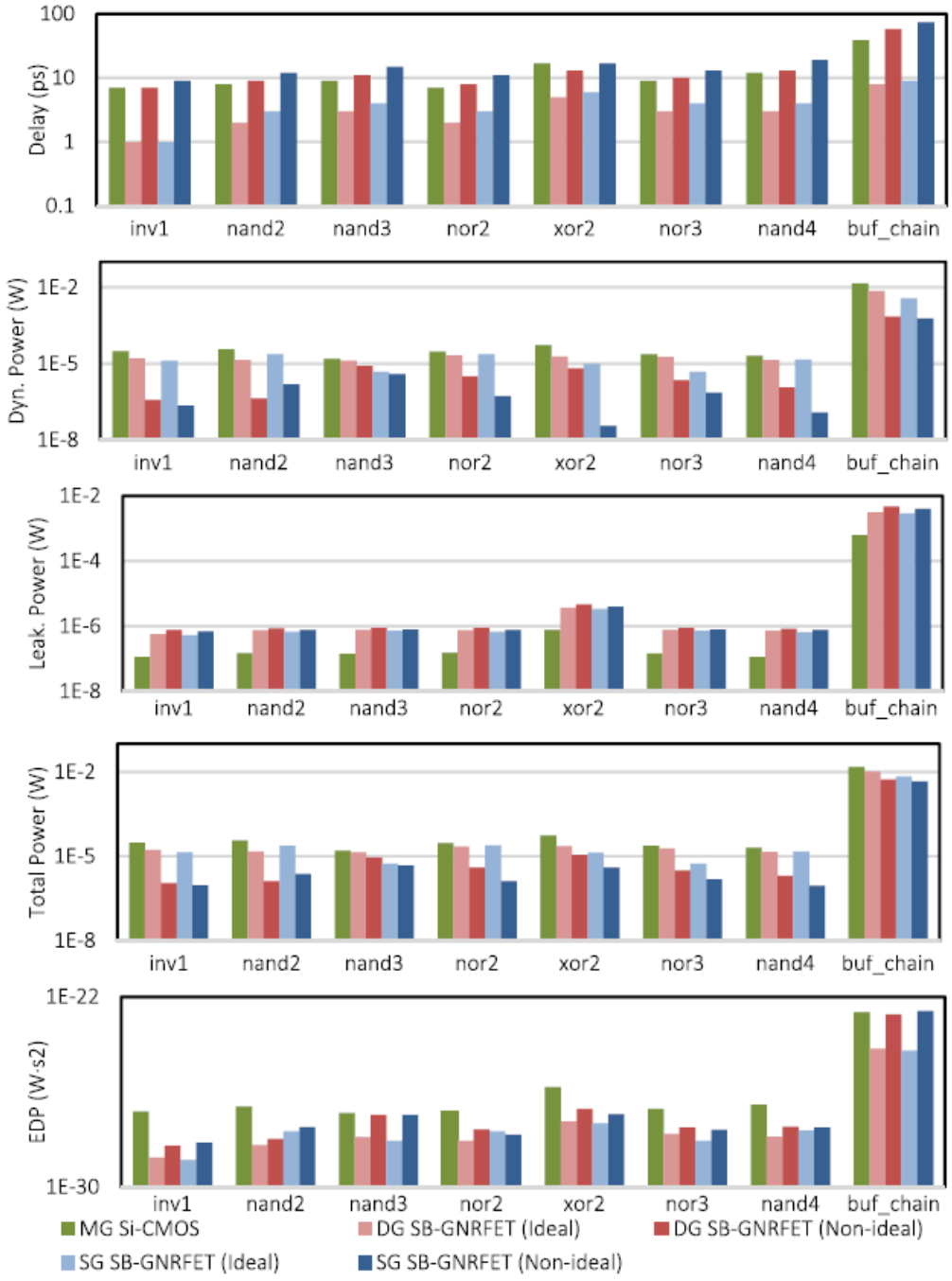

Figure 3.27: Delay, power, and EDP of basic logic gates.

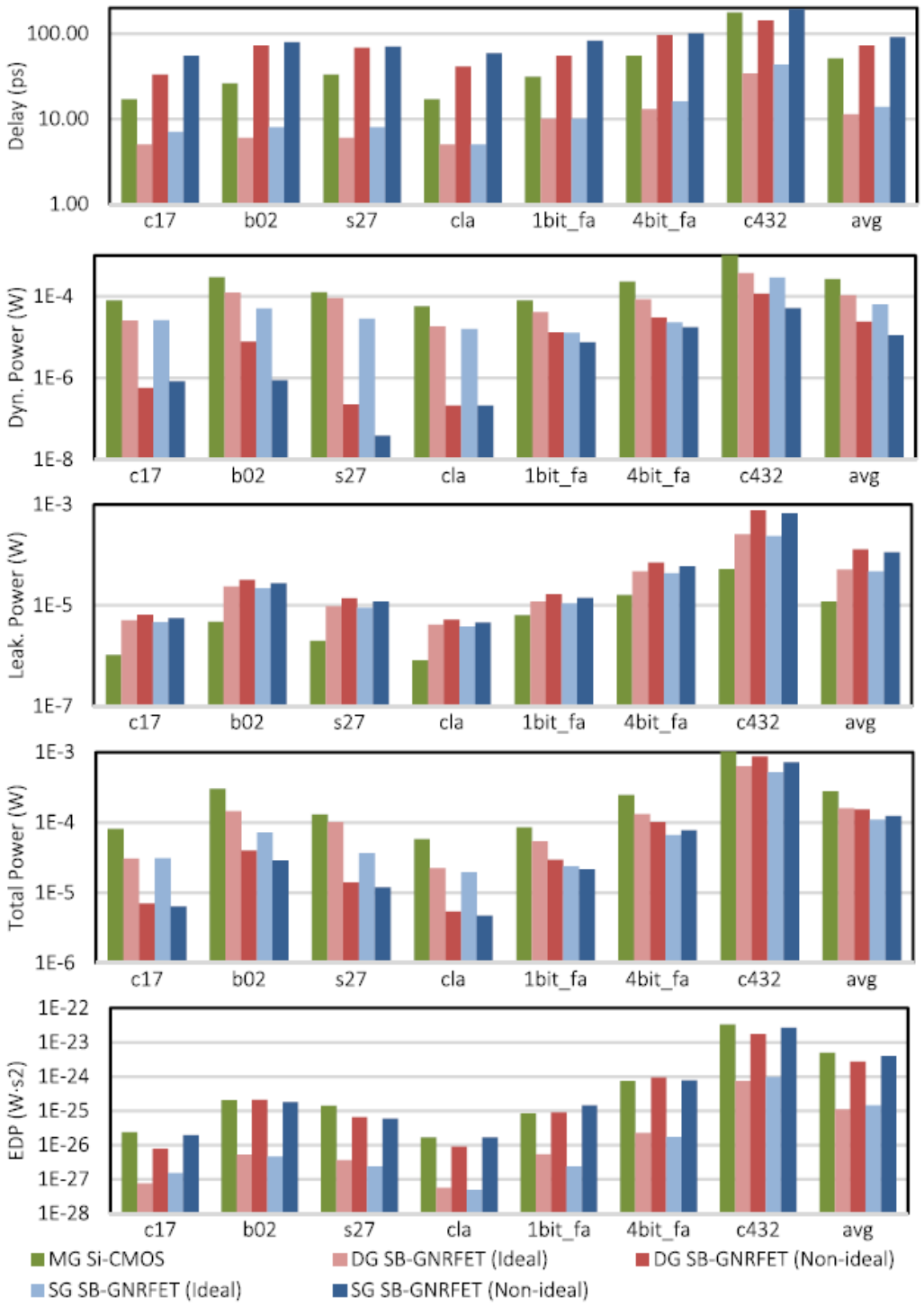

Figure 3.28: Delay, power, and EDP of benchmarks circuits.

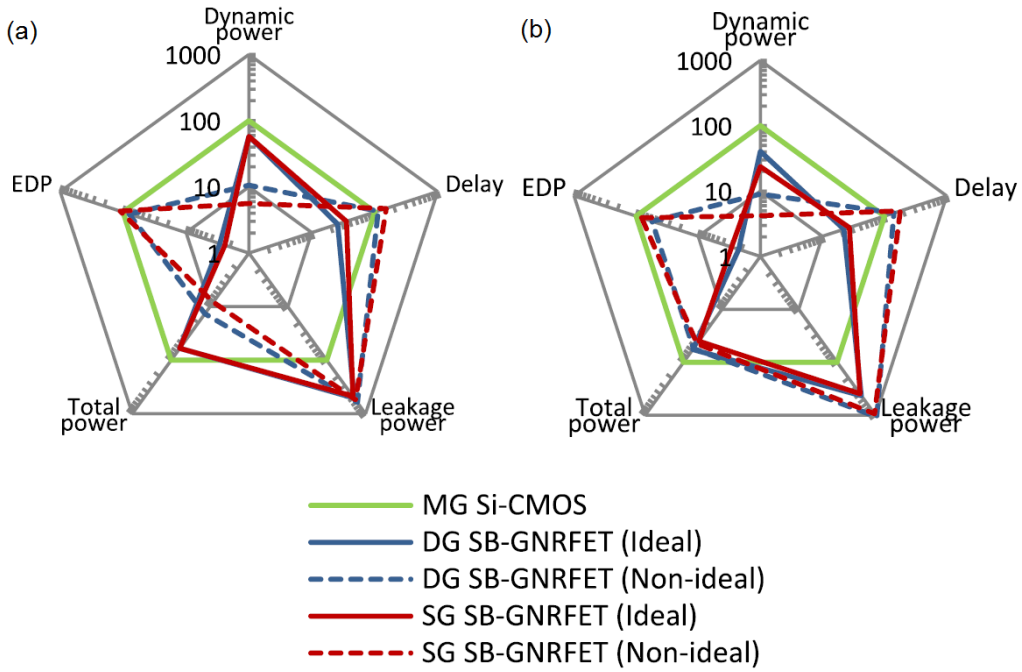

Figure 3.29: Comparison of different technologies based on five figures of merits: delay, dynamic power, leakage power, total power, and EDP. The number for each figure of merit is the average value that is normalized to the value of MG Si-CMOS in that category and presented in percentage. (a) Basic logic gates. (b) Benchmark circuits.

Table 3.3: Transistor Sizes

Node (nm)	nRib	$W_{GNR,eff}$ (nm)	W_{Gate} (nm)	W_{MOSFET} (nm)	W_{MG} (nm)
7	3	4.80	16.80	14	22
10	4	6.40	22.40	20	28
14	5	8.00	28.00	28	32
16	6	9.60	33.60	32	42

appropriate numbers of ribbons in the transistor of Figure 3.3. The scale values, as given in Table 3.3, are chosen in such a way that the width of different transistors (W_G, W_{MOSFET}, and W_{MG}) are almost the same. Figure 3.30 shows the delay, dynamic power, leakage power, total power, and EDP of Si-CMOS, ideal, and non-ideal DG and SG SB-GNRFET as the transistor size scales down. We show that delay, dynamic power, leakage power, total power, and EDP all scale down consistently across different technologies as the transistor size scales down, except for the leakage power of Si-CMOS, which increases with the downscaling.

As a result, SB-GNRFETs show an advantage on the trend of power over Si-CMOS; Si-CMOS has almost constant total power with technology scaling, while the total power of SB-GNRFETs reduce as the transistor size scales down. Moreover, ideal SB-GNRFETs give one to two orders of magnitude lower EDP than that of Si-CMOS. This indicates SB-GNRFETs potential in HP and LP computing. Nonideal SB-GNRFETs have reduced current and consume less power than the ideal ones, resulting in a very low total power. However, due to the significant increase in delay, the EDP of non-ideal SB-GNRFETs is much worse than that of the ideal ones [2]. This brings up a pressing challenge for the fabrication technology to significantly improve the quality of GNR, especially investigating new techniques that can help to produce GNRs with much smoother edges.

Nonetheless, an advantage of GNRFETs in terms of transistor size scaling is that they can scale based on the number of ribbons in one transistor. Therefore, only the driving strength of the transistor is scaled down with the transistor size, and the effect from scaling is linear. This results in stable and consistent circuit performance after scaling, which is also helpful in the circuit design process.

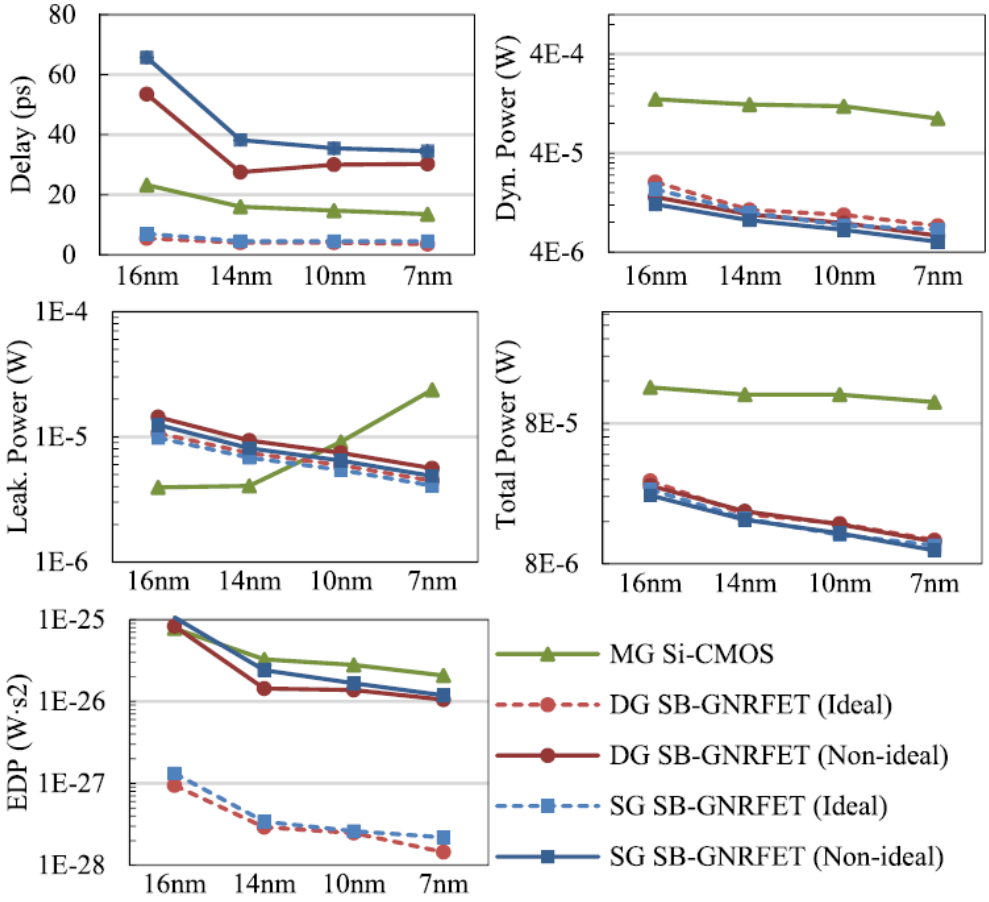

Figure 3.30: Average delay, power, and EDP with technology scaling on the four benchmark circuits (*c17*, *b02*, *s27*, and *cla*).

3.5 Conclusion

We present a physics-based analytical model for the current-voltage (I-V) characteristics of SB-GNRFETs. This model captures the effects of different parameters and process variation effects, such as channel width W_{CH}, channel length L_{CH}, oxide thickness T_{ox}, temperature T, and line edge roughness. We analytically approximate both carrier charge density and carrier current in order to achieve closed-form expressions that make compact SPICE-compatible modeling of SB-GNRFETs possible. This model enables accurate and fast circuit simulation of both SG and DG SB-GNRFET circuits.

Based on this model, we performed device-level performance evaluations. It is shown that SB-GNRFET has the highest current, highest subthreshold swing, and lowest I_{on}/I_{off} ratio among the studied transistors. We also performed circuit-level performance evaluations on SG and DG SB-GNRFETs, with and without the impact of process variation. SB-GNRFET circuits are also compared with Si-CMOS-based circuits.

We show that SG and DG SB-GNRFETs do not have a significant difference in delay, power, and EDP performance. SB-GNRFETs are also shown to perform better than Si-CMOS in terms of EDP. In the 16-nm node, the EDP is only $\sim 2.5\%$ of that of Si-CMOS for the ideal case, and $\sim 68\%$ for the non-ideal case. In the 7-nm node, the EDP is $\sim 0.88\%$ of that of Si-CMOS for the ideal case, and $\sim 54\%$ for the non-ideal case. These results indicate that the ideal SB-GNRFET has great performance and scalability, demonstrating its potential in becoming a next-generation device. However, advanced fabrication techniques are required to remove the non-idealities faced by GNR fabrication now, before GNRFETs can become a competitive alternative solution beyond Si-CMOS.

CHAPTER 4

GNRFET AS FUTURE LOW-POWER DEVICES

4.1 Introduction

Although conventional Si-CMOS devices have prevailed in the semiconductor industry for decades, it has been increasingly difficult to keep up with Moore's law due to the various challenges imposed by the extremely small feature sizes, including increased wire resistivity, significant mobility degradation, and large dopant fluctuations. Various new materials and devices have emerged as potential successors of Si-CMOS. Among them, graphene has drawn a lot of attention in recent years because of its outstanding electrical properties [11, 12, 13, 14]. Intrinsic graphene has been reported to have high carrier mobility, high carrier density, long mean free path, high thermal conductivity, and high robustness [15]. The thin, planar, and robust lattice structure of graphene makes it potentially more controllable and scalable for mass production and integration with existing Si-CMOS fabrication technology. For the same reason, graphene is also a candidate material for making flexible electronics.

Successful fabrication of graphene-based electronics has been demonstrated [55, 21, 40, 37, 41, 17, 39, 73, 74]. Sub 10-nm transistors based on graphene nano-ribbons (GNRs) have been fabricated in [21] with promising measurement results, indicating the feasibility of nanoscale production of graphene-based electronics. In [17], a fabrication methodology allowing graphene to be integrated with existing Si-CMOS technology is demonstrated. In addition, existing work on simulations of graphene-based circuits has shown that they can achieve lower energy-delay product (EDP) compared to Si-CMOS [26, 27, 1, 2]. As a result, graphene-based nanoelectronics are regarded as an emerging next-generation technology that is worth investigating.

While intrinsic 2-D graphene sheets have outstanding electrical properties,

they have zero band gaps, which make them excellent conductors instead of semiconductors. In order to open up band gaps and make it semiconducting, graphene can be patterned into 1-D narrow strips known as graphene nano-ribbons (GNRs). The band gap of a GNR is inversely proportional to its width. With width < 2 nm, GNRs exhibit good semiconducting properties. Transistors made of GNRs are called graphene nano-ribbon field effect transistors (GNRFETs). Both theoretical and experimental results have shown that GNRFETs can potentially be good transistors with high I_{on}/I_{off} ratio and low subthreshold swing [21], [23, 1, 2]. Note that 2-D graphene can also be made into graphene field-effect transistors (GFETs), which have a low I_{on}/I_{off} ratio due to the limited band gaps. They are more suitable for analog applications and thus are not further discussed in this chapter.

There are mainly two types of GNRFETs: Metal-Oxide-Semiconductor-(MOS-)type GNRFETs (MOS-GNRFETs) and Schottky-Barrier-type GNRFETs (SB-GNRFETs). MOS-GNRFETs have GNR-based drains, channels, and sources with an *n-i-n* or *p-i-p* doping profile. Current conduction inside MOS-GNRFETs is mostly based on thermionic conduction. SB-GNRFETs also have intrinsic GNR channels, but drains and sources are made of metal. As a result, Schottky barriers occur at the graphene-metal junctions, and the current conduction is mainly based on Schottky barrier tunneling. Studies have shown that MOS-GNRFETs have a higher I_{on}/I_{off} ratio and are more robust to the effect of process variation compared to SB-GNRFETs [24, 2], while SB-GNRFETs have the advantages of not introducing high contact resistance on the vias and a higher I_{on} [2]. There are also GNR tunneling field-effect transistors (GNRTFETs), which have an *n-i-p* or *p-i-n* doping profile, which creates significant band bending between drain, channel, and source. As a result, GNRTFETs operate based on the band-to-band tunneling effect of carriers. In this chapter, we focus on MOS-GNRFETs and SB-GNRFETs due to the availability of SPICE-compatible compact models that enables higher-level simulations and evaluations [1, 2].

In this chapter, we study and compare circuits made of MOS-GNRFETs and SB-GNRFETs in the following aspects:

- Practical issues of building circuits from GNRFETs.
- Device-level characteristics of MOS- and SB-GNRFETs.

- Gate-level analysis of an inverter of MOS- and SB-GNRFETs, especially in terms of noise margin.
- Circuit-level simulation results on benchmark circuits, which provide insightful information on how GNRFET-based circuits may perform when fabrication technology becomes mature.

The rest of the chapter is organized as follows: In Section 4.2, we review the background of graphene and GNR. In Section 4.3 we discuss how GNRFET circuits are built. In Sections 4.4 and 4.5, we analyze the transistor-level and gate-level properties of GNRFET circuits, respectively. In Section 4.6, we report circuit-level simulations results on delay and power. Finally, Section 4.7 concludes the chapter.

4.2 Background

4.2.1 Graphene and GNR Properties

Graphene is a sheet of carbon atoms tightly packed into a 2-D honeycomb lattice. It is a zero band-gap material, which makes it an excellent conductor by nature [11, 12, 13, 14]. Depending on the number of layers, graphene can be categorized into monolayer, bilayer, or multilayer graphene. The unbounded edges of graphene are usually passivated by absorbents such as hydrogen, oxygen, hydroxyl group, carboxyl group, and ammonia [14]. Graphene must be processed into GNRs in order to open a band gap and turn into a semiconductor. Theoretical work has shown that GNRs have band gaps inversely proportional to their widths [19].

In addition, chirality of GNRs define the energy gaps and determine the conductivity. GNRs are categorized into two types: armchair-GNRs (AGNR) and zigzag-GNRs (ZGNR) based on the chirality [20]. In AGNRs, the band gap follows a periodic pattern based on N [19]. For $N = 3p$ and $N = 3p + 1$ ($p \in \mathbf{N}$), the band gap is finite, and the GNR is semiconducting. For $N = 3p + 2$, the band gap is very small, making the GNR metallic [19]. There is a decreasing trend in band gaps as N increases. In other words, band gaps of AGNRs are generally inversely proportional to the widths [19]. In ZGNRs, metallic properties are observed when the edges are pristine, although the

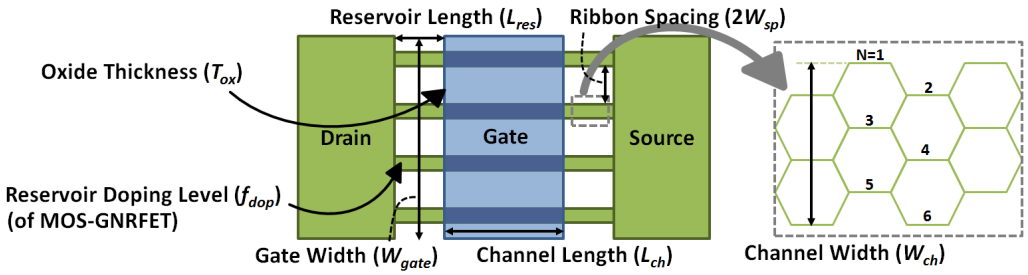

Figure 4.1: The structure of a four-ribbon GNRFET. Each ribbon is of armchair chirality. The four parallel ribbons have an equal width of W_{ch} and an equal spacing of $2W_{sp}$. The width of a GNR W_{ch} is commonly defined via the number of dimer lines N in the lattice structure. An example of an $N = 6$ GNR is illustrated on the right. A single metal gate is placed on top of the parallel ribbons. The gate of this four-ribbon GNRFET has a width of $W_{gate} = 4 \times (W_{ch} + 2W_{sp})$. The channels are defined as the portions of GNRs underneath the gate, with a length of L_{ch}. The reservoirs of a MOS-GNRFET are the portions or GNRs not covered by the gate. They have a length of L_{res} and are doped with molecular doping fraction f_{dop}. A wide common drain and a wide common source are shared by the ribbons.

band gap can be opened for zigzag GNRs with rough edges or those passivated with hydrogen atoms [22, 34]. In this chapter, we focus on GNRFETs made of AGNRs.

In this chapter, we denote the width of an AGNR to be W_{ch}. W_{ch} is commonly defined via the number of dimer lines N in the lattice structure of an AGNR [23], as illustrated on the right of Figure 4.1.

Due to process variation and limitation of manufacturing technology, GNRs with perfectly smooth edges may not always be produced. The absence of some atoms on the edges may result in fundamental changes in the properties of a GNR. Figure 4.2 shows an example of a GNR with edge roughness (top) as compared to one with perfectly smooth edges (bottom). Some segments of the GNR become narrower in width. Also, some segments become ZGNR. As discussed above, the change in width and chirality result in significant changes in the band structure. Therefore, edge roughness has a great impact on the properties of a GNR. The degree of roughness of the edges of a GNR can be defined through the *edge roughness probability* p_r, which is the probability of an atom on the edge being missing [23].

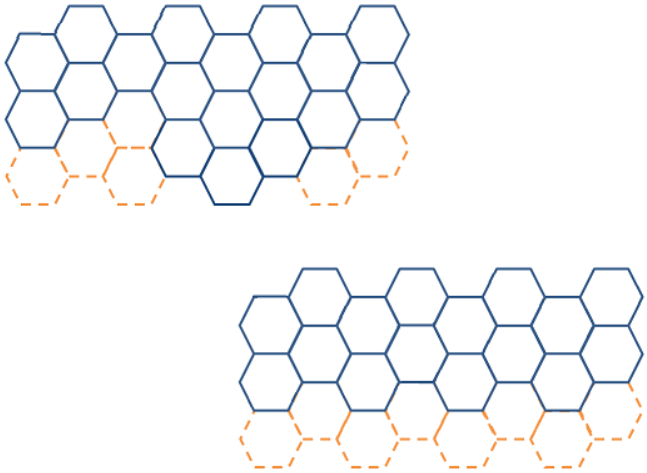

Figure 4.2: **(Top)** A GNR with a rough edge. The missing atoms not only cause the effective width in each segment to vary, but also cause some segments to become ZGNR. **(Bottom)** A GNR with width variation but with smooth edges, shown here as a comparison.

4.2.2 Fabrication of GNRs and GNRFETs

Fabrication of GNRs can be accomplished by techniques such as lithography [37, 41, 74], chemical synthesis [21, 37, 55, 73], or unzipping from carbon nanotubes [40]. In [37, 21], GNRs of sub-10-nm widths were successfully manufactured and demonstrated outstanding properties. In addition, the work [17] demonstrated the technique to integrate GNRs with existing Si-CMOS fabrication technology. The transfer-free, *in situ* fabrication of GNRFETs opened new opportunities for future GNRFET applications. All these examples of successful fabrications indicate that the manufacturing technology of GNRFETs is becoming more and more practical.

4.3 GNRFET Circuits

Before we start to evaluate and simulate GNRFET circuits, we propose a circuit architecture that is scalable to various technology nodes and practical in terms of manufacturing. In this way, the subsequent circuit evaluations are more realistic and representative.

4.3.1 Device Structure

The device structure under study is planar in order to be compatible with existing fabrication technology, as demonstrated in [17]. Both MOS-GNRFET and SB-GNRFET have a single metal top gate. Under the gate are multiple parallel GNRs with uniform spacing between them. The parallel GNRs increase the driving strength of the transistor, and the number of GNRs can be scaled to match a specific technology node. This structure is proposed and evaluated in [26, 27, 1, 2]. In addition, the technique of fabricating parallel GNR arrays has been demonstrated in [37]. For MOS-GNRFET, the drain and the source, called reservoirs, are heavily-doped GNRs. Depending on the dopant type, a MOS-GNRFET is either N-type or P-type. For SB-GNRFET, the drain and source are made of metal. Because SB-GNRFETs have an ambipolar I-V curve, they are neither N-type nor P-type by nature. With additional work function engineering, the I-V curve of SB-GNRFETs can be shifted such that they work as either N-type or P-type.

An example GNRFET with four parallel ribbons is shown in Figure 4.1. All ribbons are of armchair chirality in order for them to be semiconducting. Here, we define the gate width to be W_{gate}, the width of each ribbon to be W_{ch}, the channel length to be L_{ch}, the reservoir length to be L_{res}, the spacing between ribbons to be $2W_{sp}$, and the doping level of the drain and source of a MOS-GNRFET to be f_{dop}. The width of a GNR W_{ch} is commonly defined via N as described in Section 4.2.1 and illustrated on the right of Figure 4.1.

4.3.2 Circuit Architecture

We adopt a circuit architecture that integrates the transistor devices introduced in Section 4.3.1 with GNR and metal-based interconnects, as proposed in [1]. In the chosen circuit architecture, there are multiple metal (e.g. Cu) layers on top of a single graphene layer. The metal layers comprise of the metal gates of transistors and most of the interconnects. The single layer of graphene is placed on top of the bulk. It can be patterned into either the GNR parts of transistors or some of the interconnects in the case of MOS-GNRFETs. For MOS-GNRFETs, the drain and source are made of GNRs. As a data path has to connect from the output (the drain of a transistor) of a logic gate to the input (the gate of a transistor) of the next logic gate,

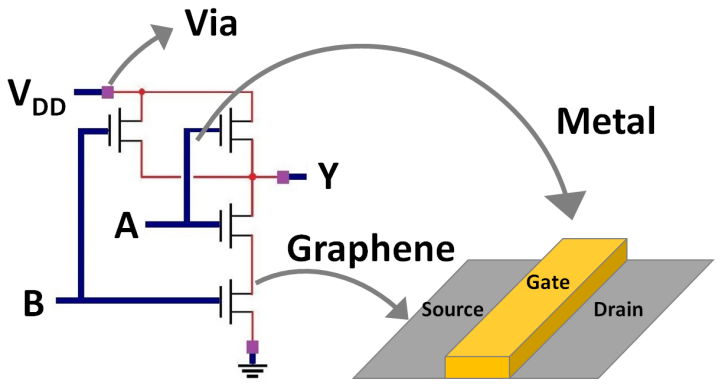

Figure 4.3: A *nand2* gate implemented in the proposed architecture of MOS-GNRFET circuits. Inputs A and B, output Y, and power rails V_{DD} and *gnd* are distributed on the metal layers (bold blue lines). Vias (purple squares) are needed to connect graphene and metal layers. Local interconnects between drains and sources are made of graphene (thin red lines), in order to avoid extra vias.

vias that connect the graphene layer and the upper metal layers are needed. Vias are assumed to be metal because vertical graphene vias have not been well studied [15]. Studies have shown that high contact resistance is introduced at graphene-metal junctions [42]. For a 50-nm wide via that connects the graphene layer to the upper metal layer, the contact resistance is estimated to be 20 kΩ [42]. As this can severely degrade the circuit performance, we choose to have local interconnects between drains and sources of MOS-GNRFETs to be made of graphene in order to avoid introducing extra vias and contact resistance. In Figure 4.3, we show a *nand2* gate implemented in the architecture introduced above to illustrate the concept.

On the other hand, SB-GNRFETs have metal drains and sources as opposed to MOS-GNRFETs. Schottky barriers are introduced at the junctions between the metal drain/source and the GNR-based channel. The effects of the graphene-metal contacts are considered in the transistor model as the Schottky barriers. As a result, no extra graphene-metal contact resistance is introduced in the circuit. Also, all interconnects are metal-based because all terminals of SB-GNRFETs (gates, drains, and sources) are made of metal. The concept is illustrated in Figure 4.4.

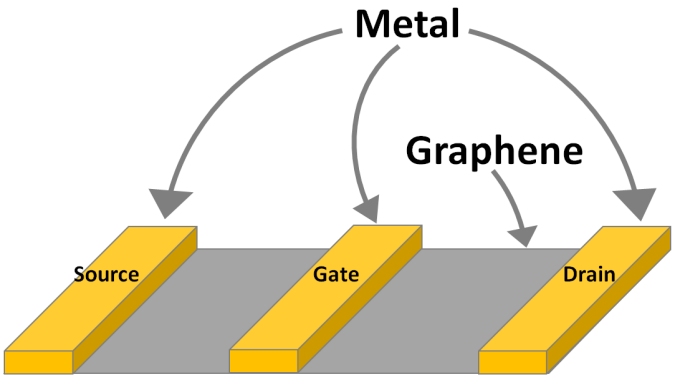

Figure 4.4: SB-GNRFETs have metal-based drains and sources as opposed to the GNR-based ones in MOS-GNRFETs. The effects from graphene-metal contact is modeled as the Schottky barrier inside the transistor, and no extra graphene-metal vias are needed.

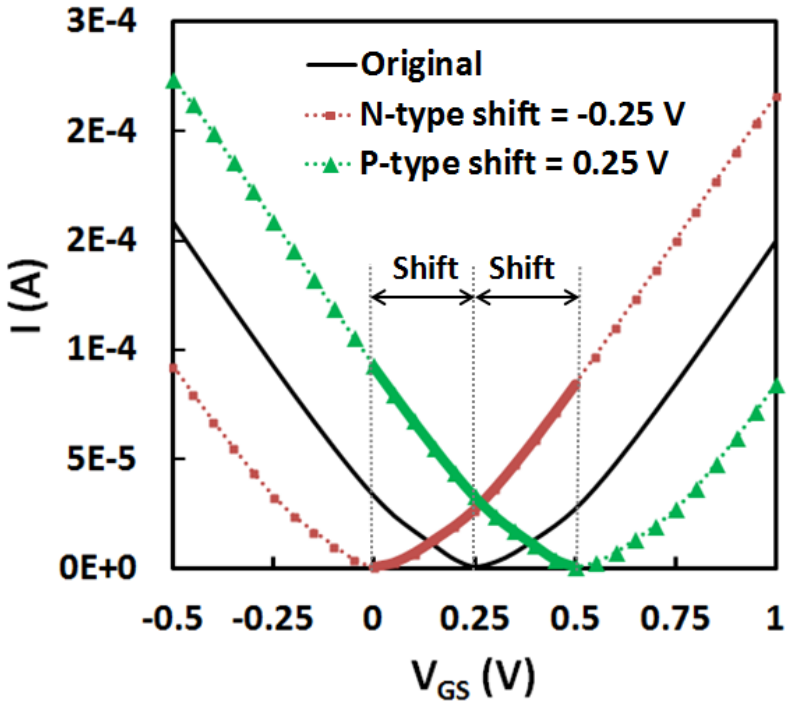

Figure 4.5: I-V curve shifting of SB-GNRFET in order to have proper I-V characteristics for P-type and N-type transistors. Ideally, the shifting amount should be $\frac{1}{2}V_{DS}$ since I_{off} occurs when $V_{GS} = \frac{1}{2}V_{DS}$. Legal operating regions after shifting are marked in red/green solids.

4.3.3 MOS-GNRFETs vs. SB-GNRFETs

In this subsection, we discuss the differences and address the practical issues in implementing circuits with MOS-GNRFETs and SB-GNRFETs.

First of all, MOS-GNRFETs have a monotonic I-V curve due to doping in the reservoirs. The dopants absorb minority carriers such that they do not create a large current. By choosing the type of the dopants, MOS-GNRFETs can be made into N-type or P-type transistors. On the contrary, SB-GNRFETs have an ambipolar I-V curve. By assuming a mid-gap Schottky barrier, SB-GNRFETs have minimum current when $V_{GS} = \frac{1}{2}V_{DS}$. This is because of the symmetrical Schottky barrier profiles for both holes and electrons imposed by this applied voltage. In this way, the currents contributed by holes and electrons have a similar magnitude to result in the minimum current for SB-GNRFETs. The ambipolar I-V curve of SB-GNRFETs is not suitable for conventional complementary MOS (CMOS) style logic design, although there have been studies on logic designs based on ambipolar transistors [69, 75]. In this work, we focus on CMOS-style designs for the widely available technologies that are applicable. In order to obtain a proper N-type or P-type I-V curve that is compatible with CMOS-style designs, work function engineering is applied to SB-GNRFETs to shift the I-V curve.

In terms of I-V curve shifting, practical techniques are limited and may not be able to shift any arbitrary amount, which leads to unbalanced N-type and P-type characteristics. For example, the best shifting achieved in [69] for SB-type carbon nanotube transistors was $\sim$ 0.25 V for P-type and $\sim$ -1.0 V for N-type, by using Pd and Al as gates, respectively. GNRFETs are expected to work similarly. When the P-type and N-type transistors are extremely imbalanced, the circuit becomes less robust or even nonfunctional. In Section 4.5.2, we analyze an example of an *inv* built with SB-GNRFETs of imbalanced I-V curve shifting to demonstrate the potential performance loss. In the rest of the chapter, we assume a perfect balanced shifting can be achieved for both P-type and N-type transistors in order to have a fair comparison of SB-GNRFET circuits with other technologies.

MOS-GNRFETs have a higher I_{on}/I_{off} ratio than SB-GNRFETs, which means they can be turned on or off more properly [24]. SB-GNRFETs can however exhibit a higher I_{on} after I-V curve shifting. For these reasons, MOS-GNRFETs are generally regarded as more suitable for digital circuit

applications.

On the other hand, the performance of MOS-GNRFET circuits is limited by the inevitable graphene-metal contact resistance introduced by vias, as discussed in Section 4.3.2. The absence of graphene-metal contact resistance in SB-GNRFET circuits is potentially a strength compared to MOS-GNRFET circuits. In addition, the doping level inside MOS-GNRFET reservoirs is potentially susceptible to process variation, which results in significant changes in transistor characteristics and circuit performance [1]. SB-GNRFETs have undoped metal drain and source and are therefore free of this problem.

To summarize, MOS-GNRFETs and SB-GNRFETs have their respective strengths and weaknesses. It is difficult to draw conclusion on which type is the more competitive device based on only the transistor-level analysis. The subsequent simulations compare MOS- and SB-GNRFET circuits in detail in order to provide an accurate insight on how these circuits perform and compare.

4.4 Transistor-Level Characteristics

In this section, we review the transistor-level properties of MOS-GNRFETs and SB-GNRFETs. Based on the explorations in [26, 27, 1, 2], both MOS-GNRFETs and SB-GNRFETs work well under a low V_{DD} around 0.5 V. Therefore, we choose a nominal $V_{DD} = 0.5$ V for both MOS-GNRFETs and SB-GNRFETs.

Because of the ambipolar nature of SB-GNRFETs, I_{off} does not occur naturally when $V_{GS} = 0$. Voltage shifting is required to create proper I-V curves for PMOS and NMOS transistors. Ideally, the shift amount should be $\frac{1}{2}V_{DS}$ since I_{off} occurs when $V_{GS} = \frac{1}{2}V_{DS}$, and the direction of shifting should be opposite for PMOS and NMOS. An example of SB-GNRFETs working under $V_{DD} = 0.5$ V with ideal amount of shifting that produces desired PMOS and NMOS characteristics is illustrated in Figure 4.5.

Figure 4.6 shows the I-V curves of MOS-GNRFET and SB-GNRFET, as well as the 16-nm High-Performance (HP) Si-CMOS and 16-nm Low-Power (LP) Si-CMOS transistors from Predictive Technology Models (PTM) for comparison. The transistor dimensions of the GNRFETs are scaled to

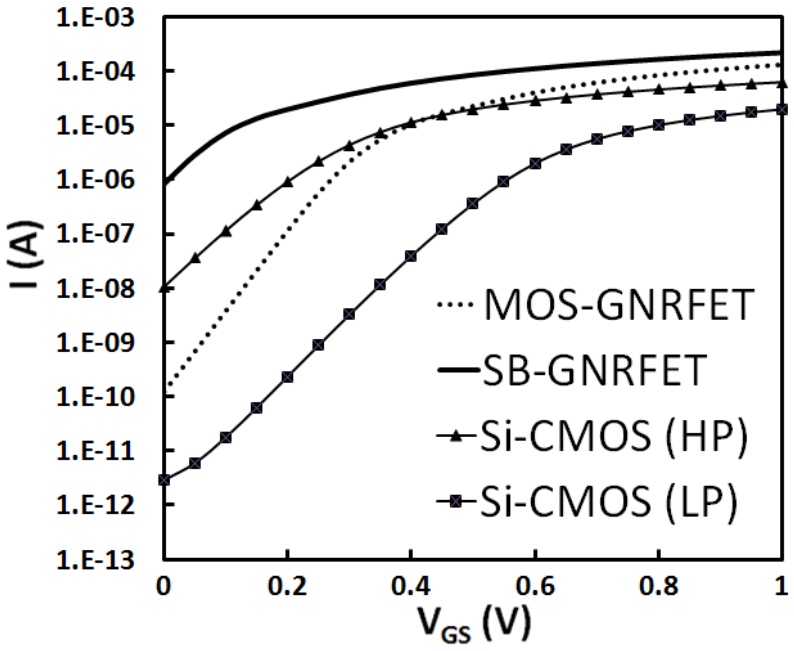

Figure 4.6: I_{DS} vs. V_{GS} for MOS-GNRFET, SB-GNRFET, 16-nm High-Performance (HP) Si-CMOS, 16-nm Low-Power (LP) Si-CMOS, respectively.

match the PTM libraries. SB-GNRFET has a shifted I-V curve in order to obtain minimum current at $V_{GS} = 0$ V. Overall, SB-GNRFET has the highest current, while the LP Si-CMOS has the lowest.

Table 4.1 shows the subthreshold swing S and I_{on}/I_{off} ratio of each device under respectively chosen V_{DD}. It is shown that ideal MOS-GNRFETs have the lowest subthreshold swing and I_{on}/I_{off} ratio. However, as edge roughness comes into play, the transistor characteristics become comparable or even worse than Si-CMOS. SB-GNRFETs have comparable subthreshold swing to that of Si-CMOS, and they have the lowest I_{on}/I_{off} ratio.

4.5 Gate-Level Analysis

4.5.1 Properties of an Inverter

In this section, we analyze the gate-level properties of an inverter built with either MOS-GNRFETs or SB-GNRFETs under $V_{DD} = 0.5$ V. Figure 4.7 (left) shows the voltage transfer curves of inverters built with MOS-GNRFETs and

Table 4.1: Transistor Properties

Device	p_r	S (mV/dec)	I_{on}/I_{off}	V_{DD} (V)
Si-CMOS (HP)	–	93.46	3.49E+03	0.7
Si-CMOS (LP)	–	86.96	5.12E+06	0.9
MOS-GNRFET	0	66.67	1.81E+05	0.5
	0.1	140.85	9.85E+01	0.5
SB-GNRFET	0	87.72	1.02E+02	0.5
	0.1	76.92	7.64E+00	0.5

Subthreshold swing and I_{on}/I_{off} ratio of each device. For GNRFETs, devices of different line edge roughness (p_r) are listed as well.

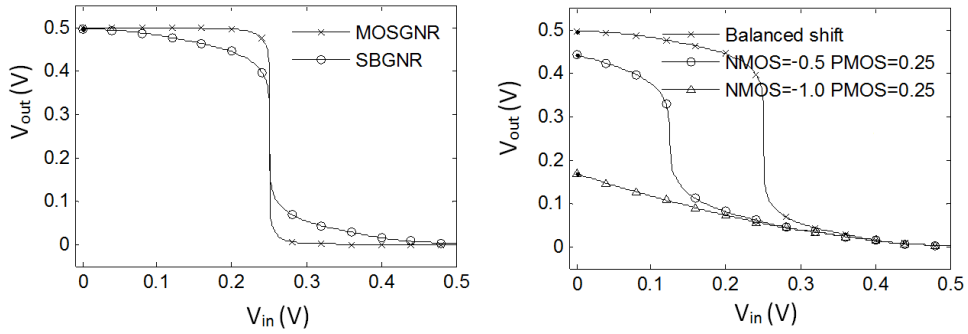

Figure 4.7: **(Left)** Voltage transfer curves of MOS-GNRFET and SB-GNRFETs inverters, respectively. **(Right)** Voltage transfer curves of inverters built with SB-GNRFETs with different voltage shifting.

SB-GNRFETs, respectively. V_{in} and V_{out} are the input and output voltages of the inverter, respectively. Both inverters have full voltage swings. The ranges of V_{in} that result in correct operations are indicated by V_{IL} and V_{IH}, the maximum voltage for a valid low input and the minimum voltage for a valid high input, respectively. For the MOS-GNRFET inverter, $V_{IL} =$ 0.23 V, $V_{IH} =$ 0.27 V, and the noise margin is 92% of V_{DD}. For the SB-GNRFET inverter, $V_{IL} = 0.22$ V, $V_{IH} = 0.28$ V, and the noise margin is 88% of V_{DD}. Note that the V_{out} of the SB-GNRFET inverter is very sensitive to the change in V_{in}, and hence it is more susceptible to noise. On the other hand, the MOS-GNRFET inverter has a sharp voltage transfer curve, which makes it more robust, as V_{out} almost stays the same as V_{in} approaches V_{IL} or V_{IH}.

4.5.2 Effects of Voltage Shifting for SB-GNRFETs

In this section, we emphasize the importance of balanced voltage shifting for SB-GNRFETs by using an inverter as an example. Figure 4.7 (right) shows the voltage transfer curves of SB-GNRFET inverters with N-type transistors of different shifting. With a balanced shifting, the voltage transfer curve is symmetric. However, as the I-V curve of the N-type transistor is shifted more, it becomes more difficult for the transistor to be turned off because I_{off} is increased. Eventually, V_{out} corresponding to $V_{in} = 0$ V does not rise to $V_{DD} = 0.5$ V when the N-type transistor is shifted for -1.0 V. From this example, we show that SB-GNRFET circuits require proper shifting in order to function correctly.

4.6 Circuit-Level Evaluation

We performed circuit-level simulations by using the SPICE models of MOS- and SB-GNRFETs from [1, 2]. The SB-GNRFET model from [2] is further calibrated for higher accuracy in the I_{off} region. We simulated the ideal cases of MOS- and SB-GNRFETs and the non-ideal cases with edge roughness $p_r = 0.1$. In both ideal and non-ideal MOS-GNRFET circuits, contact resistance of 20 kΩ is added to all graphene-metal vias, as explained in Section 4.3.2. The 16-nm HP and LP Si-CMOS libraries from PTM are adopted as comparisons, and the GNRFETs are set to have matching dimensions. In Sections 4.6.1 and 4.6.2, the impacts of supply voltage and process variation are evaluated on seven-stage, fanout-of-four buffer chains. In Section 4.6.3, benchmark circuits such as *c17* and *c432* from ISCAS '85, *b02* from ITC '99, *s27* from ISCAS '89, carry generator for the third bit of a carry look-ahead adder (*cla*), and a 4-bit full adder (*4bit_fa*) are implemented in SPICE. Sequential circuits *b02* and *s27* are converted into combinational circuits by the pseudo prime input method. We report delay, dynamic power, leakage power, total power, and energy-delay product (EDP) from circuits implemented in the six technology nodes in the following experiments.

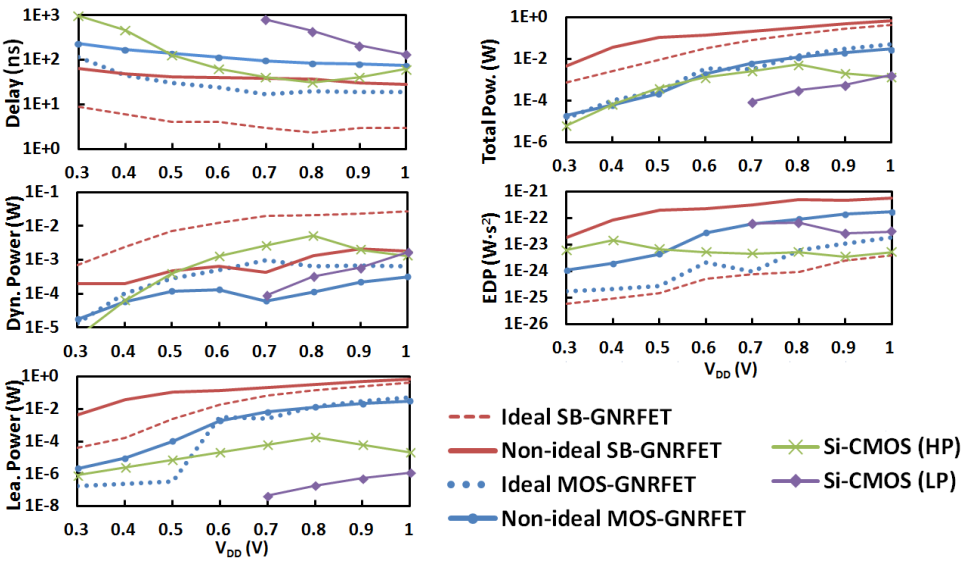

Figure 4.8: Delay, power, and EDP vs. V_{DD}. $V_{DD} < 0.7$ V results in incorrect functions for Si-CMOS (LP).

4.6.1 Impact of Supply Voltage

Figure 4.8 shows the impact of supply voltage V_{DD}. Lower V_{DD} results in higher delay and lower power in general. For GNRFETs, lower V_{DD} also results in lower EDP, which indicates better overall performance. However, considering factors such as noise margins and limitation on voltage shifting of SB-GNRFETs, $V_{DD} = 0.5$ V is chosen as the operating V_{DD} of GNRFETs. Si-CMOS (HP) has the lowest EDP at its designed nominal $V_{DD} = 0.7$ V, while for Si-CMOS (LP) it is $V_{DD} = 0.9$ V. Note that ideal GNRFETs outperform non-ideal ones significantly in terms of delay and EDP, but dynamic power reduces for non-ideal SB-GNRFETs due to the drop in I_{on}.

4.6.2 Impact of Process Variation

Figure 4.9 shows the impact of ribbon width N (W_{CH}). The results are consistent with the periodic band gaps in terms of N as reported in [19]. For examples, $N = 3p + 2$ (8, 11, 14) gives a small band gap, resulting in almost equally high I_{on} and I_{off}, corresponding to low delay and high power. $N = 3p+1$ (10, 13, 16) gives the largest band gap with low I_{on} and very low I_{off}, resulting in the highest I_{on}/I_{off} ratio. Therefore, the power, especially the leakage power, is the lowest. $N = 3p$ (9, 12, 15) gives a moderate band

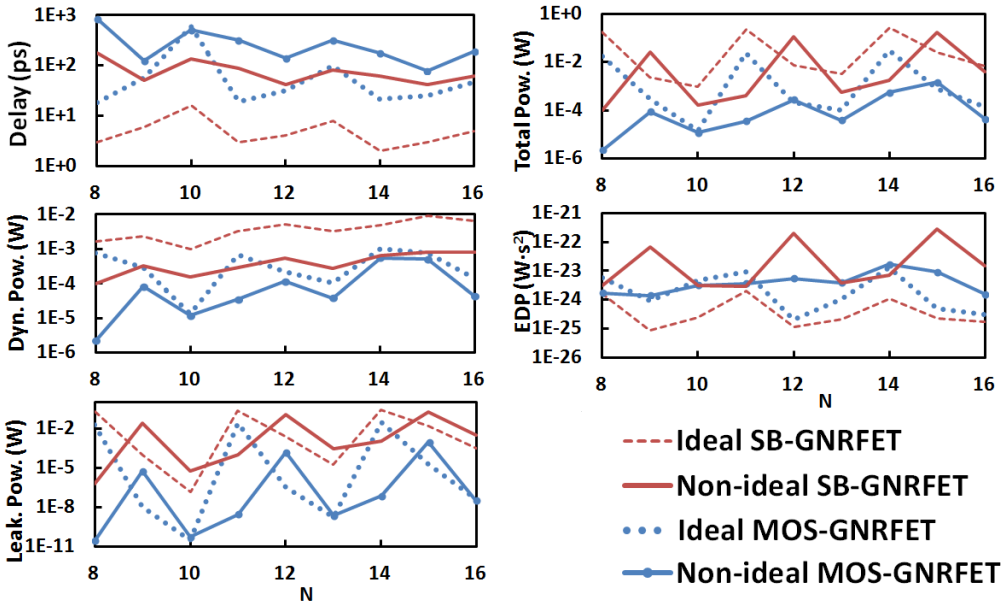

Figure 4.9: Delay, power, and EDP vs. N.

gap, and the delay and power performance is between the other two cases, with EDP being the lowest. Under the influence of edge roughness, the effective band gaps fall between the band gaps corresponding to an effective width N_{eff} between N and $N-2$, making the periodic effect not so significant. Also, the scattering effect causes the current to drop. As a result, delay is generally higher and power is generally lower compared to the ideal cases.

Figure 4.10 shows the impact of oxide thickness T_{ox} and channel length L_{CH}. In general, changes in T_{ox} or L_{CH} affect delay, power, or EDP only within one order of magnitude. The increase in T_{ox} causes I_{on} to drop, and thus increases the delay. The increase in L_{CH} results in larger gate capacitance, and therefore it increases the delay as well. Leakage power is not significantly affected by L_{CH} except for the case of MOS-GNRFETs, in which I_{off} is increased due to less control of the channel from the gate. On the other hand, edge roughness has a very high impact on delay, dynamic power, and EDP for SB-GNRFETs. Also, it significantly increases the leakage power for MOS-GNRFETs.

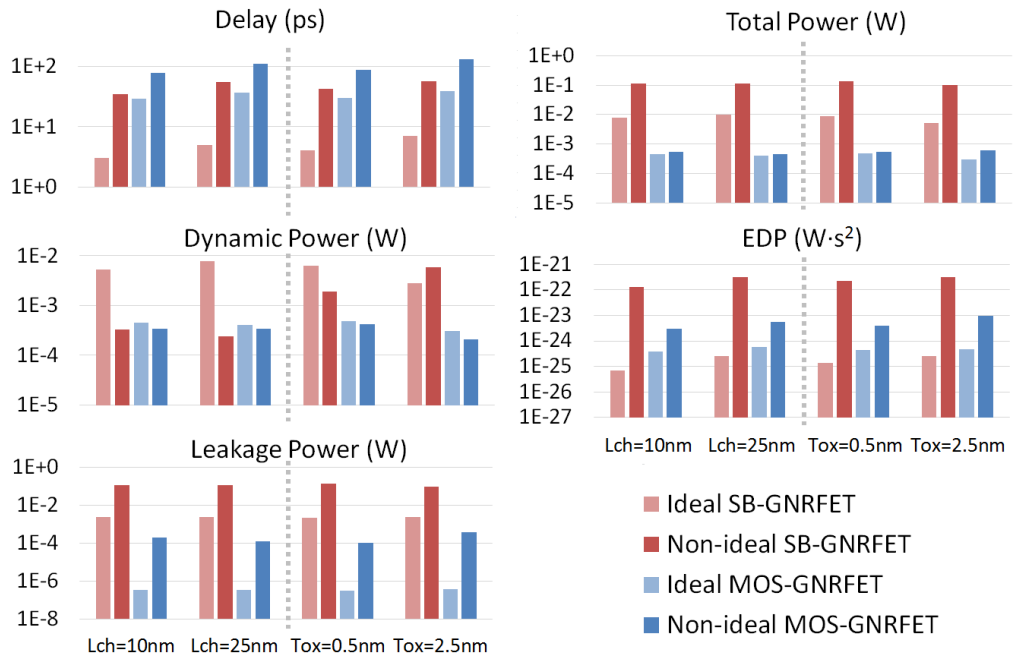

Figure 4.10: Delay, power, and EDP vs. T_{ox} and L_{CH}. For T_{ox} variation, L_{CH} and other parameters are set to the default device dimensions in [1, 2]. Similarly, for L_{CH} variation, T_{ox} and other parameters are set to default device dimensions in [1, 2].

4.6.3 Cross-Technology Comparison

Figure 4.11 shows the circuit performance of different technology nodes: Si-CMOS (HP), Si-CMOS (LP), ideal MOS-GNRFET, non-ideal MOS-GNRFET, ideal SB-GNRFET, and non-ideal SB-GNRFET. Ideal SB-GNRFET has the lowest delay and EDP and the highest power. It is suitable for high-performance, high-energy-efficiency applications. Ideal MOS-GNRFET has comparable delay with Si-CMOS (HP) but consumes much lower power. Compared with Si-CMOS (LP), ideal MOS-GNRFET has similar power but lower delay. Compared with both Si-CMOS technology nodes, MOS-GNRFET has better potential in low-power applications. Again, edge roughness significantly degrades the delay and EDP advantage of SB-GNRFET, making the EDP highest among all technologies except for Si-CMOS (LP). Also, it degrades the delay and leakage power advantage of MOS-GNRFET.

To summarize, ideal MOS-GNRFET consumes 18% and 54% total power as compared to Si-CMOS (HP) and Si-CMOS (LP), respectively, while for non-ideal MOS-GNRFET it is 35% and 102%. Ideal SB-GNRFET consumes

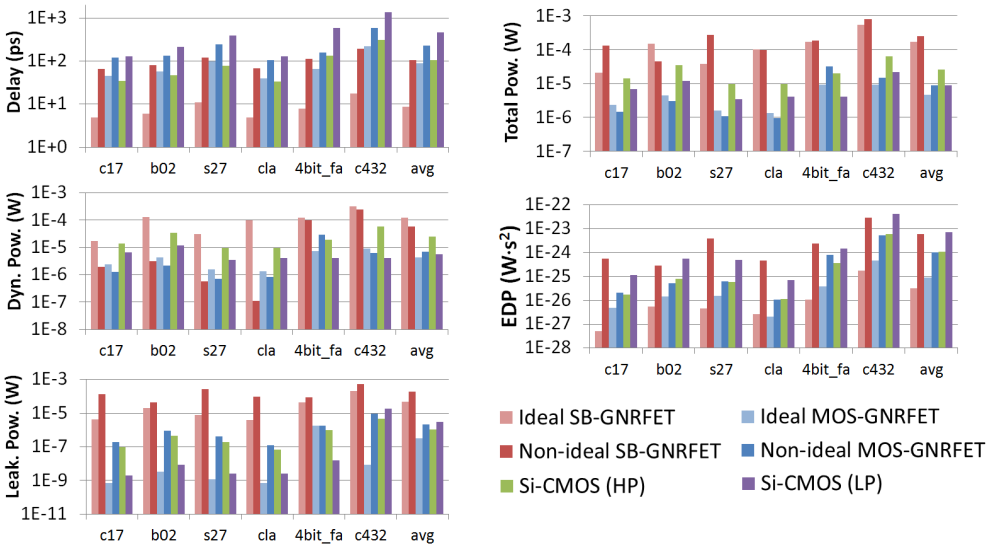

Figure 4.11: Delay, power, and EDP from different technology nodes.

6.6X and 19.4X total power as compared to Si-CMOS (HP) and Si-CMOS (LP), respectively, while for non-ideal SB-GNRFET it is 9.8X and 28.8X. Meanwhile, ideal MOS-GNRFET has 8% and 1.25% EDP compared to Si-CMOS (HP) and Si-CMOS (LP), respectively, while for non-ideal MOS-GNRFET it is 93% and 14.3%. Ideal SB-GNRFET has 3% and 0.45% EDP compared to Si-CMOS (HP) and Si-CMOS (LP), respectively, while for non-ideal SB-GNRFET it is 5.4X and 83.5%. SB-GNRFET has a much lower I_{on}/I_{off} ratio (~ 100) to begin with than MOS-GNRFET ($\sim 2 \times 10^5$). As the I_{on}/I_{off} ratio worsens with edge roughness, SB-GNRFET's performance becomes very bad (with $I_{on}/I_{off} < 10$), degrading faster than that of MOS-GNRFET, which still maintains a reasonable I_{on}/I_{off} ratio ~ 100 under the effect of edge roughness.

From the above simulations, we have predicted that MOS-GNRFET consumes lower power than Si-CMOS (HP) and has lower delay than Si-CMOS (LP). Also, SB-GNRFET has a very low delay although the power consumption is high. There are a variety of reasons for this.

First of all, GNRFET is a 1-D quantum wire, while Si-CMOS is a 3-D bulk device. The density of states for carriers to occupy in a 1-D quantum wire is much lower than that of a 3-D bulk device. Therefore, the number of carriers present in GNRFET is fewer. Secondly, GNRFET has a long mean free path. Therefore, the current conduction in GNRFET is mainly based on ballistic transport. On the contrary, bulk Si has a shorter mean free

path, and the drift current induced by the applied electric field dominates the current conduction in Si-CMOS. Moreover, MOS-GNRFETs operate mainly based on ballistic transport, while SB-GNRFETs operate mainly based on Schottky barrier tunneling. The different mechanisms of current conduction make these devices inherently different.

GNRFET operates at a lower V_{DD} of 0.5 V compared to 0.7 V or 0.9 V in the case of Si-CMOS. The lower V_{DD} of GNRFET reduces both dynamic and leakage power. The I-V curves of these technology nodes are shown in Figure 4.6.

It is shown that MOS-GNRFET has a lower subthreshold swing, which gives a lower I_{off} and makes its I_{on}/I_{off} ratio higher than that of Si-CMOS given the same voltage range. Moreover, a GNRFET is composed of multiple thin and narrow ribbons, while Si-CMOS is made of bulk Si. This makes the load capacitance of GNRFET smaller than that of Si-CMOS. As a result, dynamic power of MOS-GNRFET is reduced due to the smaller load capacitance C_L. SB-GNRFET consumes comparable dynamic power with that of Si-CMOS (HP) even though it has a higher I_{on}, also because of the smaller C_L.

For the same reason, the effective area of the current conduction in GNRFET is smaller than that of Si-CMOS. In the OFF state, current conduction in both MOS-GNRFET and Si-CMOS is based on diffusion. The smaller effective area and the lower number of states for carriers in MOS-GNRFET results in its OFF current lower than that of Si-CMOS. On the contrary, the ON current in MOS-GNRFET is dominated by ballistic transport, while the ON current in Si-CMOS is mainly based on drifting. Due to the lack of scattering, carriers in ballistic transport move faster than the drifting carriers. This makes the MOS-GNRFET and Si-CMOS (HP) having similar ON currents despite having different effective areas of current conduction. The Si-CMOS (LP) technology has a higher threshold voltage than Si-CMOS (HP), making both I_{on} and I_{off} lower.

On the other hand, SB-GNRFET has a high I_{off} due to the lack of doped reservoirs to absorb minority carriers to reduce current. The shifted I-V curve of SB-GNRFET also gives it a higher I_{on}. This results in higher power and lower delay of SB-GNRFETs.

Figure 4.12 summarizes the performance of each technology node. To sum up, MOS-GNRFET demonstrates a higher subthreshold swing and a higher

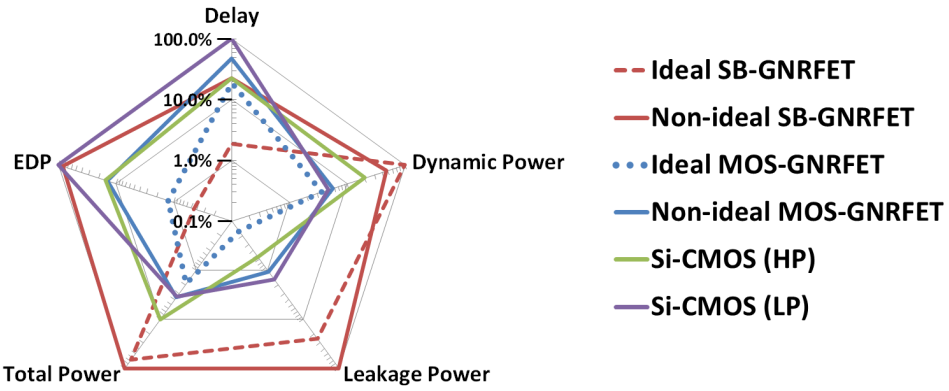

Figure 4.12: Comparison of six different technologies based on five performance parameters: delay, dynamic power, leakage power, total power, and EDP. The numbers for performance category are average values that are normalized to the maximum value of that category and presented in percentage. Each axis is in log scale and has a maximum value of 100%. Ideal MOS-GNRFET has the lowest star, which means the best overall performance, while ideal SB-GNRFET has the best EDP and delay.

I_{on}/I_{off} ratio compared to Si-CMOS. Also, MOS-GNRFET has a lower I_{off} than Si-CMOS (HP). As a result, MOS-GNRFET has lower dynamic power, lower leakage power, and comparable delay compared to Si-CMOS (HP), as well as similar power and better delay compared to Si-CMOS (LP). Despite the presence of contact resistance in MOS-GNRFET circuits, the EDP is only slightly higher than that of SB-GNRFET. SB-GNRFET gives a very low EDP, indicating excellent performance and efficiency. However, the high power dissipation due to higher current, especially in terms of leakage power, restricts its use in power-critical applications.

4.7 Conclusion

In this chapter, we reviewed practical issues in the implementation of GNRFET-based circuits, discussed transistor-level and gate-level properties of GNRFET circuits, and reported circuit-level simulations results on delay and power performance of GNRFET circuits. Our simulations show that both MOS-GNRFET and SB-GNRFET perform better than Si-CMOS in terms of EDP under ideal cases. Also, MOS-GNRFET has great potential in low-power applications, while SB-GNRFET is suitable for high-performance ap-

plications with its excellent EDP. However, when edge roughness is present, the delay and power benefits from both types of GNRFETs are significantly reduced. Future refinement in GNRFET fabrication techniques is critical in order to make GNRFET a competitive technology.

CHAPTER 5

ASYMMETRIC GATE SB-GNRFET FOR LOW-POWER DESIGN

5.1 Introduction

Graphene has received much attention as a base material for nanoelectronic devices because of the outstanding physical and electrical properties. There are two varieties of graphene nanoribbon field-effect transistors (GNRFETs): Schottky-barrier (SB)-type and MOSFET-type [3]. In MOS-type GNRFETs, the reservoirs are doped with donors or acceptors. In SB-type devices, metals are used for contacts and graphene as the base channel material, which results in the formation of SBs at the interfaces. One advantage of SB-GNRFETs is that they require no additional doping in the contacts or the channel. Therefore, it reduces the technical difficulties in the fabrication and eliminates doping variation. Most reported GNRFETs are the SB type [17]. However, one drawback of SB-GNRFET is its ambipolar behavior that results in performance limitation, and SB-GNRFETs demonstrate a low I_{on}/I_{off} ratio in comparison with their MOS-type counterparts. Meanwhile, a relatively large I_{off} results in large power consumption in the OFF-state.

The ambipolar current conduction in SB-GNRFETs is due to the parasitic tunneling current through the SB at the drain contact. This problem exists also in carbon nanotube-based FETs [76]. To suppress the parasitic tunneling current in SB carbon nanotube FETs, a double-gate structure has been proposed [76, 77, 78]. In [76], the first gate controls carrier injection from the source contact, whereas the second gate makes the band-edge profile near the drain nearly flat. Therefore, the parasitic tunneling current is reduced and the ambipolar behavior is suppressed. An additional gate, however, poses some fabrication difficulties. In this chapter, we propose a SB-GNRFET with a single asymmetric gate (AG) and show that this avoids parasitic carrier injection at the drain and the device characteristics are im-

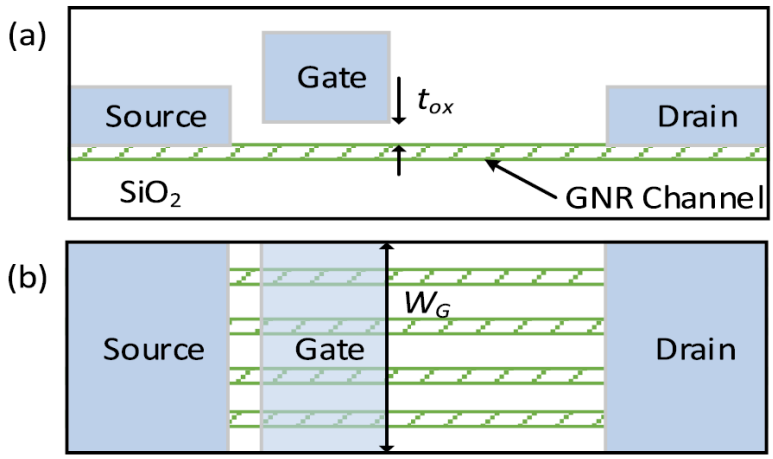

Figure 5.1: (a) Structure of an AG SB-GNRFET device. (b) Structure of a four-ribbon SB-GNRFET.

proved. A semi-analytical model for this structure is derived, implemented in Simulation Program with Integrated Circuit Emphasis (SPICE), and is applied to evaluate the circuit-level performance. Our results indicate that the AG device outperforms the symmetric gate (SG) structure.

5.2 Device Structure and Modeling

The proposed AG device is shown in Figure 5.1 (a) (cross-sectional view) and Figure 5.1 (b) (top view). In an SG device, where the gate covers the whole channel, as the voltage difference between the gate and drain increases, the Schottky barrier at the drain contact gets thinner, and as a result, the tunneling current increases, as shown in Figure 5.2. For AG structure, the thickness of the SB at the drain contact is only weakly affected by the gate voltage, and as a result, the tunneling current is significantly smaller than that of a SG structure. In the double-gate structure proposed in [76], the parasitic tunneling current can be completely suppressed, whereas for the AG structure, the parasitic current is not completely suppressed, as shown in Figure 5.3. However, in comparison with the double-gate structure proposed in [76], the AG structure can be more easily fabricated while still having an acceptable performance.

The proposed structure is similar to conventional MOSFETs, except that the gate only partially covers the channel. Therefore, its scaling is similar

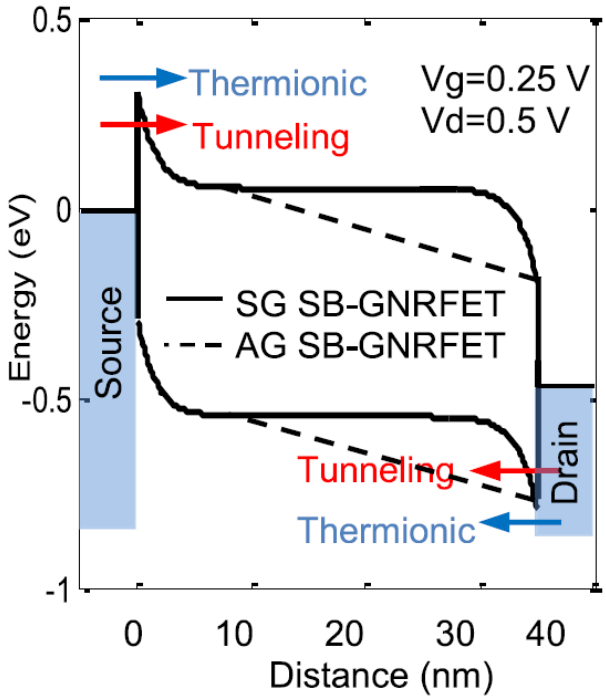

Figure 5.2: Band-edge profile along the channel of the AG and SG devices.

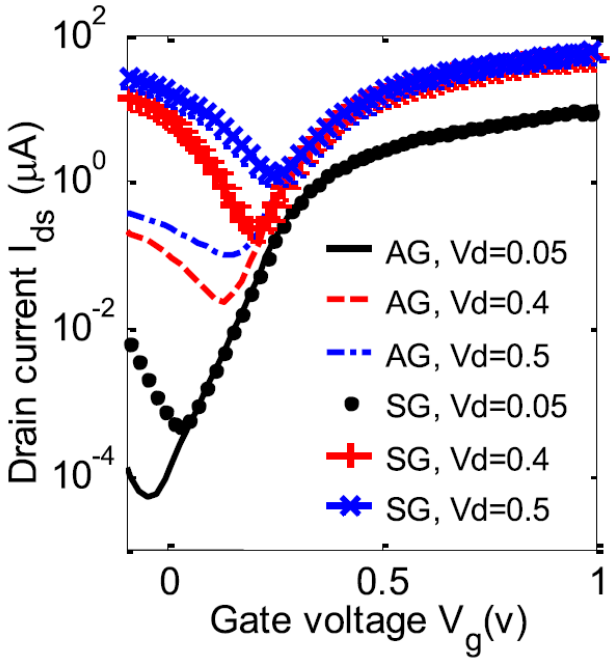

Figure 5.3: Transfer characteristics of the AG and SG devices. Device simulations are performed by employing an atomistic tight-binding model for the electron band structure along with the non-equilibrium Green's function formalism for the electronic transport [77].

to that of MOSFETs. Furthermore, this structure has the advantage that it can be scaled on the basis of the number of ribbons in each transistor [5].

We proposed a semi-analytical model for the current-voltage (I-V) characteristics of SB-GNRFETs, which allows complete and thorough exploration and evaluation of SB-GNRFET circuits [5]. This is a physics-based semi-analytical model for the I-V characteristics of SB-GNRFETs. We carry out accurate approximations of SB tunneling, channel charge, and current, which provide improved accuracy while maintaining compactness. The proposed model considers various design parameters and process variation effects, including GNR-specific line edge roughness, which allows complete and thorough exploration and evaluation of SB-GNRFET circuits.

The tunneling through the Schottky barrier of the device with the oxide thickness T_{ox} is computed using the Wentzel-Kramers-Brillouin (WKB) approximation based on the barrier profile

$$E_{SB}(z) = A_s e^{\frac{-\pi z}{2T_{ox}}} \tag{5.1}$$

with $A_s = q\varphi_{ch}$, the classical turning points $z_1 = 0$ and $z_2 = -2T_{ox}\ln(E/A_s))/\pi$, and the wavevector

$$k_z(E) \simeq \sqrt{\frac{M_\alpha}{\hbar^2\varepsilon_\alpha}}\sqrt{E_\alpha^2 - \varepsilon_\alpha^2} \tag{5.2}$$

which is obtained by the second-order expansion of the GNR E - k dispersion relationship

$$E_\alpha(k) = \pm t\sqrt{(1 + 4A_\alpha \cos\frac{\sqrt{3}a_1 k}{2}) + 4A_\alpha^2} \tag{5.3}$$

The resulting transmission coefficient is obtained as

$$T(E) = \exp\left\{-8T_{ox}\sqrt{\frac{M_\alpha}{h^2\varepsilon_\alpha}}\left[(E+\varepsilon_\alpha)(\frac{\pi}{2} - \arctan\frac{\varphi_\alpha}{\gamma_1}) \right.\right. \tag{5.4}$$

$$\left.\left. +\gamma_1 + \gamma_2\left(\arctan\left(\frac{\gamma_1\gamma_2}{A_s(E+\varepsilon_\alpha) - E(E+2\varepsilon_\alpha)}\right)\right)\right]\right\} \tag{5.5}$$

where h is the Planck's constant, $E = E_\alpha(k) - \varepsilon_\alpha$ is the energy with respect to the band edge energy $\varepsilon_\alpha = E_\alpha(0)$, M_α is the effective mass, and

$$\gamma_1 = \sqrt{\varepsilon_\alpha^2 - \varphi_\alpha^2} \tag{5.6}$$

$$\gamma_2 = \sqrt{(\varphi_\alpha + A_s)^2 - \varepsilon_\alpha^2} \tag{5.7}$$

$$\varphi_\alpha = \epsilon_\alpha + E - A_s \tag{5.8}$$

$$\theta_0 = \begin{cases} \pi & E(2\varepsilon_\alpha + E) < A_s(\varepsilon + E) \\ 0 & \text{otherwise} \end{cases} \tag{5.9}$$

In the case of φ_{ch} greater than the $E_g = 2\varepsilon_\alpha$, the spatial band diagram curvature becomes high enough to trigger band-to-band tunneling. In this case, a carrier with energy $0 < E < A_s - 2\varepsilon_\alpha$ experiences an Schottky barrier of a height $As = E + 2\varepsilon_\alpha$.

The effect of the asymmetric gate is considered in the hole's tunneling through the Schottky barrier at the drain side. Tunneling through this barrier is proportional to the drain voltage (V_d). Lower drain voltage results in a flatter band diagram at the drain contact, which in turn reduces the tunneling current, as shown in Figure 5.2. We derived an empirical equation for the effective tunneling coefficient of the holes at drain contact as

$$T_{eff}(E) = T(E) \cdot V_d^3/70 \tag{5.10}$$

The equivalent circuit of the GNRFET, which is shown in Figure 5.4 (a), consists of channel current source I_{ds}, parasitic capacitors $C_{ch,d}$, $C_{ch,s}$, $C_{g,ch}$, and $C_{sub,ch}$, and the voltage-controlled voltage source V_{ch} representing the channel voltage V_{ch}. The capacitors C_{gd} and C_{gs} are modeled using FastCap. Intrinsic capacitor $C_{ch,d(s)} = \partial C_{ch}/\partial V_{d(s)}$ is implemented in SPICE as voltage-controlled capacitor by defining the charge equations. The total channel charge Q_{ch} of carriers subband α can be expressed as

$$Q(\varphi_{ch}) = q \int D \cdot G_Q dE \tag{5.11}$$

where q is the electron charge, $D = (2/\pi\hbar)/(M_\alpha/2E)^{1/2}$ is the density of states, and G_Q is defined as

$$G_Q(E) = \frac{T_s(2 - T_d)f(E_{\alpha,s}) + T_d(2 - T_s)f(E_{\alpha,d}}{T_s + T_d + T_s T_d} \tag{5.12}$$

in which

$$f(x) = \frac{1}{1 + \exp(x/k_B T)} \tag{5.13}$$

is the Fermi-Dirac distribution function with the Boltzmann constant k_B and the temperature T. The term G_Q should be computed for both electrons and holes with

$$E^e_{s(d)} = E + \varepsilon_\alpha - q\varphi_{ch} + V_{s(d)} \tag{5.14}$$

$$E^h_{s(d)} = E - \varepsilon_\alpha + q\varphi_{ch} - V_{s(d)} \tag{5.15}$$

In our model, all subbands are considered and different Schottky barriers

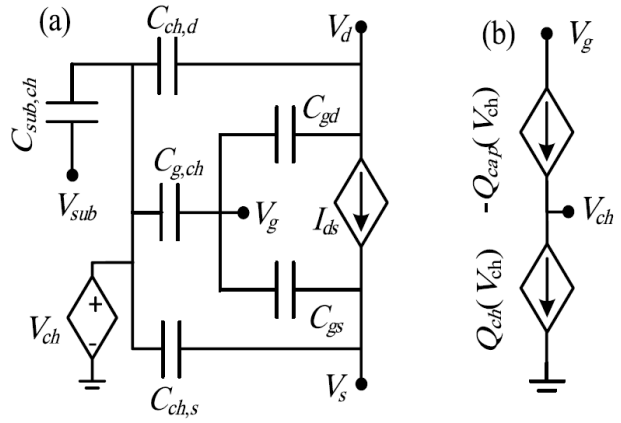

Figure 5.4: (a) Schematic of the developed compact circuit model of a SB-GNRFET. (b) SPICE setup for self-consistent solution of the channel potential V_{ch}.

exist for each subband. The first subband, however, contributes the most to the total current and the contribution of higher subbands exponentially decreases as the energy of the subband increases, as shown in the inset of Figure 5.5.

We analytically calculated GQ by piecewise linear approximation defined by four values E_{c1}, $E_{c2,Q}$, $G_{Q,0} = G_Q(0)$, and $G_{Q,1} = G_Q(E_{c1})$. The local maximum point of G_Q, the energy E_{c1}, is obtained as

$$E_{c1} = k_B \cdot T[W(\exp(E_f/k_BT - 1) + 1)] \tag{5.16}$$

where $W(\cdot)$ is the Lambert W function that is approximated as constant, parabolic, or linear functions according to the typical ranges of T and $E_f = \varepsilon_\alpha + q\varphi_{ch}V_s$ as

$$E_{c1}(E_f, T) = \begin{cases} k_BT & E_f < -0.05 \\ p_1E_f^2 + p_2E_f + p_3 & -0.05 < E_f < 0.145 \\ p_4E_f + p_5 & E_f > 0.145 \end{cases} \tag{5.17}$$

where $p_i, i = 1, 2, ..., 5$, are temperature-dependent coefficient as $p_i = \eta_{i,1} \cdot T + \eta_{i,2}$. Values of $\eta_{i,1}$ and $\eta_{i,2}$ obtained by curve fitting are given in Table 5.1.

The term $E_{c2,Q}$ is approximated using

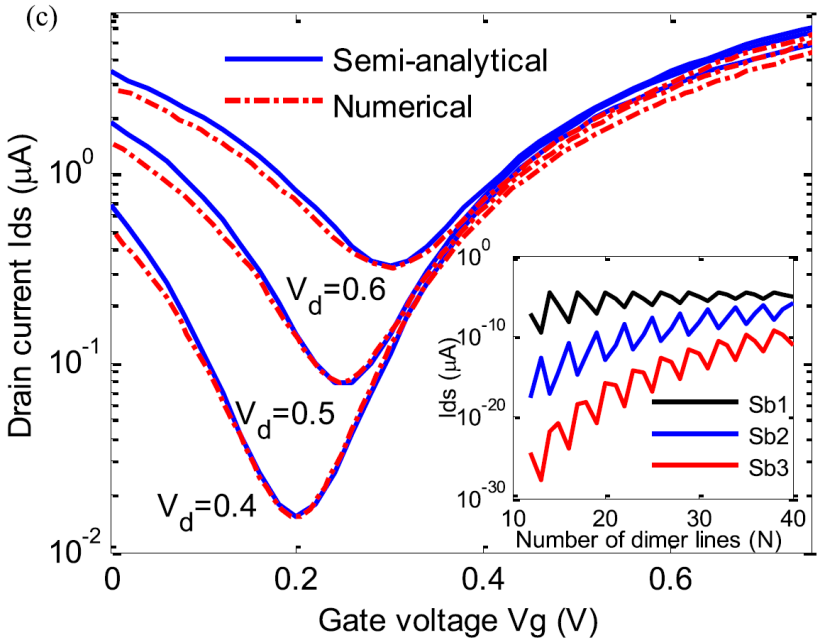

Figure 5.5: Comparison with atomistic device simulation ($N = 12$, $T_{ox} = 2$ nm) [5]. Inset: contribution of three lowest subbands to the current of GNRFETs with different widths.

Table 5.1: Values of Temperature-Dependent Coefficients

i	1	2	3	4	5
$\eta_{i,1}$	-0.0041	-1.33e-4	1.016e-4	-4.47e-4	6.29e-5
$\eta_{i,2}$	3.4092	0.2827	0.0035	1.0082	-0.0315

$$f(E_{c2,Q} - E_f) \simeq G_{Q,1}/10 \tag{5.18}$$

The integration introduced by Q_α can be therefore analytically computed as

$$Q = \frac{4q\sqrt{2M_\alpha}}{3\pi\hbar}[\sqrt{E_{c1}} \cdot G_{Q,0} + G_{Q,1} \cdot \frac{E_{c2,Q}}{\sqrt{E_{c1}} + \sqrt{E_{c2,Q}}}] \tag{5.19}$$

The same method can be used to compute the hole's charge by using $E_f = (\varepsilon_\alpha + q\varphi_{ch}V_d)$. The total mobile charge $Q_{ch} = \sum_\alpha (Q^h - Q^e)$ must be equal to the charge Q_{cap} across the gate, source, and drain capacitors that couple with the channel and are modeled empirically from data extracted from FastCap. Equating Q_{ch} and Q_{cap} yields a solution of φ_{ch}, which can be obtained using the equation solver circuit as shown in Figure 5.4 (b) [29, 30]. Given φ_{ch}, the current through the channel is computed using the Landauer-Büttiker formalism

$$I = \frac{q}{\pi\hbar}\int G_I(E)dE \tag{5.20}$$

$$G_I(E) = T_I[f(E - E_{\alpha,s}) - f(E - E_{\alpha,d})] \tag{5.21}$$

$$T_I = \frac{TsTd}{T_s + T_d - T_sT_d} \tag{5.22}$$

which can be analytically approximated using the same method for channel charge as

$$I = \frac{q}{2\pi\hbar}[G_{I,0} \cdot E_{c1} + G_{I,1} \cdot E_{c2,I}] \tag{5.23}$$

The absence of some atoms at the edges of GNR can significantly affect

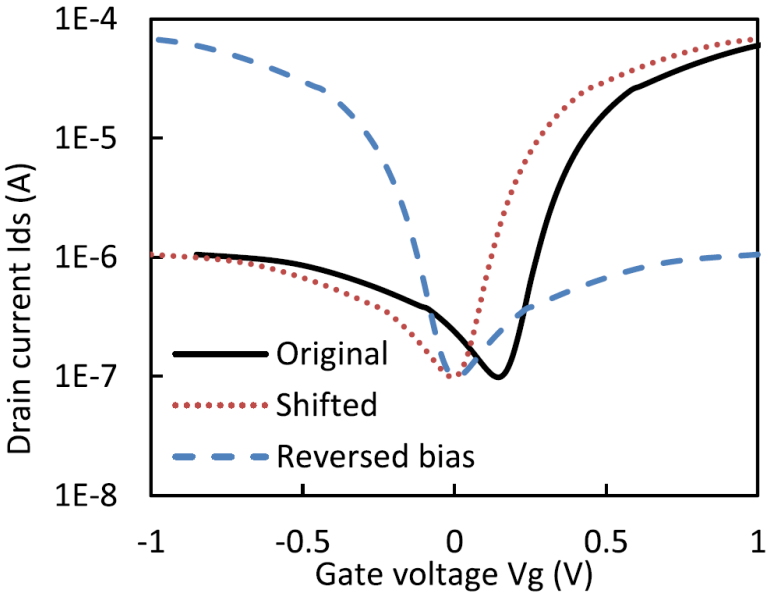

Figure 5.6: Transfer characteristics of the AG SB-GNRFET device with proper shifting and reversed bias.

electronic transport in GNRs, [3, 77, 5, 1]. The degree of roughness of the edges of a GNR is considered in our model through the line edge roughness probability p_r, which is the probability that any atom at the edges of a GNR is removed [1]. The effects of line edge roughness are modeled as $I_{rough} = A \cdot I_{DS}(\varepsilon_{\alpha,eff})$, where the scattering coefficient A, and the effective subband $\varepsilon_{\alpha,eff}$), are empirically obtained [1]. The accuracy of the developed compact model is verified with the atomistic non-equilibrium Green's function device simulator NanoTCAD ViDES [35], as shown in Figure 5.3.

We consider ambipolar devices, where the metal Fermi level is located in the middle of the GNR band gap at each contact. The minimum current in SB-GNRFETs occurs at the so-called ambipolar conduction point [5]. Ideally, the minimum current should occur in the OFF-state when $V_{GS} = 0$. The minimum current point, however, can be shifted to a different V_{GS} by tuning the gate work function by using various gate materials [26, 27, 69]. The AG SB-GNRFET can operate as a p-type device just by reversing the polarity of the applied voltages, which suppress the electron's parasitic current [76, 78]. An example of AG SB-GNRFETs working under $V_{DD} = 0.5$ V with ideal amount of 0.15 V shifting is shown in Figure 5.6.

5.3 Simulation Results and Performance Assessment

For comparison purpose, we used the 16-nm high-performance (HP) CMOS from predictive technology models (PTM). Minimum CMOS transistor dimension is chosen as (W/L) = (32 nm/16 nm). The transistor dimensions of the GNRFETs are scaled to match the PTM libraries. We choose an SB-GNRFET device with $T_{ox} = 1$ nm, and six ribbons in the channel each with $N = 12$ dimer lines (with the band gap $E_g \simeq 0.6$ eV) for the rest of this chapter, as shown in Figure 5.3 [3, 5, 1]. Both the ideal cases ($p_r = 0$) and the non-ideal cases with $p_r = 0.1$ are investigated. The supply voltage for 16-nm CMOS and SB-GNRFET are 0.7 and 0.5 V, respectively.

Figure 5.7 shows the comparison of the transfer characteristics of SG and AG SB-GNRFETs for the ideal (GNR edges are smooth) and non-ideal cases (GNR edges are rough). I_{on} of the AG is nearly the same as the SG device, whereas the I_{off} of AG is considerably smaller ($\sim 11\times$) than that of the SG device. As a result, the AG SB-GNRFET in the ideal case shows about a $10\times$ improvement in the I_{on}/I_{off} ratio; however, as shown in Figure 5.8, the I_{on}/I_{off} ratio improvement is smaller ($\sim 5\times$) in the presence of GNR line edge roughness. Figure 5.9 shows that by employing the AG device, the subthreshold swing is improved by at least 40%. Table 5.2 shows the subthreshold swing S and I_{on}/I_{off} ratio of each device under, respectively, chosen V_{DD}. It is shown that ideal AG SB-GNRFETs have the lower subthreshold swing and higher I_{on}/I_{off} ratio than SG SB-GNRFETs. However, the AG SB-GNRFET device has a still lower I_{on}/I_{off} ratio than that of the CMOS because of the large Vd and thin Schottky barrier tunneling distance due to the scaled oxide thickness.

Figure 5.10 shows the effect of T_{ox}. A higher T_{ox} implies a smaller tunneling probability through the Schottky barrier, which results in the lower current. However, the I_{on}/I_{off} ratio increases with the increase in oxide thickness. Furthermore, AG device has much higher ratio than SG. Figure 5.11 shows the effect of number of dimer lines N, which tracks the periodic effect on band gaps [19]. For $N = 3q + 2, (q \in N)$, the band gap is very small, resulting in a low I_{on}/I_{off} ratio. For $N = 3q$, there is a moderate band gap, which results in a high I_{on}/I_{off} ratio and a high I_{on}. For $N = 3q + 1$, the band gap is the largest, which results in the highest I_{on}/I_{off} ratio. Also note that the I_{on}/I_{off} ratio tends to increase as N decreases.

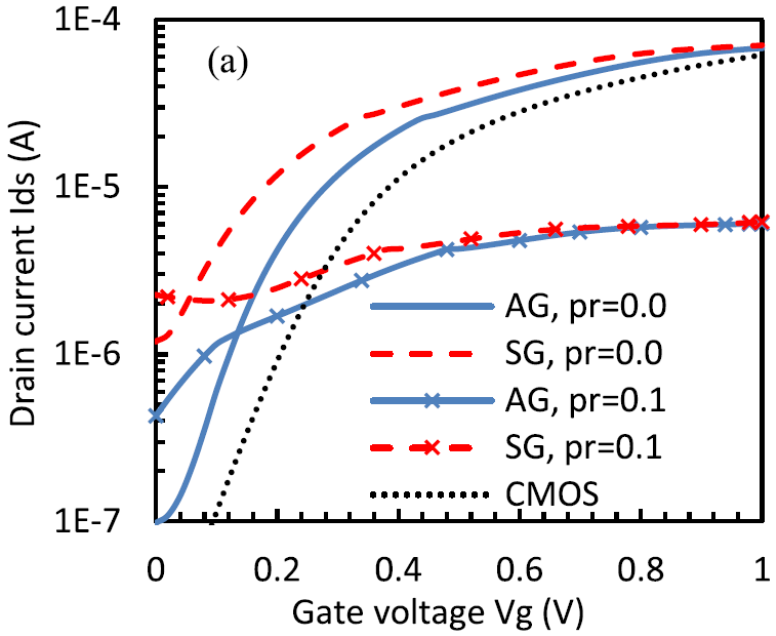

Figure 5.7: Comparison between the transfer characteristics of the AG and SG devices.

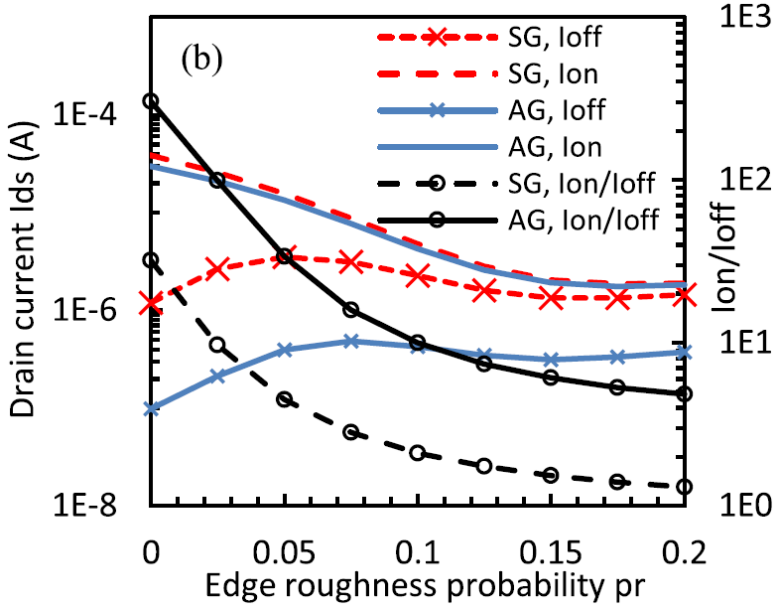

Figure 5.8: I_{on} and I_{off} of AG and SG devices and I_{on}/I_{off} ratio as functions of p_r.

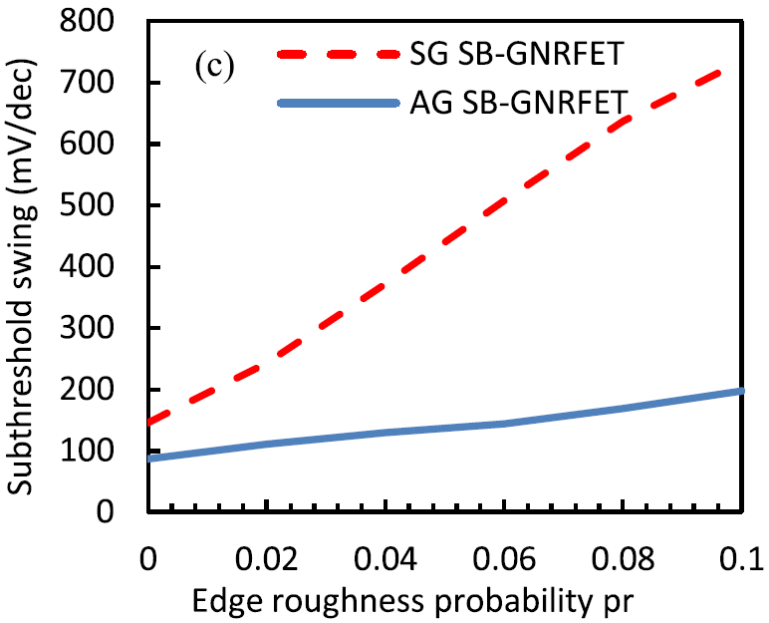

Figure 5.9: Subthreshold swing of the AG and SG devices as functions of p_r.

Table 5.2: Transistor Properties

Device	p_r	S (mV/dec)	I_{on}/I_{off}	V_{DD} (V)
Si-CMOS (HP)	–	93.46	3.49E+03	0.7
SG-GNRFET	0	145.14	3.21E+01	0.5
	0.1	735.29	2.11E+00	0.5
AG SB-GNRFET	0	86.96	3.04E+02	0.5
	0.1	197.24	9.98E+00	0.5

Subthreshold swing and I_{on}/I_{off} ratio of each device. For GNRFETs, devices of different line edge roughness (p_r) are listed as well.

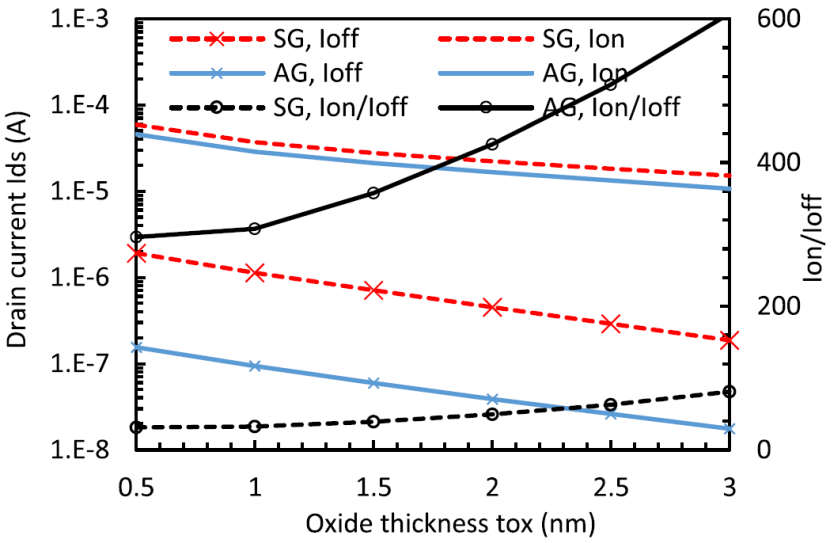

Figure 5.10: I_{on}, I_{off}, and I_{on}/I_{off} vs. oxide thickness T_{ox}.

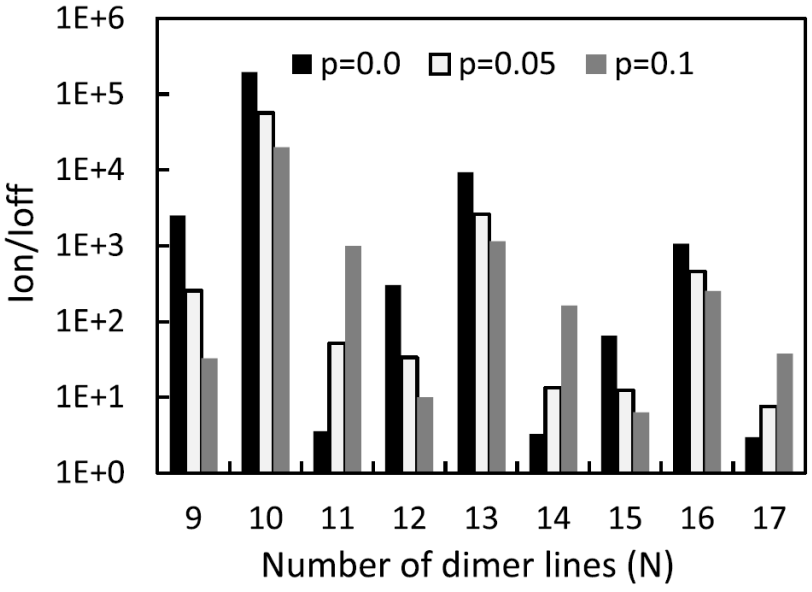

Figure 5.11: I_{on}/I_{off} vs. number of dimer lines N of AG SB-GNRFET.

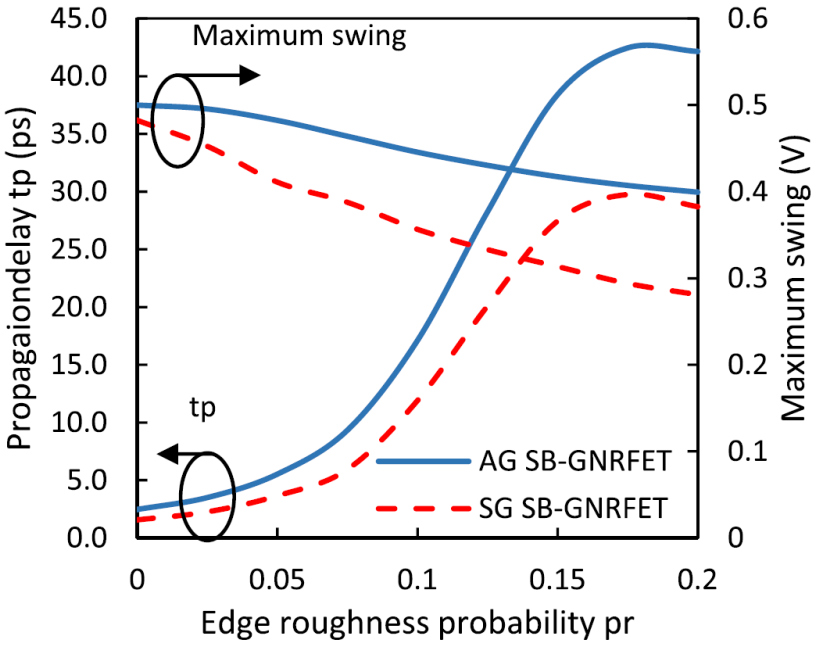

Figure 5.12: Effect of line edge roughness on inverter propagation delay, t_p, and maximum output swing of an inverter.

We analyze properties of an inverter built with SG and AG SB-GNRFETs under $V_{DD} = 0.5$ V. We used our SPICE model to perform DC and transient analysis of the inverter. As shown in Figure 5.12, high line edge roughness probability p_r results in a higher propagation delay due to the smaller transistor current. Both low and high output voltage levels of the inverter degrade with line edge roughness, which results in the lower maximum swing. The AG-based inverter has better properties than the SG inverter.

Figure 5.13 shows the voltage transfer curves of an inverter built with AG and SG SB-GNRFETs with different line edge roughness settings. Vin and Vout are the input and output voltages of the inverter, respectively. High line edge roughness probability pr results in a lower voltage swing. Figure 5.14 shows the voltage transfer curves of the inverters in different technologies which are normalized to the corresponding V_{DD}. Figure 5.15 shows the normalized noise margin of different inverters. AG-based inverter has better noise margin than the SG. The line edge roughness significantly reduces the noise margin of SG-based inverter, while it has small effect on AG.

To evaluate the AG SB-GNRFET performance on the circuit level, basic gates (*inv*, *nand2*, *nand3*, *nor2*, *xor2*, *nor3*, *nand4*), and benchmark circuits

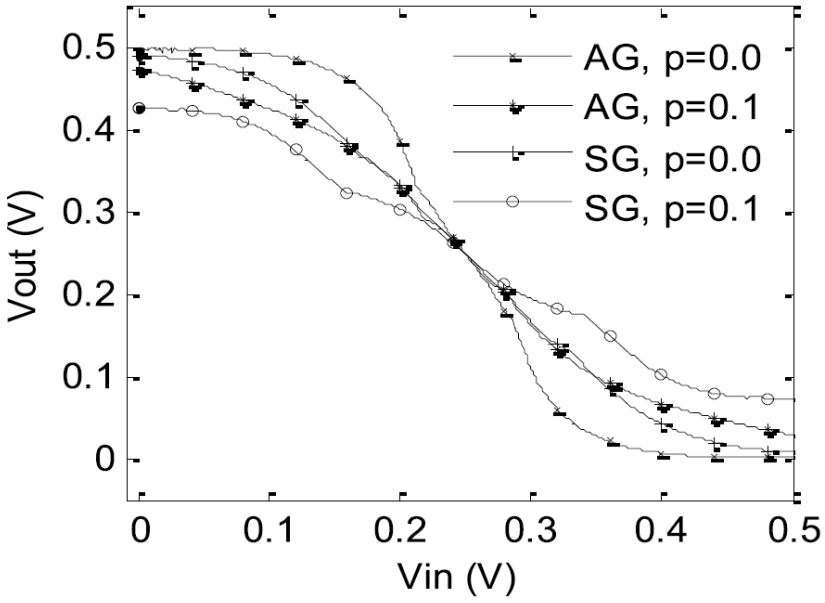

Figure 5.13: Effect of p_r on AG- and SG-based inverter DC characteristics.

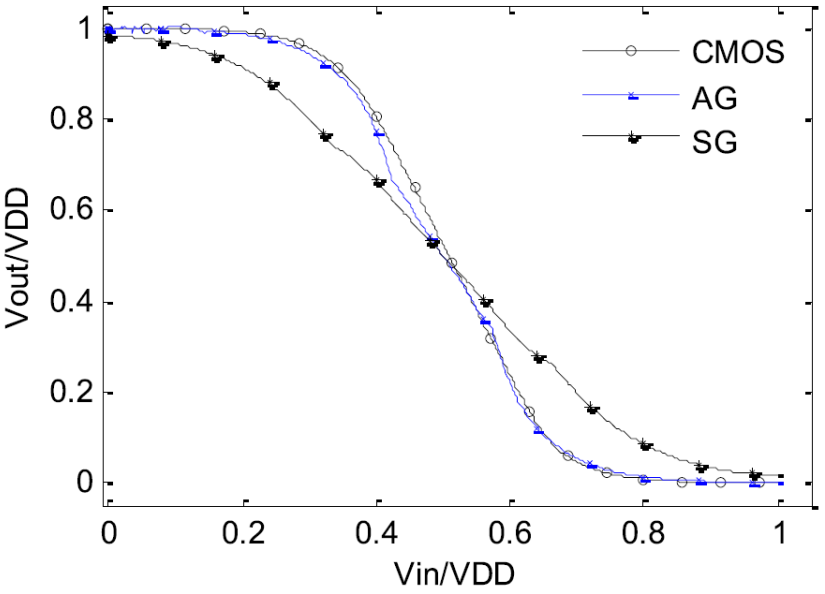

Figure 5.14: Comparison of inverter DC characteristics in different technologies. Voltages are normalized to V_{DD} in each technology.

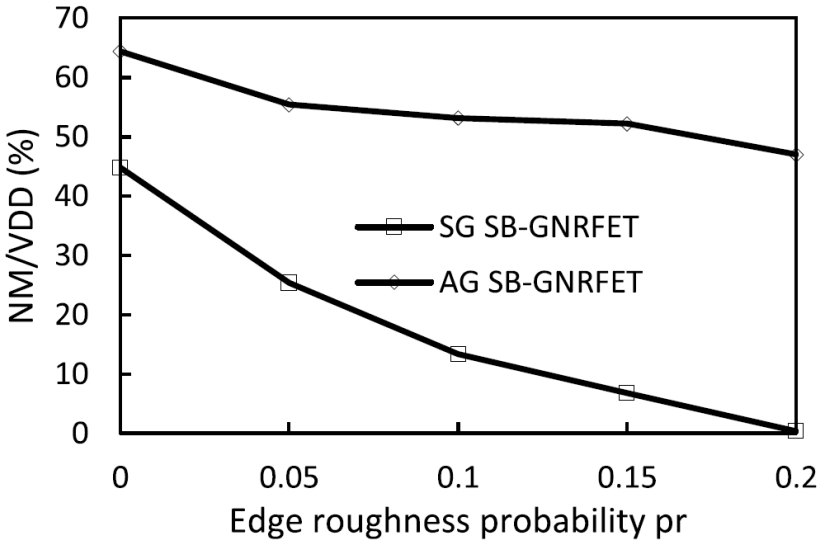

Figure 5.15: Normalized noise margin of AG- and SG-based inverters.

Table 5.3: Simulation Results of Basic Logic Gates

	Delay (ps)				Dynamic power (μW)				Leakage power (μW)				EDP (J.s)			
	SG SB-GNRFET		AG SB-GNRFET		SG SB-GNRFET		AG SB-GNRFET		SG SB-GNRFET		AG SB-GNRFET		SG SB-GNRFET		AG SB-GNRFET	
Circuit	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1
inv1	1.00	9.00	3.30	17.00	3.47	0.82	0.13	0.64	0.52	0.69	0.10	0.35	3.47E-30	6.64E-29	1.46E-30	1.85E-28
nand2	3.00	12.00	5.20	20.30	3.28	1.06	0.15	0.72	0.70	0.79	0.17	0.43	2.95E-29	1.52E-28	3.98E-30	2.97E-28
nand3	4.00	15.00	6.00	23.00	1.54	0.87	0.10	0.56	0.67	0.75	0.20	0.39	2.47E-29	1.96E-28	3.70E-30	2.95E-28
nor2	3.00	11.00	5.00	22.00	2.26	1.15	0.15	0.28	0.67	0.75	0.13	0.42	2.03E-29	1.39E-28	3.83E-30	1.37E-28
xor2	6.00	17.00	8.00	23.00	5.01	2.50	0.79	1.02	3.37	4.07	0.64	2.38	1.80E-28	7.23E-28	5.05E-29	5.40E-28
nor3	4.00	13.00	7.00	29.00	1.28	1.25	0.37	0.37	0.72	0.80	0.42	0.42	2.05E-29	2.11E-28	1.81E-29	3.10E-28
nand4	4.00	19.00	6.00	25.00	1.22	1.04	0.10	0.39	0.65	0.76	0.26	0.36	1.96E-29	3.75E-28	3.47E-30	2.45E-28
Average	3.57	13.71	5.79	22.76	2.58	1.24	0.26	0.57	1.04	1.23	0.27	0.68	4.26E-29	2.66E-28	1.21E-29	2.87E-28

are studied, including circuits *c17* from ISCAS '85, *b02* from ITC '99, *s27* from ISCAS '89, and carry generator for the third bit of a carry look-ahead adder (*cla*). Simulation results of basic gates and benchmark circuits are presented in Tables 5.3 and 5.4, respectively. The reported values are the maximum delay, dynamic power, leakage power, and energy-delay product (EDP) values.

Figure 5.16 shows the performance of each technology node in terms of maximum delay, dynamic power, leakage power, total power, and EDP values. An ideal AG SB-GNRFET has the best EDP, but its leakage power is higher than that of Si-CMOS. The non-ideal AG SB-GNRFET has the worst delay because of its relatively lower I_{on}. Because of a lower I_{off} of AG SB-GNRFET, the circuits, either ideal or non-ideal, have lower power than that of the SG SB-GNRFET (e.g., 73% or 44% lower leakage power

Table 5.4: Simulation Results of Benchmark Circuits

	Delay (ps)				Dynamic power (μW)				Leakage power (μW)				EDP (J.s)			
	SG SB-GNRFET		AG SB-GNRFET		SG SB-GNRFET		AG SB-GNRFET		SG SB-GNRFET		AG SB-GNRFET		SG SB-GNRFET		AG SB-GNRFET	
Circuit	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1	pr=0.0	pr=0.1
c17	10.00	62.00	10.00	70.00	5.57	6.74	1.13	4.24	4.65	5.77	0.91	3.34	5.57E-28	2.59E-26	1.13E-28	2.08E-26
b02	10.00	72.00	14.30	116.00	53.99	60.10	5.32	20.37	21.99	28.95	4.09	16.37	5.40E-27	3.12E-25	1.09E-27	2.74E-25
s27	12.20	92.00	17.00	114.00	11.78	13.19	1.98	7.95	8.95	12.61	1.65	6.98	1.75E-27	1.12E-25	5.72E-28	1.03E-25
cla	7.00	59.00	8.00	49.00	9.97	4.62	0.83	3.76	3.80	4.56	0.82	2.51	4.88E-28	1.61E-26	5.34E-29	9.02E-27
Average	9.80	71.25	12.32	87.25	20.33	21.16	2.31	9.08	9.85	12.97	1.87	7.30	2.05E-27	1.16E-25	4.56E-28	1.02E-25

for basic gates). Non-ideal SB-GNRFET (both AG and SG) consumes more leakage power than the ideal one because its I_{off} is increased in the presence of line edge roughness. This also reduces the I_{on}/I_{off} ratio and results in the degradation of the delay and EDP.

5.4 Conclusion

To improve the performance of SB-GNRFETs, the ambipolar behavior of these devices should be suppressed. We propose a device with an AG which covers only some part of the channel close to the source contact. The newly proposed design effectively suppresses the ambipolarity and reduces the I_{off} by 11×. A SPICE-compatible compact model of the proposed device is developed. Simulation results show significant improvement in device and circuit characteristics, which render AG SB-GNRFET as a potential candidate for the next-generation high-performance/low-power device. However, advanced fabrication techniques are required to remove the non-idealities faced by GNR fabrication.

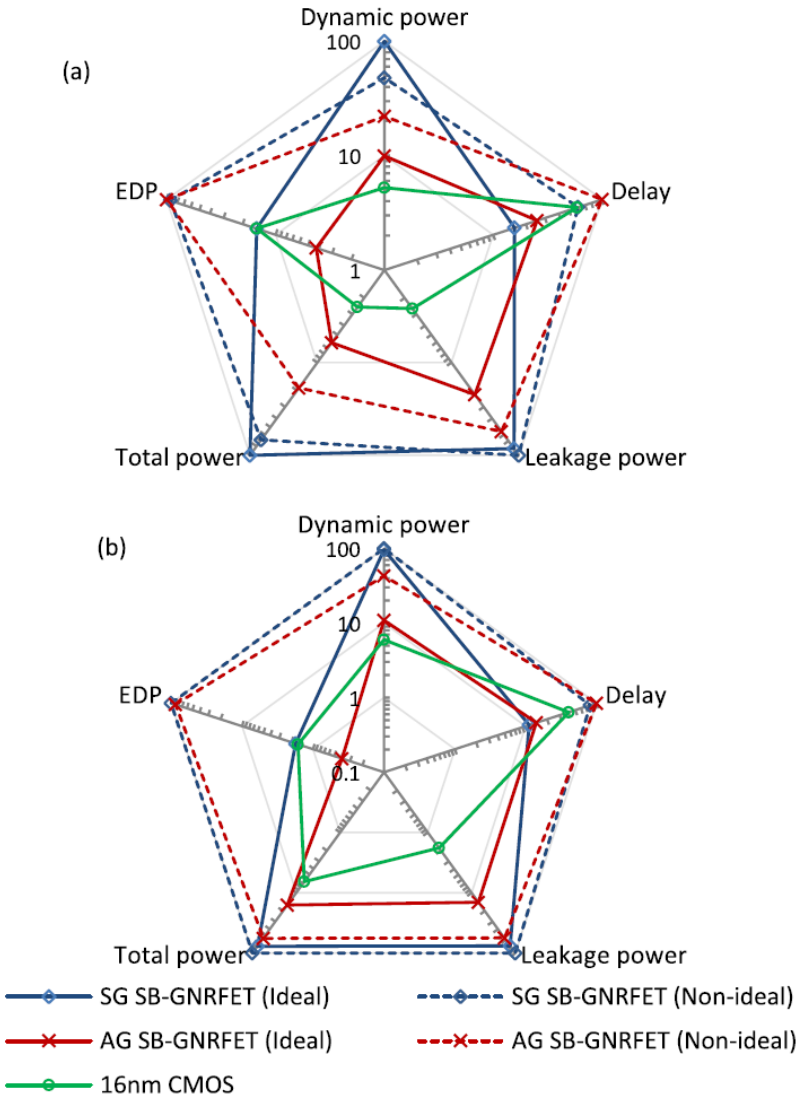

Figure 5.16: Comparison of different technologies based on five figures of merits: delay, dynamic power, leakage power, total power, and EDP. The number for each figure of merit is the average value that is normalized to the maximum value of that category and presented in percentage. Each axis has a maximum value of 100%. (a) Basic logic gates. (b) Benchmark circuits.

CHAPTER 6

TMDFET MODELING AND SIMULATION

6.1 Introduction

Conventional silicon-based CMOS transistor scaling has become increasingly difficult due to increased wire resistivity, significant mobility degradation, and large dopant fluctuations. Researchers began to turn to various emerging materials in order to keep up with Moore's law. In particular, two-dimensional (2-D) materials such as graphene and transition metal dichalcogenides (TMDs) with a chemical formula of MX_2 (e.g. MoS_2, WSe_2, etc.) have drawn a great deal of attention because of their planar and robust honeycomb lattice structure and outstanding properties [79]. Graphene in its pristine 2-D form is semi-metallic and has no band gap. While band gap can be opened by narrowing graphene into graphene nanoribbons under 10 nm, issues such as process variation, mobility degradation, and line edge roughness severely degrade its performance [80, 71, 3, 1, 5, 7, 9]. TMDs, on the other hand, have a band gap of 1-2 eV by nature without scaling down to the small nanometer range, which greatly reduces the difficulty of production and also mitigates undesirable effects from process variation. As a result, TMD field-effect transistors (FETs) have been regarded as a promising transistor design in the post-CMOS era [79, 81, 82, 83].

Moreover, the ultra-thin, planar structure of 2-D materials opens up the possibility of flexible electronics, which can be an important component in wearable technology, electronic paper, and sensors. Successful manufacturing of flexible 2-D transistors have been demonstrated in [84, 85, 86, 87, 88]. Measurement data collected from these transistors with different extent of bending are also reported. In these examples, TMDs are used as the channel material, while the substrate, gate oxide, and electrodes of the transistor are made of various different materials. These successful demonstrations show

the great potential of producing flexible electronics with TMDFETs.

While TMDFETs show great potential, fabrication techniques are not yet mature. Most fabricated TMDFETs that are reported, whether flexible or not, are in the micrometer or even millimeter scale [84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94], far behind today's sub-45-nm technology nodes. Besides, only few instances of fabricated TMDFET circuits consist of more than one transistor [93, 94]. As a result, it is unclear at the moment how well TMDFETs perform on the circuit level, especially when they are sized down to match contemporary transistor sizes. Also, it will be interesting to learn how flexible circuits perform under different bending settings, which is our main focus in this work. Before such futuristic flexible TMDFET circuits can be manufactured, simulation plays an important role in evaluating the emerging technology. In fact, there is abundant work in the theoretical and computational studies of transistor-level properties of nanoscale TMDFETs based on non-equilibrium Green's function (NEGF) formalism and/or Schrödinger-Poisson solvers at the cost of very high computational complexity [81, 82, 83, 95, 96, 97], in which detailed transistor-level transfer characteristics is reported. However, it is difficult to scale to circuit-level simulations with these approaches due to the high computation time. In [98], a physics compact model of TMDFET is proposed, which simplifies the computation down to relatively simple equations that can be computed more efficiently. With a compact model, circuit-level simulations become feasible. However, the focus of the work was the model derivation rather than simulation, and there has been no circuit-level simulation of TMDFETs to date to the best of our knowledge.

In this work, we further simplify the compact TMDFET model proposed in [98] by deriving more closed-form approximations, reducing the computational complexity such that it becomes SPICE-compatible. In addition, we explore and model the effects induced by bending in order to create a model for flexible TMDFETs. The SPICE model is parameterized in design parameters such as length, width, oxide thickness, channel material, and applied strain, which are what circuit designers are familiar with. We also modeled parasitics such that circuit simulations are more realistic. We implement our model in SPICE and perform circuit-level simulations to evaluate the performance of TMDFET-based circuits as a whole. This enables design space exploration and process variation evaluation on the circuit level.

To summarize, the main contributions of this chapter are as follows:

- Developing the first SPICE-compatible TMDFET model.
- Modeling and evaluating the effect of bending in terms of *applied strain*.
- Modeling and evaluating process variation in W and L.
- Evaluating flexible TMDFET circuits in terms of delay and power performance and scalability.
- Comparing TMDFET circuits with Si-CMOS circuits.

The rest of the chapter is organized as follows: Section 6.2 provides additional background on TMDFETs. Section 6.3 presents our SPICE-compatible model that supports flexibility. Section 6.4 presents the experimental results. Section 6.5 draws conclusions.

6.2 Background

6.2.1 TMDs and TMDFETs

Transition metal dichalcogenides (TMDs) have the chemical form of MX_2, where M is a transition metal such as Mo, W, Ti, and X is a chalcogen such as S, Se. A TMD monolayer is an atomically thin planar honeycomb lattice structure similar to that of graphene. The electrical properties of TMDs vary for different molecular compositions, crystal structures, and number of layers [79]. For example, the band gap decreases as the TMD thickness goes from monolayer, bilayer, to multilayer [99], and the mobility generally increases with the number of layers [100].

Table 6.1 shows the band gap E_g and electron/hole effective mass m^* of different monolayer MX_2. Here, m_0 is the electron mass. Monolayer TMDs with M = Mo or W have a direct band gap ranging from 1.10 eV to 1.93 eV, very suitable for making transistors for digital applications. Therefore, they are of very high interest in the research community. In particular, MoS_2 is the most studied TMD material to date [79], while WSe_2 starts to draw attention as there are examples of WSe_2 transistors that outperform their MoS_2 counterparts [92].

Table 6.1: Properties of monolayer MX_2 [96]

MX_2	Band Gap E_g [eV]	Effective Mass	
		Electron ($\frac{m_e^*}{m_0}$)	Hole ($\frac{m_h^*}{m_0}$)
MoS_2	1.80	0.56	0.64
$MoSe_2$	1.51	0.62	0.72
$MoTe_2$	1.10	0.64	0.78
WS_2	1.93	0.33	0.43
WSe_2	1.62	0.35	0.46

A few examples of fabricated TMDFETs are reviewed here. In the work of [89], a 4 um × 500 nm n-type MoS_2 transistor was fabricated, which achieved an ON/OFF current ratio I_{on}/I_{off} of 10^8, a subthreshold swing of 74 mV/dec, with mobility of 217 $cm^2V^{-1}s^{-1}$. In a more recent work [91], a MoS_2 transistor made from chemical vapor deposition (CVD) was presented. It achieved an I_{on}/I_{off} ratio of 10^4 over a voltage range of 2-5 V, and a subthreshold swing of $\sim$ 200 mV/dec. The mobility was no greater than 25 $cm^2V^{-1}s^{-1}$. Although the performance was worse than the transistor in [89], the yield was improved. In [92], a 1 um × 9.4 um p-type monolayer WSe_2 transistor was fabricated. It achieved an I_{on}/I_{off} ratio of 10^6 over a voltage range of less than 1 V and a subthreshold swing of 60 mV/dec. The mobility was 250 $cm^2V^{-1}s^{-1}$. To date, this is one of the best TMDFETs reported [79]. It was also used as the validation data set for the compact modeling work of [98].

6.2.2 TMDFETs under Bending

When discussing flexible transistors, two metrics are often used to describe the amount of bending: *bending radius* and *strain*, as illustrated in Figure 6.1. Here, R is bending radius, τ is half film thickness, and ϵ is applied strain. The relationship is formulated as $\epsilon = \tau/R$ [101]. Moreover, strain can be further classified as uniaxial or biaxial based on the direction of bending [102]. Bending the TMD lattice causes the M−M, M−X, and X−X bonds to change, and thus affecting the electrical properties [103].

In [86], a flexible transistor with a MoS_2 channel and graphene electrodes is fabricated and measured, and the mobility and threshold voltage are reported. At bending radii ranging from 2.2 to ∞, both mobility and threshold

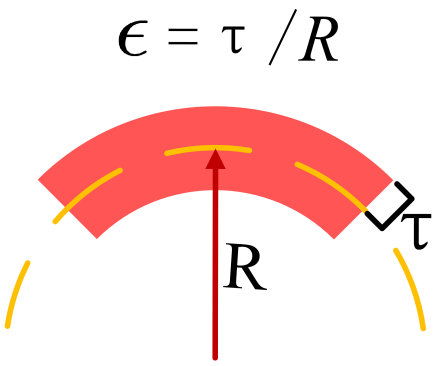

Figure 6.1: The relationship between bending radius R, half film thickness τ, and strain ϵ [101].

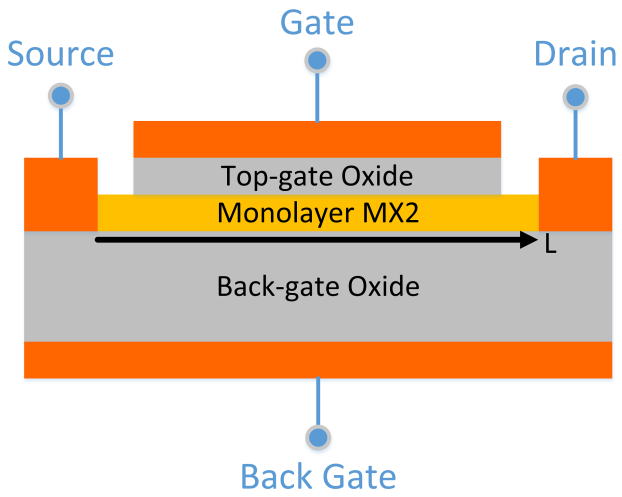

Figure 6.2: Cross section of a TMDFET, in which L is the channel length, $T_{ox,tg}$ ($T_{ox,bg}$) are the top (back) gate oxide thicknesses.

voltage showed little correlation to bending. In [102], the band gap of different MX_2 material is computed from first-principles density-functional-theory (DFT) based methodology. The band gap under a strain of 0 to 10% is reported: it generally decreases linearly with respect to strain, as further discussed in Section 6.3.2. In addition, the amount of change in band gap under uniaxial strain is the same regardless of the bending direction. In [103], a flexible MoS_2 transistor is fabricated, and the band gap is measured by observing the absorption and photoluminescence spectra. The band gap also shows a clear linear trend with respect to strain, consistent with the computational results in [102]. The range of the applied strain is from 0 to 0.52%. The work [101] also reported similar results.

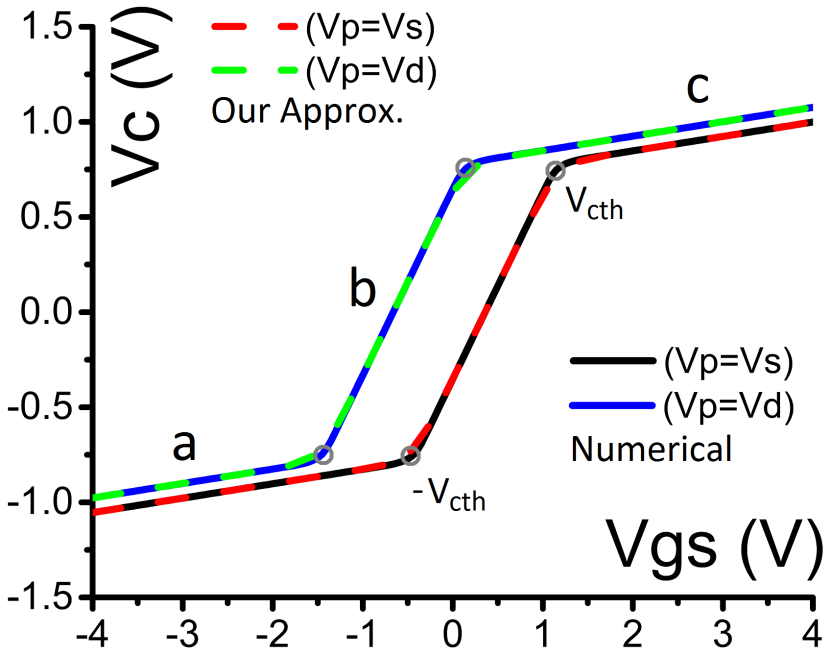

Figure 6.3: V_c vs. V_{gs} at different V_{ds}, divided into three regimes. V_p is from equations (6.1) and (6.5).

6.3 Flexible TMDFET Modeling

In this section, we will discuss the derivation of our compact model, modeling of flexibility, and the full transistor model implemented in SPICE.

6.3.1 SPICE-Compatible Current Modeling

To the best of our knowledge, the only existing TMD transistor compact modeling work is [98]. In [98], a physics-based analytical model of a generic double-gate monolayer TMDFET is developed. While most of the computational studies [81, 82, 83, 95, 96, 97] assume sub-100-nm channel length and ballistic transport, the work [98] adopts a classical drift-diffusion current model, which is more accurate when describing transistor sizes above 100 nm [104]. As most of the fabricated TMDFETs reported to date have sizes greater than 0.5 μm [84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94], the classical model is more suitable to describe the transfer characteristics in these transistors. Although the model in [98] is not fully SPICE-compatible, it provides a good basis that we can further simplify and build upon.

In the TMDFET modeled in [98] (as shown in Figure 6.2), the channel

is made of MX_2, and the source and drain electrodes are assumed to be ohmic contacts. The top and bottom gates that modulate the electrostatics in the channel are made of metal. The source is treated as a reference node. The channel length is defined as L, the gate width being W, and the top (bottom) gate oxide thickness is $T_{ox,tg}$ ($T_{ox,bg}$). In our work, we follow the same framework.

The drain current I_{ds} model in [98], based on the classical drift-diffusion theory, takes the following form:

$$I_{ds} = \mu \frac{W}{L} \int_{V_{cs}}^{V_{cd}} Q_c \frac{dV_p}{dV_c} dV_c = \mu \frac{W}{L} \{g(V_c)\}_{V_{cs}}^{V_{cd}} \tag{6.1}$$

where μ is the carrier mobility. V_c is the channel voltage. V_{cd} and V_{cs} are the channel voltages at the drain and source, respectively. Q_c is the channel charge. $V_p(x)$ is applied voltage at channel position x. $g(V_c)$ is a function defined in [98] as

$$g(V_c) = \begin{cases} g_1(V_c) & \texttt{if } |qV_c| < |E_0| \\ g_2(V_c) & \texttt{if } |qV_c| \geq |E_0| \end{cases} \tag{6.2}$$

$$g_1(V_c) = \left(1 + \frac{q^2 D_0}{C_t + C_b}\right) \left(\frac{q^2 D_0 V_c^2}{2} + qD_0(E_0 - kT)V_c\right) \tag{6.3}$$

$$g_2(V_c) = D_0 (kT)^2 e^{\frac{-qV_c - E_0}{kT}} \left(1 + \frac{q^2 D_0}{2(C_t + C_b)} e^{\frac{-(qV_c + E_0)}{kT}}\right) \tag{6.4}$$

The channel voltage V_c can be solved from the following equation [98]:

$$\begin{aligned} V_c(x) = & \frac{Q_c(V_c)}{C_t + C_b} + \frac{C_t}{C_t + C_b} (V_{gs} - V_{gs0} - V_p(x)) \\ & + \frac{C_b}{C_t + C_b} (V_{bs} - V_{bs0} - V_p(x)) \end{aligned} \tag{6.5}$$

where x is the position in the channel. $C_t = \epsilon_{tg}/T_{ox,tg}$ is the unit area top gate capacitance, where ϵ_{tg} is the permittivity of the top gate dielectric. C_b, the bottom gate capacitance can be computed similarly. V_{gs} is the top-gate-to-source voltage, V_{bs} is the bottom-gate-to-source voltage, V_{gs0} and V_{bs0} are flat-band voltages, and $V_p(x = 0) = V_s = 0$ and $V_p(x = L) = V_{ds}$, the drain-to-source voltage.

The channel charge Q_c can be computed from the density of states $D(E)$ and the Fermi-Dirac distribution $f(E)$:

$$\begin{aligned} Q_c = Q_p + Q_n =& q \int_{-\infty}^{0} D(E) f(E_F - E) dE \\ & - q \int_{0}^{\infty} D(E) f(E - E_F) dE \end{aligned} \tag{6.6}$$

where E is the energy, and Q_p and Q_n are the charge contributed from holes and electrons, respectively. In [98], the following approximations are adopted in order to obtain a closed-form Q_p and Q_n: $f(E) \sim 1$ for $E < E_F$ and $f(E) \sim \exp(\frac{E_F - E}{kT})$ for $E \geq E_F$, and $D(E) = D_0 \cdot H(E - E_0)$. Here, E_F is the Fermi level, k is the Boltzmann constant, T is temperature, $H(E)$ is the Heaviside function, $E_0 = E_g/2$ is the energy of the zeroth (most dominant) subband, and $D_0 = \frac{m^*}{\pi\hbar^2}$ where m^* is the effective mass and $\hbar$ is the reduced Planck's constant. The resulting Q_p and Q_n are:

$$Q_p = \begin{cases} -q^2 D_0 V_c - q D_0 (E_0 - kT) & \text{if } qV_c \leq -E_0 \\ q D_0 kT e^{\frac{-qV_c - E_0}{kT}} & \text{otherwise} \end{cases} \tag{6.7}$$

$$Q_n = \begin{cases} -q^2 D_0 V_c + q D_0 (E_0 - kT) & \text{if } qV_c \geq E_0 \\ -q D_0 kT e^{\frac{qV_c - E_0}{kT}} & \text{otherwise} \end{cases} \tag{6.8}$$

Unfortunately, even after plugging in the above simplified Q_p and Q_n, equation (6.5) does not have an analytical solution. The work [98] does not specifically provide a method to solve for V_c. It is assumed that V_c can be solved numerically. Figure 6.3 shows the numerical solution of V_c with respect to different V_{gs}. In SPICE, a solver subcircuit can be built to solve for V_c iteratively as the simulation goes, as demonstrated in SPICE models designed

for other emerging transistors [29, 30, 1, 31]. The solver approach, although often feasible, significantly increases the computation time. Therefore, we seek a closed-form approximation to V_c.

As $Q_c(V_c)$ is linear in V_c when $|qV_c| \geq |E_0|$, V_c has an exact solution in this region as follows:

$$V_c = \begin{cases} V_{c-} & \text{if } qV_c \leq -E_0 \\ V_{c+} & \text{if } qV_c \geq E_0 \end{cases} \tag{6.9}$$

$$V_{c-} = \frac{-qD_0(E_0 - kT) + C_t(V_g - V_{gs0} - V_p) + C_b(V_b - V_{bs0} - V_p)}{C_t + C_b + q^2 D_0} \tag{6.10}$$

$$V_{c+} = \frac{qD_0(E_0 - kT) + C_t(V_g - V_{gs0} - V_p) + C_b(V_b - V_{bs0} - V_p)}{C_t + C_b + q^2 D_0} \tag{6.11}$$

In the region where $|qV_c| \leq |E_0|$, $Q_c(V_c)$ is an exponential function in V_c, and thus V_c has no closed-form analytical solution. However, in the region where V_c is close to 0, V_c is almost linear (shown as segment b in Figure 6.3), which allows us to approximate V_c in this region as a linear function V_{c0}:

$$V_{c0} = \frac{2V_{cth}(V_{c-} + V_{cth})}{2V_{cth} + V_{c-} - V_{c+}} - V_{cth} \tag{6.12}$$

where $V_{cth} = (E_0 - kT)/q$ is a pseudo-boundary that separates the regions where $V_{c\pm}$ or V_{c0} is dominating, shown as segment a and segment c in Figure 6.3, respectively.

Finally, we combine the V_{c0} region and approximate the curvature on the boundary between V_{c0} and $V_{c\pm}$ into a single function as follows:

$$V_{c,p} = V_{c\pm}[1 - \frac{1}{\alpha}\log(1 + e^{\alpha(1-\frac{V_{c0}}{V_{c\pm}})})] \tag{6.13}$$

where $V_{c,p}$ is the V_c in a p-type TMDFET, and α is a scaling parameter. $V_{c,n}$

is the V_c in an n-type TMDFET and can be derived similarly. In the end, $V_{c,p}$ and $V_{c,n}$ are further combined into a single function as follows:

$$V_c = \beta V_{c,p} + (1 - \beta) V_{c,p} \tag{6.14}$$

where β is a sigmoid function that smoothly connects $V_{c,p}$ and $V_{c,n}$ such that V_c transitions from $V_{c,p}$ to $V_{c,n}$ at the point where $V_c = 0$. This concludes the closed-form approximation of V_c. Figure 6.3 shows a good agreement between the numerically computed V_c and the V_c computed from our approximation.

Note that in our approach, equations (6.12) to (6.14) are all derived from V_{c+} and V_{c-}, which are solved analytically from the given design parameters and the applied voltage. Therefore, the approximation remains when input parameters are changed.

After V_c is computed for $x = 0$ and $x = L$ (denoted as V_{cs} and V_{cd}, respectively), the current can be computed by equation (6.1). In our SPICE implementation, $g(V_c)$, which is a piecewise function in [98], is further combined into a single smooth function by applying the sigmoid function, similar to what is done in the V_c derivation.

6.3.2 Flexibility Modeling

From the discussion in Section 6.2.2, we know that the applied strain ϵ mainly causes the band gap E_g to decrease without affecting the mobility μ. Both theoretical and experimental results show that the E_g mainly follows a linear relationship with respect to ϵ [102, 103]. Therefore, we approximate E_g as a linear function of the uniaxial strain ϵ, while μ is treated as a constant.

Figure 6.4 shows the band gap E_g vs. uniaxial strain ϵ relationship reported in [102] and our model obtained from linear regression. Given the original band gap without bending, E_{g0}, we have $E_g = E_{g0} - 0.1046\epsilon$ for MoS_2, and $E_g = E_{g0} - 0.06778\epsilon$ for WSe_2. Note that the band gaps computed in [102] are based on the DFT methods, which are known to be slightly underestimated, and there is a small difference between the E_{g0} shown here and the band gaps in Table 6.1.

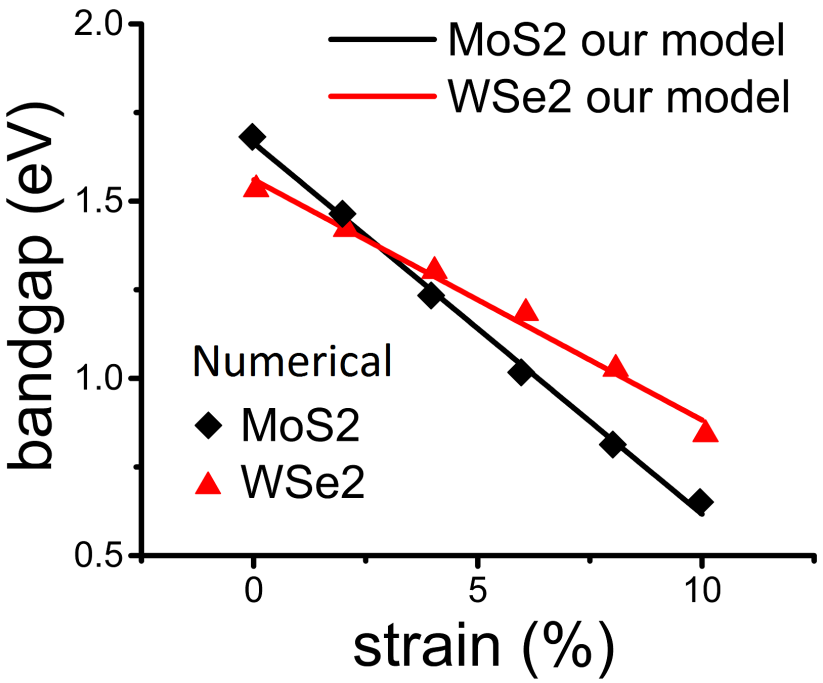

Figure 6.4: Linear model of band gap E_g vs. uniaxial strain ϵ for monolayer MoS_2 and WSe_2.

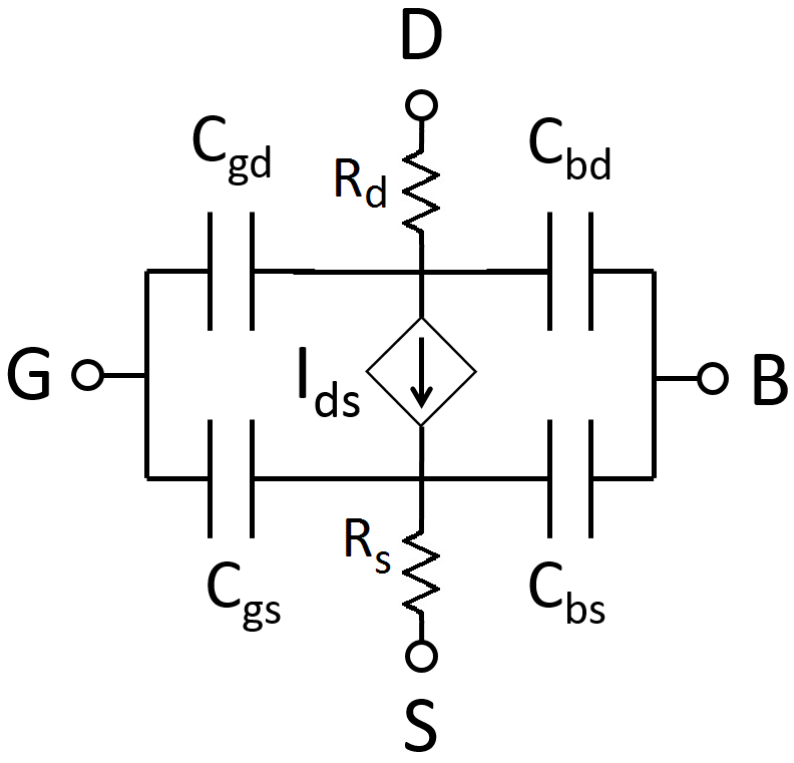

Figure 6.5: Equivalent circuit of TMDFET. G, D, S, B are electrodes. I_{ds} is the drain current.

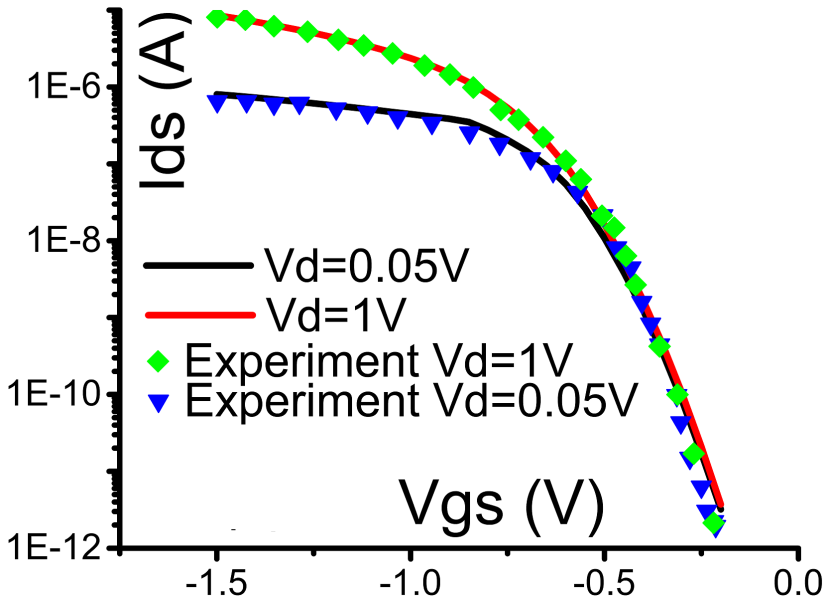

Figure 6.6: I_{ds} vs. V_{gs} at V_{ds} = -0.05 V and -1 V for the p-type WSe_2 transistor in [92].

6.3.3 Full Transistor Model in SPICE

The full transistor SPICE model is depicted in Figure 6.5, in which G (gate), D (drain), S (source), B (bulk or bottom gate) are electrodes, and I_{ds} is the drain current. Parasitic capacitors C_{gd}, C_{gs}, C_{bd}, and C_{bs} exist between neighboring electrodes. Considering the large transistor size, these capacitors are treated as parallel plate capacitors and are computed as $C_{gd} = C_{gs} = C_t \cdot W \cdot L$ and $C_{bd} = C_{bs} = C_b \cdot W \cdot L$.

The zeroth subband energy E_0, which plays an important role in the V_c and $g(V_c)$ computation, is derived from the band gap E_g, and E_g is determined by the channel material and the applied strain. In this way, our model supports different channel material and bending.

In addition, channel length modulation from V_{ds} is considered in our model by replacing the original channel length L with the effective channel length $L_{eff} = L + \mu\frac{|V_{ds}|}{v_{sat}}$ in equation (6.1), as in [98]. Here, v_{sat} is the saturation velocity of the carrier. Finally, we characterize the contact resistance introduced by the interface between the metal drain/source and the TMD channel as $R_{d/s} = 10\ \Omega{\cdot}\text{mm}/W$ [105].

To summarize, the model incorporates the following main design parameters:

- W and L: in equation (6.1) and in parasitic capacitors and resistors.
- Channel material MX_2: in band gap E_{g0}, mobility μ, and effective mass m^*.
- Bending (as strain ϵ): in band gap E_g.

This equivalent circuit is implemented in HSPICE as a *subckt* library. As we arrived at closed-form solutions for all the equations used in the model, it is a more efficient SPICE model than those in [29, 30, 1, 31] which require additional solver structures for non-closed-form quantities.

6.3.4 Model Validation

To validate our model, we compared with two sets of experimental data from different TMDFETs. One is the p-type WSe_2 transistor in the work of [92], and the other is the n-type MoS_2 transistor in [89], which are some of the best fabricated TMDFETs reported to date. Table 6.2 shows the design parameters of the experimental data and the fitting parameters (last 3 rows) used in our model for improved agreement. To be specific, D_0 and μ are scaled by a constant, while V_{cth} is shifted by an amount to achieve an adjusted transition between V_{c0} and $V_{c\pm}$, accounting for possible charge distribution differences. These do not change the underlying physics in the drift-diffusion model.

Our model, derived from the physics of [98], shows excellent agreement with the WSe_2 transistor in [92]. For the MoS_2 transistor, the I_{ds} reported in [89] was originally overestimated by about 100×. The large I_{ds} arises from the high W/L ratio and likely an overestimated carrier mobility μ. Although mobility in TMDs usually fall in the range of a few hundreds, numbers as low as 25 $cm^2V^{-1}s^{-1}$ has been reported in [91]. After the adjustment of D_0, μ, and V_{cth}, the transfer characteristics (I-V curves) computed from our model agree with the numbers reported in the work of [92] and [89], as shown in Figures 6.6 and 6.7.

Table 6.2: Parameters Used in Model Validation

Parameters	WSe_2 [92, 98]	MoS_2 [89]
W	1 μm	4 μm
L	9.4 μm	1.5 μm
$T_{ox,tg}$	17.5 nm	30 nm
$T_{ox,bg}$	270 nm	270 nm
ϵ_{tg}	12.5 ϵ_0 (ZrO_2) [a]	25 ϵ_0 (HfO_2)
ϵ_{bg}	3.9 ϵ_0 (SiO_2)	3.9 ϵ_0 (SiO_2)
μ	250 $\frac{cm^2}{V \cdot s}$	200 $\frac{cm^2}{V \cdot s}$
m^*	0.64 m_0 [92]; 0.46 m_0 [98]	0.64 m_0
E_g	1.68 eV [92]; 1.62 eV [98]	1.8 eV
V_{bs}	-40 V	0
D_0 mult.	0.25	0.36
V_{cth} shift	–	Yes
μ mult.	–	0.01

[a] ϵ_0 is the permittivity of free space.

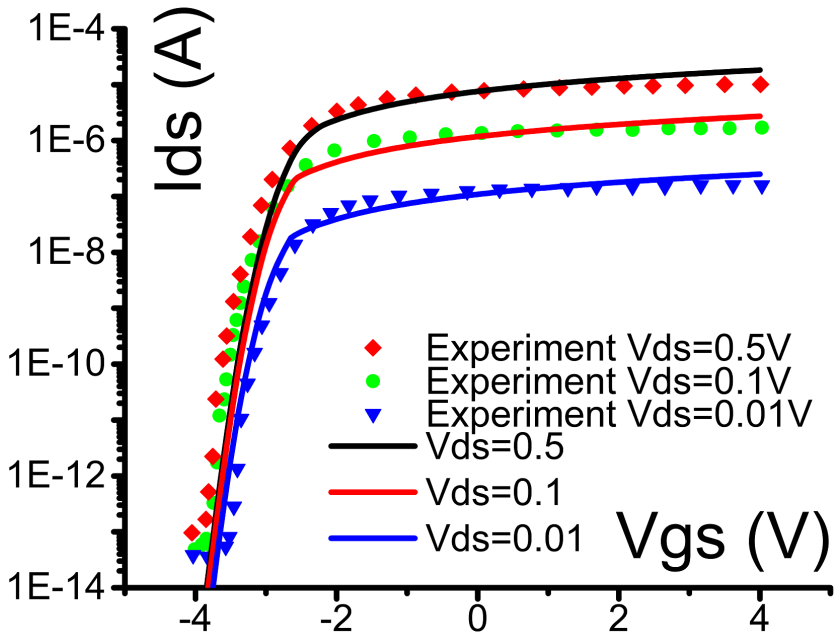

Figure 6.7: I_{ds} vs. V_{gs} at V_{ds} = 0.01 V, 0.1 V, and 0.5 V for the n-type MoS_2 transistor in [89].

6.4 Experimental Results

In the following experiments, we perform SPICE simulations using the SPICE model developed and validated in Section 6.3. Specifically, we use the two sets of parameters in Table 6.2 in order to provide realistic simulation reports that represent the best TMDFETs fabricated to date, which we investigated in Section 6.3.4. In this section, we will use the term MoS_2FET to refer to the device from [89], and the term WSe_2FET to refer to the device from [92].

As discussed in Sections 6.1 and 6.2, most fabricated TMDFETs have sizes in micrometers, which are far too large compared to the state-of-the-art technology nodes, and the aforementioned MoS_2FET and WSe_2FET are not exceptions. In our simulations, we choose to scale the MoS_2FET and WSe_2FET down to $\sim$ 100 nm in length in order to make our simulations closer to commercial reality. The work of [104] discussed the feasibility of scaling down a MoS_2 transistor and demonstrated that the short-channel effects are insignificant in 2-D transistors due to the extremely thin channel, at least down to a channel length of around 100 nm. Therefore, the drift-diffusion-based model remains accurate when it scales down to this range. To justify the scaling using the compact model, note that sizing down a monolayer TMDFET requires only the W and L to change, as the MX_2 thickness τ and oxide thickness ($T_{ox,tg}$,$T_{ox,bg}$) remain the same. W and L only affects the W/L ratio in equation (6.1) and the parasitic capacitance, and thus do not affect other parts of the model. As a result, we decide to scale the TMDFETs to the 180 nm, 130 nm, and 90 nm technology nodes for the simulation.

In addition to the unstrained MoS_2FET and WSe_2FET, we also simulated MoS_2FET and WSe_2FET with a strain $\epsilon = 10\%$ in order to observe the effects from the applied strain. The 10% strain is the maximum strain explored in the theoretical work of [102]. However, irreversible effects from bending [86] is not yet captured in our model.

A set of Si-based CMOS models of the 180 nm, 130 nm, and 90 nm technology from Predictive Technology Model (PTM) [49] serve as the baseline in our experiments. Table 6.3 gives the recommended design parameters of these technology nodes.

Table 6.3: Design Parameters of Si-Based CMOS Technology Nodes [49, 106]

Tech. Node	180 nm	130 nm	90 nm
L	180 nm	130 nm	90 nm
W_n	360 nm	260 nm	220 nm
V_{dd}	1.5 V	1.25 V	1.2 V

6.4.1 Finding Supply Voltage

Firstly, we look for the best operating supply voltage V_{dd} for the MoS_2FET and WSe_2FET. We build a seven-stage, fanout-of-four buffer chain in the 180 nm technology node in HSPICE and simulate it under different V_{dd}'s from 0.6 to 1.5 V. Delay and power are measured from the simulations. In general, the delay increases and the power decreases as V_{dd} decreases. To understand the trade-off, we compute the energy-delay product (EDP) and plot it against V_{dd}, which is shown in Figure 6.8. We observe that MoS_2FET has lower EDP as V_{dd} increases, which indicates better performance, but when V_{dd} drops to 1.2 V and below, the unstrained MoS_2FET buffer chain does not reach full swing within 100 ns due to high band gap. Therefore, we choose its nominal V_{dd} to be 1.5 V. For WSe_2FET, the EDP does not change much as V_{dd} varies, and we choose its nominal V_{dd} to be 0.8 V after considering the I-V curve in Figure 6.6. These V_{dd}'s will be used in the experiments to follow. Note that bending does not significantly affect the EDP trend with respect to V_{dd}. When scaled from 180 nm to 130 nm and 90 nm, the trend remains similar, and the average EDP becomes 29.5% and 9.32% of that of the 180-nm node for MoS_2FET at $V_{dd} = 1.5$ V, and 34.0% and 12.4% for WSe_2FET at $V_{dd} = 0.8$ V, respectively.

6.4.2 Effects of Bending

Bending the TMDFETs causes the band gap to decrease linearly, and causes I_{ds} to increase. This effect can be utilized to make the transfer characteristics tunable after the transistor is manufactured. Figure 6.9 shows the change in the I-V curves with respect to ϵ for the WSe_2FET. It shows a significant increase in I_{off} and some increase in I_{on}, which reflects as lower delay and higher power on the circuit level, which can be seen in Figure 6.10, where

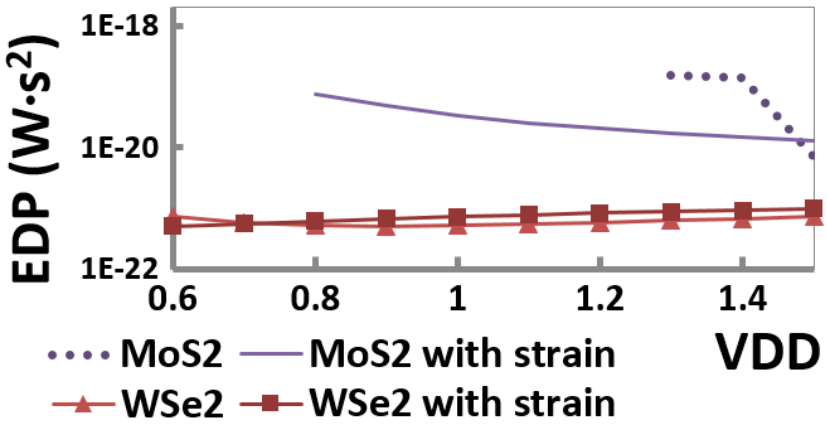

Figure 6.8: EDP vs. V_{dd} for unstrained and strained MoS_2FET and WSe_2FET. Si-based CMOS serves as a baseline.

$\epsilon = 10\%$ causes the average delay to reduce to 28.36% and 37.89% and the average power to increase to 26.8× and 9.07× for MoS_2FET and WSe_2FET, respectively.

6.4.3 Cross-Technology Comparison

In this experiment, we simulate basic logic gates *inv*, *nand2*, *nor2*, *nand3*, *nor3*, *nand4*, *xor2*, a seven-stage, fanout-of-four buffer chain, and *c17* from ISCAS'85 and report the delay, power, and EDP, shown in Figure 6.10. The delay, power, and EDP trends for each technology are consistent across all circuits. MoS_2FET has very high delay and very low power due to its small I_{ds}. WSe_2FET has comparable delay with CMOS but lower power, which makes its EDP the smallest (1 order smaller on average than CMOS and MoS_2FET), showing that it is a better material than the other two.

Next, we scale down from 180 nm to 130 nm and 90 nm for the buffer chain. The delay of unstrained MoS_2FET reduces to 52.4% and 25.3%, while power decreases 0.56% and increases 20.7%, respectively. The delay of unstrained WSe_2FET reduces to 54.9% and 29.2%, while power decreases 4.97% and increases 5.09%, respectively. The delay of CMOS reduces to 46.6% and 33.8%, while the power increases 10% and decreases 7.2%, respectively. The reduced delay can be accounted for by the quadratic reduction in gate capacitance, while the change in power can be explained by the change in I_{ds}, proportional to W/L. To summarize, all three materials' delay scale similarly,

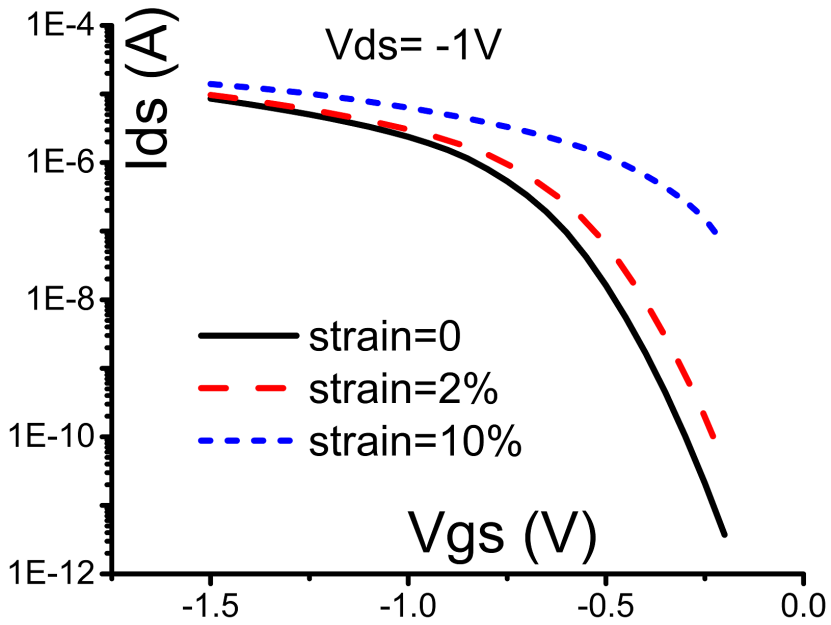

Figure 6.9: I_{ds} vs. V_{gs} at V_{ds} = -1.0 V for the p-type WSe_2 transistor under $\epsilon = 0$, 2%, and 10%.

while CMOS's power scales better.

6.4.4 Effects of Process Variation

Finally, we evaluate the effects of process variation. On the 90 nm node, we vary W and L and observe the change in delay and power of the buffer chain. As W varies by ±10%, the delay of both unstrained and strained MoS_2FET changes within 0.16%, while for WSe_2FET the change is within 0.01%, and for CMOS, 0.06%. While all show no significant change in delay, the strained TMDFETs have slightly less change in delay than unstrained. The power of MoS_2FET changes as much as 10.00%, for WSe_2FET, also 10.00%, and for CMOS, 16.64%. As L varies by ±10%, the delay of MoS_2FET changes within 20.79%, while for WSe_2FET it is within 17.22%, and for CMOS, 31.56%. The power of MoS_2FET changes as much as 10.87%, while for WSe_2FET, 6.97%, and for CMOS, 55.13%. In summary, TMDFET is more robust than CMOS in terms of both W and L, with WSe_2FET slightly more robust than MoS_2FET.

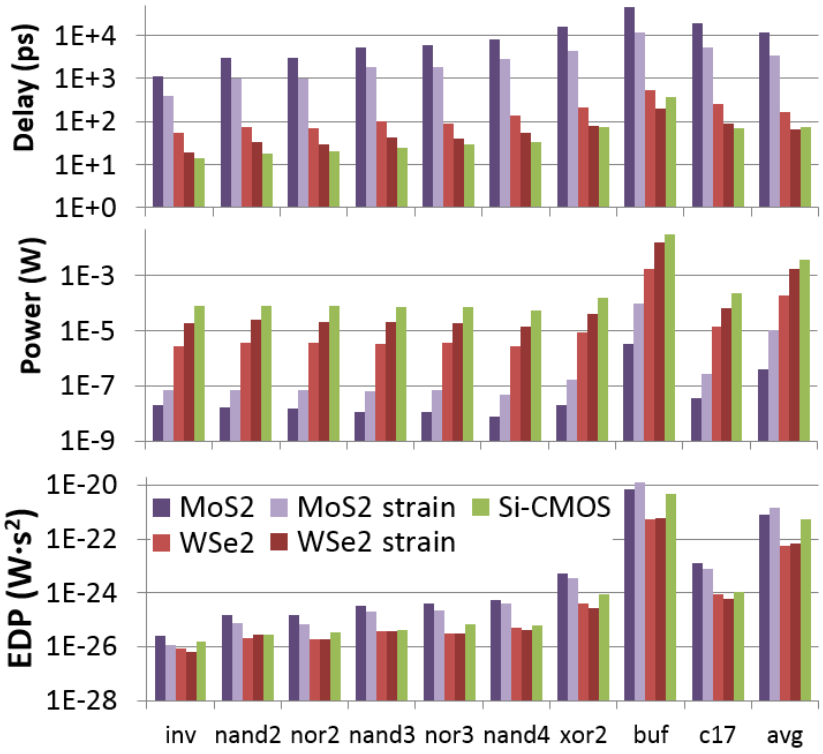

Figure 6.10: Simulation of logic gates, reporting delay, power, and EDP on the 180-nm technology node.

6.5 Conclusion

In this chapter, we presented the first SPICE model of flexible TMDFETs and simulated TMDFET circuits for the first time. We specifically modeled two successful TMDFETs in the literature, which are made of MoS_2 and WSe_2, respectively. In our simulations, we explored how these two TMDFETs perform on the circuit level. MoS_2FET consumes very low power but is very slow, while WSe_2FET outperforms Si-based CMOS in the same technology node in terms of power and EDP. Bending significantly affects the transistor characteristics and results in reduced delay and increased power. Bending also allows post-fabrication tuning for the TMDFETs. In the end, we showed the TMDFET's sensitivity to process variation in terms of delay and power. Finally, we plan to release our SPICE model as open-source in the future.

CHAPTER 7

TMDFET MODELING AND SIMULATION: A SCALING STUDY

7.1 Introduction

Flexible transistors has become a research topic of interest because they can play an essential role in the recent technology trends of wearable technology and electronic paper. Prototypes of these flexible transistors or even circuits have been successfully demonstrated in labs [107, 84, 85, 86, 87, 88]. Among the thin-film materials that can be made into flexible transistors, we focus on transition metal dichalcogenide (TMD) monolayers in this chapter.

The TMD monolayer is an emerging nano-material that garnered a lot of attention in recent years. Similar to graphene, it is a 2-D honeycomb lattice and is a robust thin-film structure, but instead of being made of carbon atoms, it consists of transition metal (denoted as M) and chalcogen atoms (denoted as X), and thus its chemical formula is expressed as MX_2. It also has outstanding electrical and physical properties like graphene. However, one major drawback of graphene is that it has no intrinsic band gap, and therefore is not semiconducting by nature. Field-effect transistors (FETs) made of 2-D graphene do not have a high I_{on}/I_{off} ratio and performs poorly in digital circuits. While graphene can be processed into nano-ribbons less than 10 nm in width in order to open up a band gap and to be made into a graphene nano-ribbon FET with higher I_{on}/I_{off} ratio, it reduces mobility, and process variation, especially line edge roughness, has been shown to severely degrade its performance [23, 67, 7, 9]. TMD, on the other hand, has a finite band gap by nature, and thus does not require further narrowing down to become semiconducting. For examples, the most well-studied TMD materials, MoS_2 and WSe_2, have band gaps of 1.80 and 1.62 eV, respectively. Table 7.1 shows some examples of TMD materials and their properties [96].

Despite successful demonstrations of flexible TMDFET instances, there

have been no circuits fabricated with flexible TMDFETs to the best of our knowledge. Therefore, there have been no experimental reports on flexible TMDFET's circuit-level performance. In order to gain a better understanding of this futuristic device and provide early assessment and evaluation studies on their potential advantages compared to conventional silicon-based technology, simulation is needed. Simulation studies of TMDFETs in the literature have so far been limited to transistor-level reports, except for the work of [8], but it was also limited to transistor sizes of 90 nm and above, far behind the state-of-the-art technology nodes.

In this work, we extend the existing TMDFET modeling work of [8] to describe the effects when scaling the transistor size down to the 16-nm technology node, and report the effects of process variation and circuit-level performance of five types TMDFETs via SPICE simulation. We also report the effects from bending. In the end, we compare the performance of TMDFET circuits with Si-based ones. To summarize, our main contribution in this work includes:

- Discussing the validity of long-channel-based models when scaling down the transistor size.
- Modeling ballistic current and validating with simulation data.
- Modeling quasi-ballisticity based on backscattering principles.
- Modeling the effect of bending in five types of TMD materials, namely MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2.
- Performing circuit-level simulations with five types of TMDFETs, focusing on supply voltage, design parameter variation, and bending.
- Presenting a study on cross-technology comparison and scaling from the 180-nm technology node down to the 16-nm technology node.

The rest of the chapter is organized as follows: Section 7.2 reviews the existing work on TMDFET modeling and discusses the issues when scaling down the transistor size below 100 nm. Section 7.3 presents our compact modeling of 16-nm TMDFETs. Section 7.4 presents the circuit-level simulation results using our developed model. Finally, Section 7.5 draws conclusions.

Table 7.1: Properties of Monolayer MX_2 [96]

MX_2	Band Gap E_g [eV]	Effective Mass	
		Electron ($\frac{m_e^*}{m_0}$)	Hole ($\frac{m_h^*}{m_0}$)
MoS_2	1.80	0.56	0.64
$MoSe_2$	1.51	0.62	0.72
$MoTe_2$	1.10	0.64	0.78
WS_2	1.93	0.33	0.43
WSe_2	1.62	0.35	0.46

7.2 Background

7.2.1 Existing Work

The work of [98] is the first compact modeling work of TMDFET. It is a long-channel drift-diffusion model and was validated with the fabricated TMDFET of [92]. Another drift-diffusion-based compact model was presented in [108]. The model's drain current was derived similarly to that in [98] except for (1) the density of states expression used in deriving channel charge Q_c is slightly different and (2) the channel charge V_c is solved differently. It was validated against the fabricated devices of [92, 89, 109, 110] and numerically solved 2-D Poisson's equation for different transistor dimension and material settings. In [8], a SPICE model based on [98] was developed. It is the first model that considers bending in flexible TMDFETs. It was also validated against the fabricated devices of [92, 89]. The validation with existing devices show that all three models yield similar results.

The model of [108] has the advantage of having a continuous expression of drain current I_D in terms of V_c that covers all three regions of the FET operation, namely linear, saturation, and subthreshold, but V_c in all these regions does not have a closed-form expression and requires a separate solver as in the work of [7, 9, 98, 29, 31]. In contrast, the model of [8] achieved all closed-form expressions. The model of [108] also features some modeling of non-idealities such as interface traps, mobility degradation, and inefficient source/drain doping. However, once these effects are included, some expressions require numerical integration and become incompatible with SPICE.

Still, all three models of [8, 98, 108] focus on long-channel devices, although the potential of the model being scaled down to sub-100 nm was discussed

in [108]. When a transistor is further scaled down to sub-20 nm, a ballistic transport model is more suitable for describing the current since the channel length becomes comparable or even less than the mean free path of TMDFET ($\sim$ 15 nm). In [83], a drain current equation based on Landauer-Büttiker formalism that describes ballistic transport is given as

$$I_D = \frac{q}{\hbar^2}\sqrt{\frac{m_y^* k_B T}{2\pi^3}} \int dE_{k_x} \left[F_{-1/2}\left(\frac{\mu_1 - E_{k_x}}{k_B T}\right) - F_{-1/2}\left(\frac{\mu_2 - E_{k_x}}{k_B T}\right)\right] T_{SD}(E_{k_x}) \tag{7.1}$$

where $F_{-1/2}$ denotes the Fermi-Dirac integral of order -1/2, T_{SD} is the transmission coefficient, μ_1 and μ_2 are electrochemical potential at the source and drain, respectively, q is the electron charge, $\hbar$ is the reduced Planck's constant, m_y^* is effective mass, q is electron charge, k_B is Boltzmann's constant, T is temperature, and E_{k_x} is the longitudinal energy. Unfortunately, $F_{-1/2}$ has no closed-form solution and must be integrated numerically [111], which makes it very difficult to develop an accurate analytical compact model. Therefore, there has been no TMDFET model based on ballistic transport to date.

In summary, existing long-channel models of [8, 98, 108] are not accurate when describing TMDFETs in modern sub-20-nm technology nodes. While a ballistic transport model would be more suitable to describe TMDFETs in this regime, it requires numerical integration and is not SPICE-compatible, making it impossible to perform circuit-level simulations. In this work, we aim to deliver a model that addresses these problems.

7.2.2 Discussion on Channel Length Scaling

7.2.2.1 Short Channel Effect

A long-channel drift-diffusion drain current model is based on gradual channel approximation (GCA), which assumes that the electric field along the channel is mostly constant as compared to the electric field perpendicular to the channel. When computing integration along the channel, the field can be treated as constant if each slice along the channel is infinitesimal. With

this assumption, the Poisson's equation that describes the electrostatics in the transistor can be reduced to 1-D. Compared with its 2-D counterpart, 1-D Poisson's equation is much simpler to solve and can be used to develop a closed-form drift-diffusion current model. For this reason, most compact models were based on GCA. [43]

When the channel length of a transistor decreases, the short-channel effect (SCE) comes into play. The dimensions of the channel length and the channel height become comparable, and the component of the electric field that is along the channel can no longer be regarded as constant. In other words, the electric field variation both along the channel and perpendicular to the channel needs to be accounted for. This makes the models based on GCA start to deviate from reality. To best describe the electric field contour in a short-channel transistor, it requires formulating and solving a 2-D Poisson's equation, which does not have a straightforward analytical solution. In some cases, it is possible to derive an analytical compact model based on 2-D Poisson's equation after a certain amount of approximation is done, as in the Si-CMOS case of [112], but the accuracy is still compromised [43].

In the work of [104], TMDFETs with channel lengths from 2 μm to 100 nm were fabricated and measured to evaluate the short-channel effect. No significant short-channel effect was observed in these transistors. Two major factors, TMD's thin-film channel and low dielectric constant, helped mitigate the short-channel effect. Furthermore, the paper speculates that the channel length can be reduced down to sub-10 nm if a state-of-the-art high-k dielectric was used in the transistor. The study also revealed that a short-channel TMDFET's current is mainly limited by the large contact resistance between TMD and metal, and this should be the main concern when scaling down TMDFETs. The contact resistance is modeled in [8] based on [105].

The work of [108] investigated the impacts from short-channel effects on the accuracy of their modeling. While an analytical current model could not be developed from 2-D Poisson's equation, they were able to compare their GCA-based closed-form subthreshold swing expression with numerical simulation based on 2-D Poisson's equation. It was shown that the computed subthreshold swing started to deviate from simulation when channel length was below 40 nm. Table 7.2 shows the extracted subthreshold swing (SS) data from [108]. It is shown that the subthreshold swing computed from the GCA-based analytical expression is an underestimation. The er-

Table 7.2: Subthreshold Swing at Different Channel Lengths

L_{CH} (nm)	SS (ana.)	SS (sim.)	diff.
6	67.83	81.72	20.48%
8	63.39	70.76	11.63%
10	61.77	65.97	6.81%
15	60.72	62.03	2.16%
20	60.21	60.95	1.22%
30	59.97	60.17	0.32%
40	59.85	60.00	0.26%
60	59.68		–
80	59.62		–
100	59.57		–

Subthreshold swing (SS, in mV/dec) from GCA-based analytical expression vs. SS from 2-D Poisson's equation simulation at different channel lengths (L_{CH}) in [108].

ror gets larger as channel length decreases, and can be as much as 20.48% at 6 nm. Nevertheless, at our target technology node of 16 (15) nm, the 2.16% error indicates that the short-channel effect is not too prominent and the long-channel model's accuracy remains reasonable at this channel length. However, in order to model technology nodes below 16 nm, new models on short channel effects should be developed.

7.2.2.2 Ballisticity and Quasi-Ballisticity

Another factor that affects the validity of a long-channel drift-diffusion model is ballisticity and quasi-ballisticity. Let L_{CH} be the channel length of the transistor and λ be the mean free path of the channel material. When $L_{CH} \gg \lambda$, the current is governed by drift-diffusion transport and the well-developed mobility theory, which leads to drift-diffusion-based compact models. When $L_{CH} \ll \lambda$, the carriers travel through the channel by ballistic transport, and the current is governed by the carrier injection from the source into the channel. On the other hand, when $L_{CH} \simeq \lambda$, the current is best described by the quasi-ballistic transport model, which is ballistic transport with some amount of scattering. [113]

The mean free path λ of TMD monolayers is $\sim$ 15 nm [83], which is close

to our target technology node with channel length $L_{CH} = 16$ nm. Therefore, we will take quasi-ballisticity into account when developing a compact model for simulation.

7.3 Compact Model

In this section, we present our approach of modeling a TMDFET with channel length < 100 nm with consideration of our target technology node of 16 nm. We choose to use the model of [8] as our base model for its simplicity and efficiency. The model includes (1) the drain current component as a function of transistor design parameters and applied voltages and (2) parasitic capacitors and resistors. Further adaptations of quasi-ballisticity and flexibility are explained in this section.

7.3.1 Adaptation for Quasi-Ballisticity

In a drift-diffusion model, the current in the saturation region can be expressed as

$$I_{D,sat} = WC_{ox}(V_G - V_t)v_{sat} \tag{7.2}$$

where W is the channel width, C_{ox} is the oxide capacitance, V_t is the threshold voltage, and v_{sat} is the saturation velocity. This equation describes the charge $Q = WC_{ox}(V_G - V_t)$ moving under velocity v_{sat}, resulting in drain current I_D.

The ballistic current can be approximated similarly as

$$I_{D,bal} = WC_{ox}(V_G - V_t)v_{inj} \tag{7.3}$$

with v_{inj} being the injection velocity, at which the carriers are injected into the channel, constituting the drain current [113]. Then, the ballistic enhancement factor (BEF) [114] can be computed as

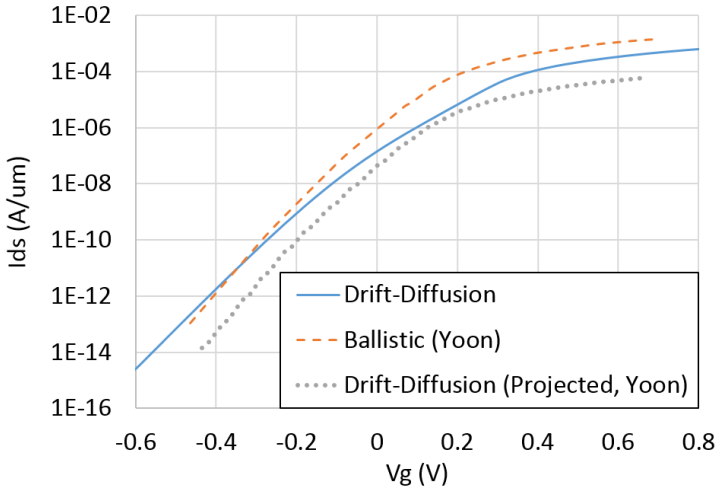

Figure 7.1: Comparison of MoS_2FET currents computed from the drift-diffusion model of [8] and from the ballistic simulation in [83] and projected drift-diffusion current estimated from their ballistic current.

$$BEF = \frac{v_{inj}}{v_{sat}} = \frac{I_{D,bal}}{I_{D,sat}} \tag{7.4}$$

which means the ballistic current can be approximated as

$$I_{D,bal} = I_{D,sat} \cdot BEF \tag{7.5}$$

Figure 7.1 shows the difference between the current computed from the model of [8], the ballistic simulation in [83], and the projected drift-diffusion current estimated from the ballistic current in [83]. The ballistic simulation was done with a device with L_{CH} = 15 nm and HfO_2 gate oxide thickness T_{ox} = 2.8 nm. The projected drift-diffusion current was estimated by assuming L_{CH} = 500 nm and using a backscattering model of $I_{proj} = I_{bal} \cdot \lambda/(L_{CH} + \lambda)$, which is an underestimation as compared to the drift-diffusion current computed from the model of [8] (which also assumes L_{CH} = 500 nm). By inspecting the current in the saturation region, we have $BEF \simeq 2.5$.

After adjusting the drift-diffusion current by multiplying with BEF, we have the results in Figure 7.2, showing good agreement. This gives a simple, efficient and yet valid approximation of ballistic current without the need to solve equation (7.1). Furthermore, as shown in Figure 7.1, the ratio between ballistic current and drift-diffusion current varies in different regions. We can

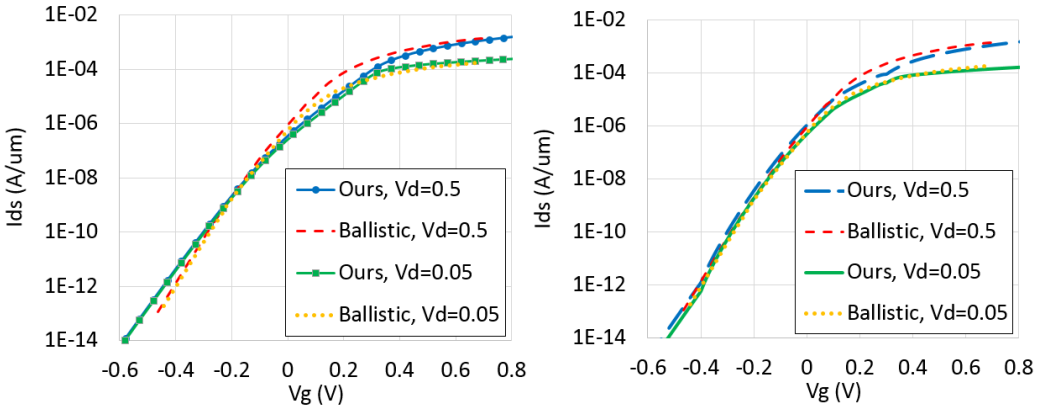

Figure 7.2: Drift-diffusion model after the BEF adjustment compared with ballistic simulation in [83] well. **(Left)**, with constant BEF. **(Right)**, with piecewise BEF.

divide the I-V curve into 12 regions by the values of gate and drain voltages V_G and V_D and apply piecewise BEF values in each region, which results in a better agreement with the ballistic current, shown in Figure 7.2 (right). The efficiency is further discussed in Section 7.4.4.

Finally, the quasi-ballistic current is computed from the ballistic current by adding the scattering effect as

$$I_{D,qb} = I_{D,bal} \cdot \left(\frac{1-R}{1+R} \right) \tag{7.6}$$

where R is the backscattering coefficient [113]. This approximation is validated with Monte Carlo simulations [115]. R can be determined empirically by comparing the theoretical ballistic current with measurement results, or can be theoretically computed by

$$R = \frac{l}{l+\lambda} \tag{7.7}$$

where λ is the mean free path and l is the critical distance [116]. It has been observed that backscattering only occurs within l from the beginning of the channel [115], and l is typically the distance where the channel potential drops by k_BT/q, a small portion of the channel length [116]. Here, we set $l = \frac{1}{6}L_{CH}$. At $L_{CH} = 15$ nm, the resulting $R = \frac{1}{7}$ and $\frac{1-R}{1+R} = 0.75$.

Note that since the current based on pure ballistic transport is independent of L_{CH}, the BEF as a constant obtained from the 15-nm simulation

results should remain valid as long as the transistor operates in the ballistic and quasi-ballistic regimes. The subsequent quasi-ballistic current computation, on the other hand, has R that depends on L_{CH}, and therefore the current computed from the final model has dependency on L_{CH}. However, for the piecewise BEF implementation, each region is determined empirically from the simulation data, and therefore one piecewise description cannot be applied universally across all devices.[1] Unfortunately, we only have the unstrained MoS_2FET simulation data from [83] and are not able to use this implementation for other types of TMDFETs. Nevertheless, if a numerical simulation framework is made available, our methodology can be adopted to create better fitting TMDFET models for other types of materials. In Section 7.4, we only report SPICE simulation data obtained from the piecewise BEF model for the MoS_2FET. Data of other TMDFETs are obtained using the constant BEF model.

7.3.2 Modeling Flexibility

With the above model now describing a quasi-ballistic TMDFET, we proceed to add the model of bending. First, we define bending in terms of the *applied strain* ϵ as follows:

$$\epsilon = \frac{\tau}{R_b} \tag{7.8}$$

where R_b is bending radius and τ is half film thickness.

Theoretical and experimental studies on TMDFET bending [101, 102, 103] show that the band gap of TMD monolayers has a linearly decreasing relationship with respect to ϵ. Other parameters such as mobility and threshold voltage may vary with bending, but they show little correlation with ϵ. We take the approach of [8] to compute an updated band gap under applied strain and then replace the original band gap in the compact model for subsequent computations. The model is expressed as

[1] Empiricism targeting specific devices is commonly done in compact modeling in order to create a *local model* that can be used to simulate the targeted device with better accuracy. In contrast, a more generic *global model* describes a larger range of devices with some compromised accuracy [117].

Table 7.3: Band Gap Computation Parameters

MX_2	E_{g0}	c
MoS_2	1.80 eV	0.1046
$MoSe_2$	1.51 eV	0.06958
$MoTe_2$	1.10 eV	0.04006
WS_2	1.93 eV	0.1078
WSe_2	1.62 eV	0.06778

Parameters for computing the band gap under bending in equation (7.9).

$$E_g = E_{g0} - c\epsilon \tag{7.9}$$

where E_g is the updated band gap, E_{g0} is the original band gap without bending, and c is a material depending coefficient. The parameters are obtained from linear curve fitting with the data from [102] and are listed in Table 7.3 for five different TMD materials. Since band gap is only dependent on material type and banding, equation (7.9) and the parameters in Table 7.3 remain valid across different simulation settings.

7.4 Experimental Results

We implemented the model presented above in SPICE and performed a series of circuit-level simulations to evaluate TMDFET circuits' performance. We define the transistor size as per the 16-nm transistors in Predictive Technology Model (PTM) [49]. We simulate five types of TMDFETs, namely MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2 FETs. We also simulate unstrained and strained versions of TMDFETs. The maximum strain we simulate is $\epsilon = 10\%$, which means the bending radius R_b is only $5\times$ of the TMD film thickness. For unstrained MoS_2, we use both the constant BEF and the piecewise BEF models, with the latter denoted as *MoS_2 adjusted*.

We first explore the optimal supply voltage for each of the five types of TMDFETs in Section 7.4.1. Next, we evaluate the effects from variation in design parameters such as transistor width W, channel length L_{CH}, and oxide thickness T_{ox} in Section 7.4.2. Next, we compare the circuit-level performance

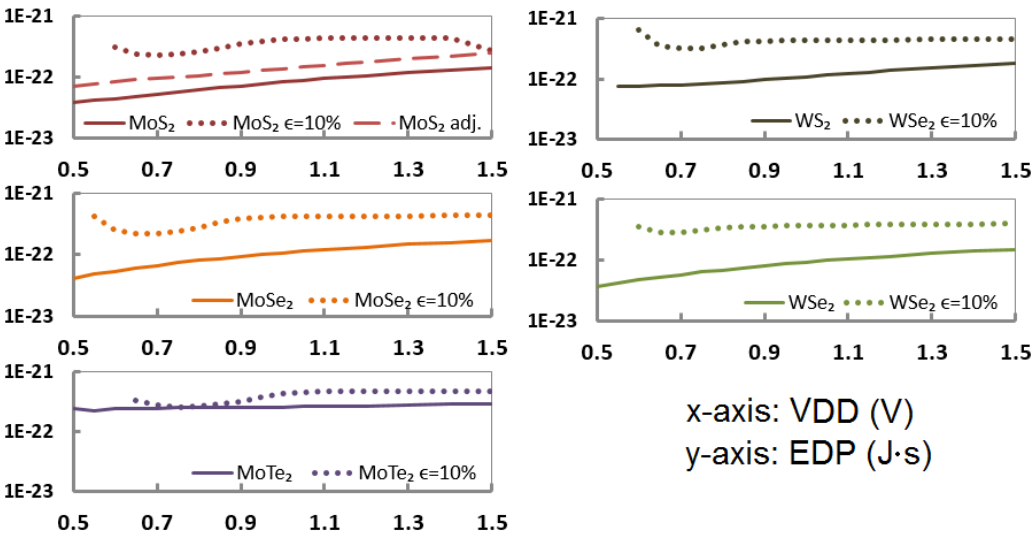

Figure 7.3: V_{DD} vs. EDP of TMDFETs.

of all five TMDFETs with those of Si-based transistors from PTM in Section 7.4.3. Finally, we discuss the running time of our SPICE model in Section 7.4.4.

7.4.1 Supply Voltage Exploration

In this section, we explore the optimal supply voltage for TMDFETs based on simulation results. We built a seven-stage fanout-of-four buffer chain with all five types of TMDFETs in SPICE. We sweeped the supply voltage V_{DD} from 0.6 V to 1.5 V and measured the energy-delay product (EDP) from SPICE simulations. The results are shown in Figure 7.3. We show that the EDP of unstrained TMDFETs mostly increases with V_{DD} within the range we sweeped. However, TMDFETs with $\epsilon = 10\%$ have a minimum EDP between $V_{DD} = 0.65$ to 0.75 V. When V_{DD} is further decreased, TMDFETs with $\epsilon = 10\%$ become very slow (with delay > 50 ps) and we omitted those data points in Figure 7.3. By averaging the EDP of $\epsilon = 0\%$ and $\epsilon = 10\%$, the resulting optimal V_{DD} for MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2 FETs are 0.70 V, 0.65 V, 0.75 V, 0.70 V, and 0.70 V, respectively. It is also shown that the EDP of TMDFETs with $\epsilon = 10\%$ is 1.26× to 3.93× higher than that of the unstrained version, depending on material type.

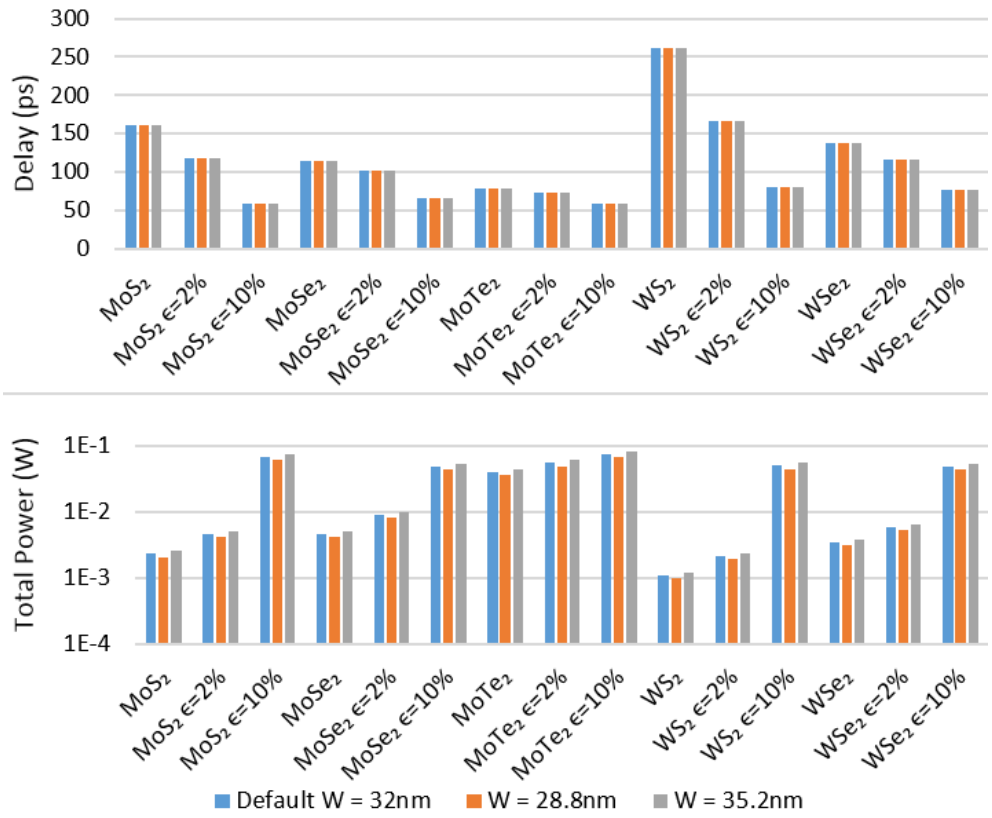

Figure 7.4: Delay and power of TMDFETs under W variation. Note that the delay variation is less than 1% and is hard to see here.

7.4.2 Variation in Design Parameters

In this section, we take the buffer chain from the previous section and vary the transistor width W, channel length L_{CH}, and oxide thickness T_{ox} by 10% to report the variation in delay and power. The default transistor parameters are $W = 32$ nm, $L_{CH} = 16$ nm, and $T_{ox} = 2.8$ nm. The results are shown in Figure 7.4 to Figure 7.6. We show that variation in L_{CH} results in the most change in delay, while variation in W and T_{ox} results in more change in power than variation in L. We also show that the change with respect to variation is proportional to the band gap, that is, materials with higher band gap or under higher applied strain suffer from the effects from process variation more significantly.

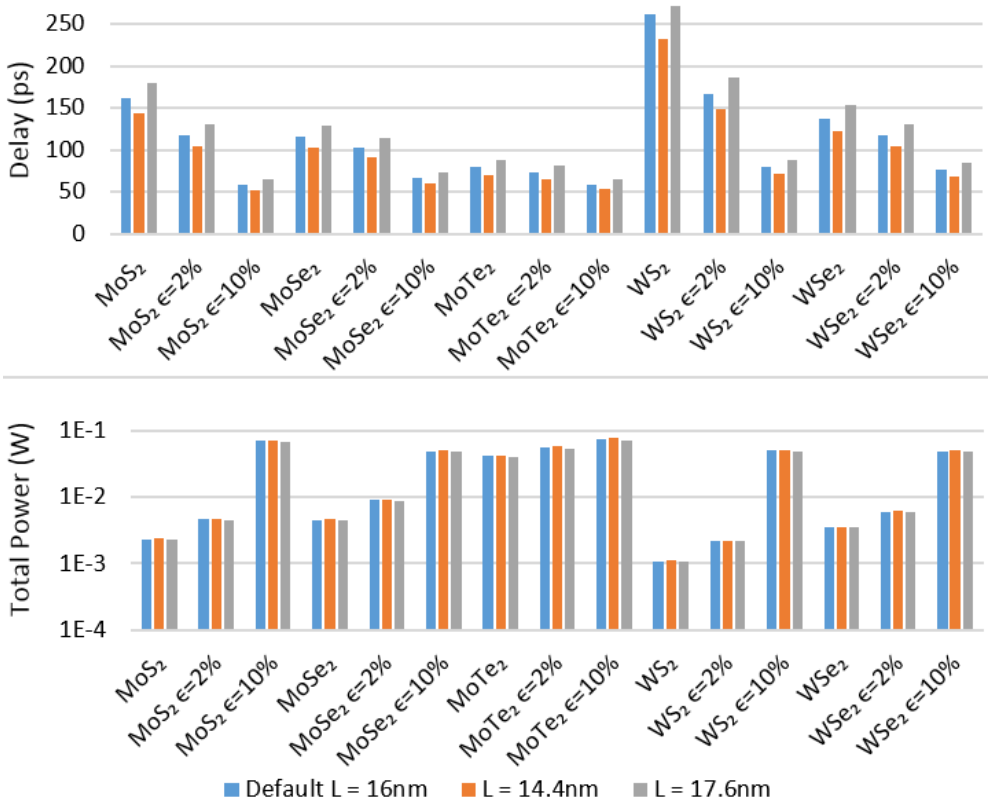

Figure 7.5: Delay and power of TMDFETs under L_{CH} variation.

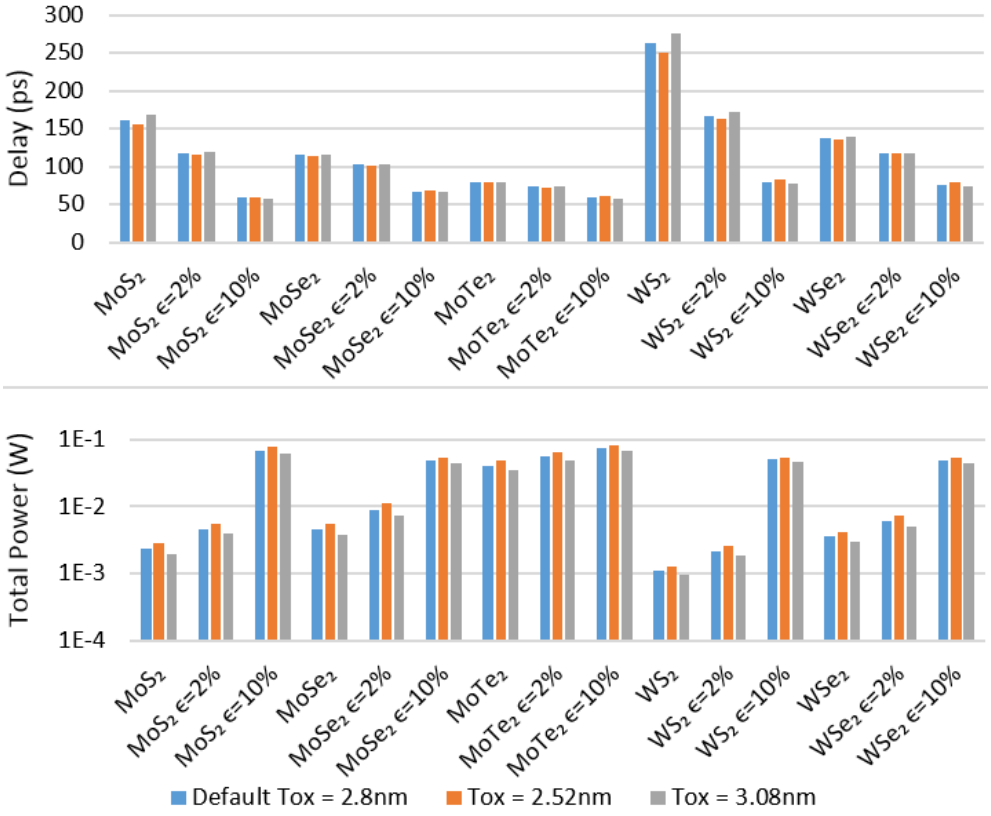

Figure 7.6: Delay and power of TMDFETs under T_{ox} variation.

7.4.3 Cross-Technology Comparison and Scaling

In this section, we perform simulations on basic logic gates of *inv*, *nand2*, *nor2*, *nand3*, *nor3*, *nand4*, *xor2*, a seven-stage, fanout-of-four buffer chain, and *c17* from ISCAS'85 and report delay, power, and EDP averaged from all nine circuits. The high-performance (HP) bulk-Si, low-power (LP) bulk-Si, high-performance (HP) Si-based FinFET, and low-standby-power (LSTP) Si-based FinFET from PTM serve as baselines for comparison. We also report the delay, power, and EDP from the 180-nm and 90-nm technology nodes from simulations using the model from [8].[2] The results are shown in Figure 7.7 to Figure 7.9.

The delay of TMDFETs almost all range from 50-95 ps except for WSe_2 FETs, which are significantly faster. Bending causes the band gap to decrease, resulting in an overall higher current, which in turn makes the delay decrease. However, when the band gap decreases too much from bending, the I_{on}/I_{off} ratio may become too low, resulting in a low noise margin of the circuit. This causes some circuits with more transistors in series such as *nand2* and *nand3* to switch slower and *nand4* fails to switch correctly, while other circuits such as *inv* and buffer chain do have lower delay than their unstrained counterparts. Take WSe_2 for example. The delays of *inv* with $\epsilon =$ 0, 2%, and 10% are 52.17 ps, 44.76 ps, and 24.47 ps, respectively. Meanwhile, the delays of *nand2* with $\epsilon = 0$, 2%, and 10% are 62.35 ps, 56.89 ps, and 64.56 ps, respectively. Note that the 180-nm and 90-nm MoS_2FET modeled in [8] was based on [89], which has a much lower current than theoretically predicted,[3] and therefore operates very slowly, as shown in the insets of Figure 7.7. The insets also show that both MoS_2FET and WSe_2FET's delay improves as the transistor size scales down.

The power of TMDFETs, as reported in Figure 7.8, is negatively correlated with the band gap of the material. For example, WS_2, which has the lowest power, has an intrinsic band gap of 1.93 eV, while $MoTe_2$, which has the highest power, has an intrinsic band gap of 1.10 eV. This is because

[2]The work of [8] only modeled MoS_2 and WSe_2 FETs, so only these two types of TMDFETs are reported.

[3]While a mobility of 200 $cm^2/(V{\cdot}s)$ was extracted in [89], a mobility of ~ 2 $cm^2/(V{\cdot}s)$ was used in the model of [8, 98] to match the I-V curve reported in [89], which implies the current is much lower than theoretically predicted. It is possible that phonon scattering reduced the mobility of the device, with the lowest reported mobility being 0.1 $cm^2/(V{\cdot}s)$ [89].

the current is inversely proportional to the band gap. Meanwhile, bending decreases the band gap, and thus increase the current and power. When compared with the 180-nm and 90-nm technology nodes, the power tends to increase as transistor size scales down, which comes from the increased total current from reduced channel length. Again, the 180-nm and 90-nm MoS_2FET from [89] has a very small current, resulting in very low power.

The EDP of TMDFETs, shown in Figure 7.9, is also negatively correlated with the band gap of the material. It generally increases with bending and decreases with transistor size scaling. It is also shown that the amount of the EDP increase with respect to bending is higher when the transistor size is smaller, that is, bending affects EDP more significantly at smaller transistor sizes. On the 180-nm and 90-nm technology nodes, WSe_2FET's EDP is only 10.6% and 31.2% of that of Si-based transistors, which means it is a better device. However, TMDFETs do not outperform Si-based transistors in terms of EDP (at least 4.9× of that of the best performing Si-based transistor) on the 16-nm technology node, although the Si-based transistors are not flexible.

Finally, with the more accurate piecewise BEF model of MoS_2FET, the delay is 19.83% higher and power is 2.32×, which comes from an overall lower ON current and higher OFF current. However, the EDP is 1.90× higher than that from the constant BEF model, which makes MoS_2FET a worse device than predicted by the constant BEF model.

From these experimental results, we show that flexible TMDFETs can be tuned by bending to achieve a lower delay at the cost of higher power and EDP. This allows post-fabrication tuning for delay-power tradeoff. It also provides opportunities to improve yield; if a circuit does not meet the timing constraints, the critical path delay may be reduced by bending such that it meets the timing constraints again. However, too much bending may result in errors in circuit operations.

7.4.4 Discussion on Running Time

As discussed previously, our SPICE model, which is based on the work of [8], has all closed-form expressions and therefore is efficient to compute. Take the *c17* circuit for example. The running time of across five TMD materials with different bending ranges from 12 minutes to 74 minutes, with the average

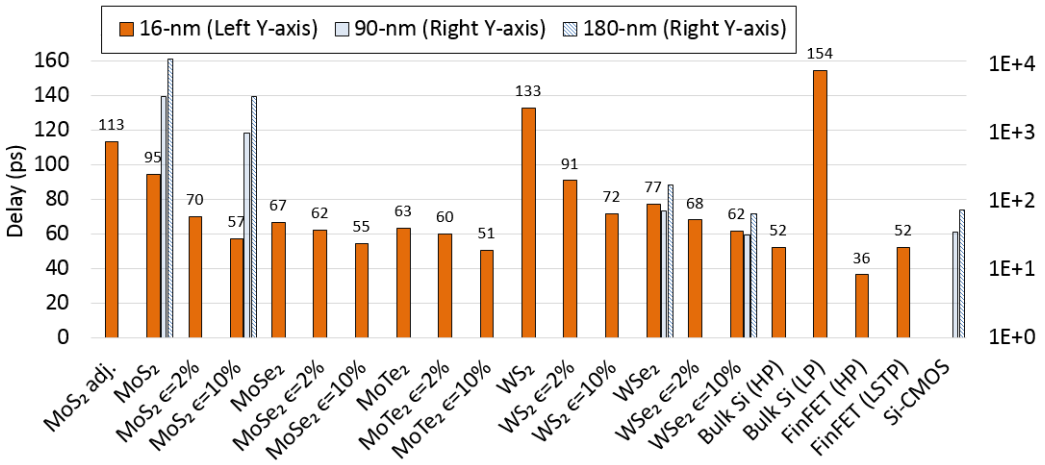

Figure 7.7: Max delay of different technologies on 16-nm, 90-nm, and 180-nm. (Average of 9 circuits.) Labels denote the 16-nm node.

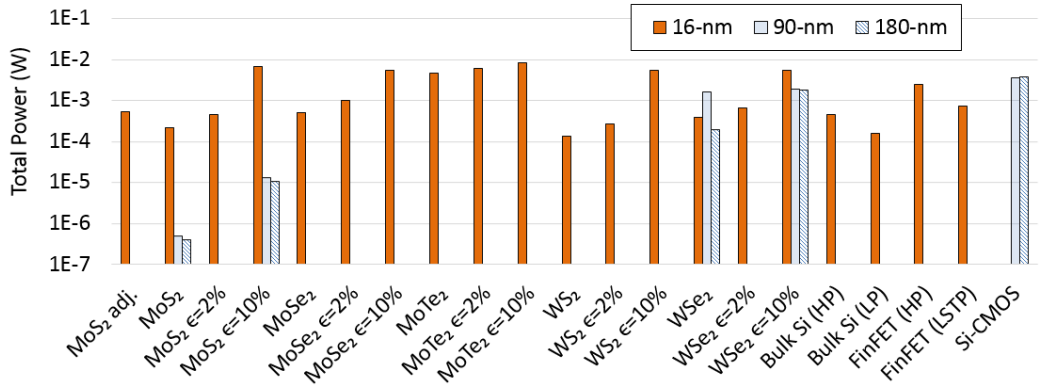

Figure 7.8: Total power of different technologies on 16-nm, 90-nm, and 180-nm. (Average of 9 circuits.)

being 38 minutes. With piecewise BEF, however, the average running time of MoS_2FET increases to 293 minutes. Meanwhile, it takes $\sim$ 8 hours for a solver-based model [9] to perform the same simulation.

7.5 Conclusion

To summarize, we scaled down the existing compact models of TMDFETs to describe transistors on contemporary 16-nm technology nodes. We performed extensive SPICE simulations on the circuit-level and explored and compared the delay and power performance of five types of TMDFETs, made of MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2, and compared them with Si-based transistors.

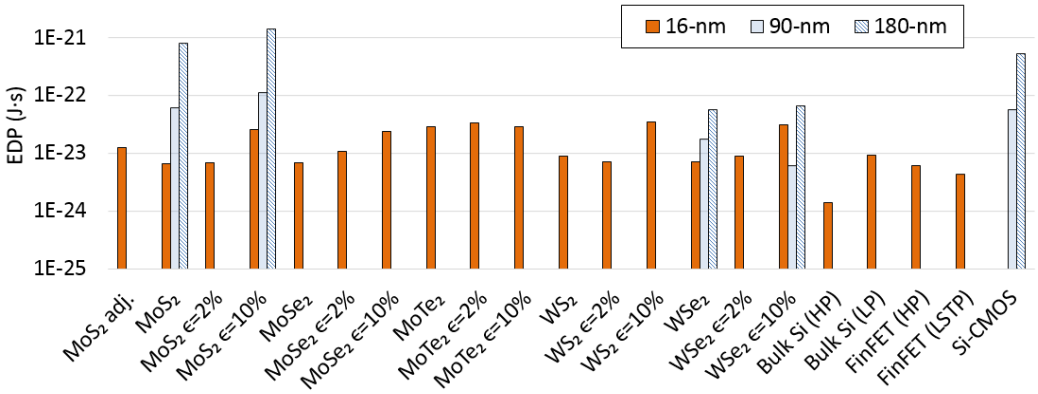

Figure 7.9: EDP of different technologies on 16-nm, 90-nm, and 180-nm.. (Average of 9 circuits.)

We also investigated the effects from bending and variation in flexible TMD-FETs. We show that bending results in lower delay at the cost of higher power and the risk of poor transistor operation. This creates a larger design space and opportunities in post-fabrication tuning. Finally, we plan to release our SPICE model as open-source in the future.

CHAPTER 8

CONCLUSION

In this dissertation, we reviewed practical issues in the implementation of GNRFET- and TMDFET-based circuits, developed SPICE-compatible compact models for GNRFETs and TMDFETs, discussed transistor-level and gate-level properties of GNRFET and TMDFET circuits, and reported circuit-level simulations results on delay and power performance of GNRFET and TMDFET circuits. Our results show that both MOS-GNRFET and SB-GNRFET perform better than Si-CMOS in terms of EDP under ideal cases. Also, MOS-GNRFET has great potential in low-power applications, while SB-GNRFET is suitable for high-performance applications with its excellent EDP. However, when line edge roughness is present, the delay and power benefits from both types of GNRFETs are significantly reduced. Future refinement in GNRFET fabrication techniques is critical in order to make GNRFET a competitive technology. For TMDFETs, not all outperform Si-CMOS in terms of power and EDP. We also showed that bending can increase the overall current in TMDFETs and thus reduce delay but increase power consumption. It provides opportunities in post-fabrication tuning for power-delay trade-off.

REFERENCES

[1] Y.-Y. Chen, A. Rogachev, A. Sangai, G. Iannaccone, G. Fiori, and D. Chen, "A SPICE-compatible model of graphene nano-ribbon field-effect transistors enabling circuit-level delay and power analysis under process variation," in *Design, Automation Test in Europe Conference Exhibition (DATE), 2013*, March 2013, pp. 1789–1794.

[2] Y.-Y. Chen, A. Sangai, M. Gholipour, and D. Chen, "Schottky-barrier-type graphene nano-ribbon field-effect transistors: A study on compact modeling, process variation, and circuit performance," in *Nanoscale Architectures (NANOARCH), 2013 IEEE/ACM International Symposium on*, July 2013, pp. 82–88.

[3] Y.-Y. Chen, A. Sangai, M. Gholipour, and D. Chen, "Graphene nano-ribbon field-effect transistors as future low-power devices," in *Low Power Electronics and Design (ISLPED), 2013 IEEE International Symposium on*, Sept 2013, pp. 151–156.

[4] Y.-Y. Chen, M. Gholipour, A. Sangai, and D. Chen, "Effects of process variation on the circuit-level performance of graphene nano-ribbon field-effect transistors." [Online]. Available: https://docs.google.com/file/d/0BzOfmcirXgUgak51a3FvYzQ5TDA/

[5] M. Gholipour, Y.-Y. Chen, A. Sangai, and D. Chen, "Highly accurate SPICE-compatible modeling for single- and double-gate GNRFETs with studies on technology scaling," in *Design, Automation and Test in Europe Conference and Exhibition (DATE), 2014*, March 2014, pp. 1–6.

[6] M. Gholipour, N. Masoumi, Y.-Y. Chen, D. Chen, and M. Pourfath, "Asymmetric gate Schottky-barrier graphene nanoribbon fets for low-power design," *Electron Devices, IEEE Transactions on*, vol. 61, no. 12, pp. 4000–4006, Dec 2014.

[7] M. Gholipour, Y.-Y. Chen, A. Sangai, N. Masoumi, and D. Chen, "Analytical SPICE-compatible model of Schottky-barrier-type GNRFETs with performance analysis," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, pp. 1–14, 2015.

[8] Y.-Y. Chen and D. Sun, Zelei Chen, "A SPICE model of flexible transition metal dichalcogenide field-effect transistors," in *Proceedings of the 52nd annual Design Automation Conference*, June 2015.

[9] Y.-Y. Chen, A. Sangai, A. Rogachev, M. Gholipour, G. Iannaccone, and D. Chen, "A SPICE-compatible model of MOS-type graphene nano-ribbon field-effect transistors enabling gate- and circuit-level delay and power analysis under process variation," *Nanotechnology, IEEE Transactions on*, 2015 (to appear).

[10] A. Javey, J. Guo, D. B. Farmer, Q. Wang, D. Wang, R. G. Gordon, M. Lundstrom, and H. Dai, "Carbon nanotube field-effect transistors with integrated ohmic contacts and high-k gate dielectrics," *Nano Letters*, vol. 4, no. 3, pp. 447–450, 2004.

[11] A. K. Geim and K. S. Novoselov, "The rise of graphene," *Nature Materials*, vol. 6, no. 3, pp. 183–191, 2007.

[12] A. K. Geim, "Graphene: Status and prospects," *Science*, vol. 324, no. 5934, pp. 1530–1534, 2009. [Online]. Available: http://www.sciencemag.org/content/324/5934/1530.abstract

[13] S. Chilstedt, C. Dong, and D. Chen, *Transistors: Types, Materials and Applications, Chapter: Carbon nanomaterials transistors and circuits.* Nova Science Publishers, 2010.

[14] D. Chen, S. Chilstedt, C. Dong, and E. Pop, "What everyone needs to know about carbon-based nanocircuits," *Online Knowledge Center, IEEE/ACM Design Automation Conference*, 2010.

[15] W. de Heer, C. Berger, E. Conrad, P. First, R. Murali, and J. Meindl, "Pionics: The emerging science and technology of graphene-based nanoelectronics," in *Electron Devices Meeting, 2007. IEEE International*, Dec 2007, pp. 199–202.

[16] M. P. Levendorf, C. S. Ruiz-Vargas, S. Garg, and J. Park, "Transfer-free batch fabrication of single layer graphene transistors," *Nano Letters*, vol. 9, no. 12, pp. 4479–4483, 2009.

[17] P. J. Wessely, F. Wessely, E. Birinci, U. Schwalke, and B. Riedinger, "Transfer-free fabrication of graphene transistors," *Journal of Vacuum Science and Technology B*, vol. 30, no. 3, 2012.

[18] A. Betti, G. Fiori, and G. Iannaccone, "Strong mobility degradation in ideal graphene nanoribbons due to phonon scattering," *Applied Physics Letters*, vol. 98, no. 21, p. 212111, 2011.

[19] Y.-W. Son, M. L. Cohen, and S. G. Louie, "Energy gaps in graphene nanoribbons," *Physical Review Letters*, vol. 97, no. 21, p. 216803, 2006.

[20] K. Nakada, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus, "Edge state in graphene ribbons: Nanometer size effect and edge shape dependence," *Physical Review B*, vol. 54, no. 24, p. 17954, 1996.

[21] X. Wang, Y. Ouyang, X. Li, H. Wang, J. Guo, and H. Dai, "Room-temperature all-semiconducting sub-10-nm graphene nanoribbon field-effect transistors," *Physical Review Letters*, vol. 100, no. 20, p. 206803, 2008.

[22] H. Raza, "Zigzag graphene nanoribbons: Bandgap and midgap state modulation," *Journal of Physics: Condensed Matter*, vol. 23, no. 38, p. 382203, 2011.

[23] G. Fiori and G. Iannaccone, "Simulation of graphene nanoribbon field-effect transistors," *Electron Device Letters, IEEE*, vol. 28, no. 8, pp. 760–762, 2007.

[24] Y. Yoon, G. Fiori, S. Hong, G. Iannaccone, and J. Guo, "Performance comparison of graphene nanoribbon FETs with Schottky contacts and doped reservoirs," *Electron Devices, IEEE Transactions on*, vol. 55, no. 9, pp. 2314–2323, 2008.

[25] P. Michetti and G. Iannaccone, "Analytical model of one-dimensional carbon-based Schottky-barrier transistors," *Electron Devices, IEEE Transactions on*, vol. 57, no. 7, pp. 1616–1625, 2010.

[26] M. Choudhury, Y. Yoon, J. Guo, and K. Mohanram, "Technology exploration for graphene nanoribbon FETs," in *Proceedings of the 45th Annual Design Automation Conference*. ACM, 2008, pp. 272–277.

[27] M. R. Choudhury, Y. Yoon, J. Guo, and K. Mohanram, "Graphene nanoribbon FETs: Technology exploration for performance and reliability," *Nanotechnology, IEEE Transactions on*, vol. 10, no. 4, pp. 727–736, 2011.

[28] S. Fregonese, C. Maneux, and T. Zimmer, "A versatile compact model for ballistic 1d transistor: Applications to GNRFET and CNTFET," in *ISDRS*, 2009, pp. 1–2.

[29] 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*, vol. 54, no. 12, pp. 3186–3194, 2007.

[30] J. Deng and H.-S. Wong, "A compact SPICE model for carbon-nanotube field-effect transistors including nonidealities and its application Part II: Full device model and circuit performance benchmarking," *TED*, vol. 54, no. 12, pp. 3195–3205, 2007.

[31] G. M. Landauer, D. Jimenez, and J. L. Gonzalez, "An accurate and verilog-a compatible compact model for graphene field-effect transistors," *Nanotechnology, IEEE Transactions on*, vol. 13, no. 5, pp. 895–904, 2014.

[32] "Spice model of graphene nanoribbon fets." [Online]. Available: https://nanohub.org/resources/17074

[33] T. Yan, Q. Ma, S. Chilstedt, M. D. F. Wong, and D. Chen, "A routing algorithm for graphene nanoribbon circuit," *ACM Transactions on Design Automation of Electronic Systems*, vol. 18, no. 4, pp. 61:1–61:18, Oct 2013. [Online]. Available: http://doi.acm.org/10.1145/2505056

[34] S. S. Chauhan, P. Srivastava, and A. K. Shrivastava, "Band gap engineering in zigzag graphene nanoribbons an *Ab Initio* approach," *Journal of Computational and Theoretical Nanoscience*, vol. 9, no. 8, pp. 1084–1089, 2012.

[35] "NanoTCAD ViDES simulator." [Online]. Available: http://vides.nanotcad.com/vides/

[36] Z. Chen, Y.-M. Lin, M. J. Rooks, and P. Avouris, "Graphene nanoribbon electronics," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 40, no. 2, pp. 228 – 232, 2007.

[37] X. Wang and H. Dai, "Etching and narrowing of graphene from the edges," *Nature Chemistry*, vol. 2, pp. 661–665, 2010.

[38] X. Li, X. Wang, L. Zhang, S. Lee, and H. Dai, "Chemically derived, ultrasmooth graphene nanoribbon semiconductors," *Science*, vol. 319, no. 5867, pp. 1229–1232, 2008.

[39] J. Cai, P. Ruffieux, R. Jaafar, M. Bieri, T. Braun, S. Blankenburg, M. Muoth, A. P. Seitsonen, M. Saleh, X. Feng, K. Mllen, and R. Fasel, "Atomically precise bottom-up fabrication of graphene nanoribbons," *Nature*, vol. 466, pp. 470–473, July 2010.

[40] L. Jiao, L. Zhang, L. Ding, J. Liu, and H. Dai, "Aligned graphene nanoribbons and crossbars from unzipped carbon nanotubes," *Nano Research*, vol. 3, no. 6, pp. 387–394, 2010. [Online]. Available: http://dx.doi.org/10.1007/s12274-010-1043-z

[41] S. Prezioso, F. Perrozzi, M. Donarelli, F. Bisti, S. Santucci, L. Palladino, M. Nardone, E. Treossi, V. Palermo, and L. Ottaviano, "Large area extreme-UV lithography of graphene oxide via spatially resolved photoreduction," *Langmuir*, vol. 28, no. 12, pp. 5489–5495, 2012.

[42] K. Nagashio, T. Nishimura, K. Kita, and A. Toriumi, "Metal/graphene contact as a performance killer of ultra-high mobility graphene analysis of intrinsic mobility and contact resistance," in *Electron Devices Meeting (IEDM), 2009 IEEE International*. IEEE, 2009, pp. 1–4.

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

[44] S. Mukhopadhyay, A. Raychowdhury, and K. Roy, "Accurate estimation of total leakage in nanometer-scale bulk CMOS circuits based on device geometry and doping profile," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 24, no. 3, pp. 363–381, March 2005.

[45] J. Knoch and J. Appenzeller, "Tunneling phenomena in carbon nanotube field-effect transistors," *Physica Status Solidi (a)*, vol. 205, no. 4, pp. 679–694, 2008. [Online]. Available: http://dx.doi.org/10.1002/pssa.200723528

[46] "FastCap 2." [Online]. Available: http://www.fastfieldsolvers.com/

[47] G. H. See, X. Zhou, K. Chandrasekaran, S. B. Chiah, Z. Zhu, C. Wei, S. Lin, G. Zhu, and G. H. Lim, "A compact model satisfying Gummel symmetry in higher order derivatives and applicable to asymmetric mosfets," *Electron Devices, IEEE Transactions on*, vol. 55, no. 2, pp. 624–631, Feb 2008.

[48] I. Hajj, "On device modeling for circuit simulation with application to carbon-nanotube and graphene nano-ribbon field-effect transistors," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, no. 99, pp. 495–499, Jan 2015.

[49] "Predictive technology model." [Online]. Available: http://ptm.asu.edu/

[50] C. C. McAndrew, "Statistical modeling for circuit simulation," in *Proc. IEEE International Symposium on Quality Electronic Design*, 2003, pp. 357–362.

[51] J. Zhang, N. Patil, A. Hazeghi, H. S. P. Wong, and S. Mitra, "Characterization and design of logic circuits in the presence of carbon nanotube density variations," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 30, no. 8, pp. 1103–1113, Aug 2011.

[52] N. Drego, A. Chandrakasan, and D. Boning, "An all-digital, highly scalable architecture for measurement of spatial variation in digital circuits," in *Solid-State Circuits Conference, 2008. A-SSCC '08. IEEE Asian*, Nov 2008, pp. 393–396.

[53] Q. Y. Tang, P. Friedberg, G. Cheng, and C. Spanos, "Circuit size optimization with multiple sources of variation and position dependent correlation," in *Proceedings of the SPIE advanced lithography*, vol. 6521, March 2007, pp. 1–11.

[54] C. Berger, Z. Song, X. Li, X. Wu, N. Brown, C. Naud, D. Mayou, T. Li, J. Hass, A. N. Marchenkov et al., "Electronic confinement and coherence in patterned epitaxial graphene," *Science*, vol. 312, no. 5777, pp. 1191–1196, 2006.

[55] C. Sire, F. Ardiaca, S. Lepilliet, J.-W. T. Seo, M. C. Hersam, G. Dambrine, H. Happy, and V. Derycke, "Flexible gigahertz transistors derived from solution-based single-layer graphene," *Nano Letters*, vol. 12, no. 3, pp. 1184–1188, 2012.

[56] G. Fiori and G. Iannaccone, "Code for the 3D simulation of nanoscale semiconductor devices, including drift-diffusion and ballistic transport in 1D and 2D subbands, and 3D tunneling," *Journal of Computational Electronics*, vol. 4, no. 1-2, pp. 63–66, 2005.

[57] X. Yang, G. Fiori, G. Iannaccone, and K. Mohanram, "Semi-analytical model for Schottky-barrier carbon nanotube and graphene nanoribbon transistors," in *Proceedings of the 20th symposium on Great Lakes Symposium on VLSI*. ACM, 2010, pp. 233–238.

[58] S. Sinha, G. Yeric, V. Chandra, B. Cline, and Y. Cao, "Exploring sub-20nm FinFET design with predictive technology models," in *Proceedings of the 49th Annual Design Automation Conference*. ACM, 2012, pp. 283–288.

[59] S. Vaziri, A. D. Smith, C. Henkel, M. Östling, M. C. Lemme, G. Lupina, G. Lippert, J. Dabrowski, and W. Mehr, "An integration approach for graphene double-gate transistors," in *Solid-State Device Research Conference (ESSDERC), 2012 Proceedings of the European*. IEEE, 2012, pp. 250–253.

[60] M. Y. Han, B. Özyilmaz, Y. Zhang, and P. Kim, "Energy band-gap engineering of graphene nanoribbons," *Physical Review Letters*, vol. 98, no. 20, p. 206805, 2007.

[61] D. Jiménez, "A current–voltage model for Schottky-barrier graphene-based transistors," *Nanotechnology*, vol. 19, no. 34, p. 345204, 2008.

[62] S. Reich, J. Maultzsch, C. Thomsen, and P. Ordejon, "Tight-binding description of graphene," *Physical Review B*, vol. 66, no. 3, p. 035412, 2002.

[63] J. Mintmire and C. White, "Universal density of states for carbon nanotubes," *Physical Review Letters*, vol. 81, no. 12, p. 2506, 1998.

[64] N. Frank and L. Young, "Transmission of electrons through potential barriers," *Physical Review*, vol. 38, no. 1, p. 80, 1931.

[65] M. Najari, S. Fregonese, C. Maneux, H. Mnif, N. Masmoudi, and T. Zimmer, "Schottky barrier carbon nanotube transistor: Compact modeling, scaling study, and circuit design applications," *Electron Devices, IEEE Transactions on*, vol. 58, no. 1, pp. 195–205, 2011.

[66] S. Datta, *Electronic Transport in Mesoscopic Systems*. Cambridge University Press, 1997.

[67] A. Betti, G. Fiori, and G. Iannaccone, "Atomistic investigation of low-field mobility in graphene nanoribbons," *Electron Devices, IEEE Transactions on*, vol. 58, no. 9, pp. 2824–2830, 2011.

[68] M. Poljak, T. Suligoj, and K. L. Wang, "Influence of substrate type and quality on carrier mobility in graphene nanoribbons," *Journal of Applied Physics*, vol. 114, no. 5, p. 053701, 2013.

[69] Z. Chen, J. Appenzeller, Y.-M. Lin, J. Sippel-Oakley, A. G. Rinzler, J. Tang, S. J. Wind, P. M. Solomon, and P. Avouris, "An integrated logic circuit assembled on a single carbon nanotube," *Science*, vol. 311, no. 5768, pp. 1735–1735, 2006.

[70] W. J. Yu, U. J. Kim, B. R. Kang, I. H. Lee, E.-H. Lee, and Y. H. Lee, "Adaptive logic circuits with doping-free ambipolar carbon nanotube transistors," *Nano Letters*, vol. 9, no. 4, pp. 1401–1405, 2009.

[71] Y. Yoon and J. Guo, "Effect of edge roughness in graphene nanoribbon transistors," *Applied Physics Letters*, vol. 91, no. 7, p. 073103, 2007.

[72] M. Luisier and G. Klimeck, "Performance analysis of statistical samples of graphene nanoribbon tunneling transistors with line edge roughness," *Applied Physics Letters*, vol. 94, no. 22, p. 223505, 2009.

[73] L. Liao, Y.-C. Lin, M. Bao, R. Cheng, J. Bai, Y. Liu, Y. Qu, K. L. Wang, Y. Huang, and X. Duan, "High-speed graphene transistors with a self-aligned nanowire gate," *Nature*, vol. 467, no. 7313, pp. 305–308, 2010.

[74] H.-Y. Chen and J. Appenzeller, "Complementary-type graphene inverters operating at room-temperature," in *69th Device Research Conference*, 2011.

[75] R. Sordan, F. Traversi, and V. Russo, "Logic gates with a single graphene transistor," *Applied Physics Letters*, vol. 94, no. 7, p. 073305, 2009.

[76] M. Pourfath, A. Gehring, E. Ungersboeck, H. Kosina, S. Selberherr, B.-H. Cheong, and W.-J. Park, "Separated carrier injection control in carbon nanotube field-effect transistors," *Journal of Applied Physics*, vol. 97, no. 10, p. 106103, 2005.

[77] A. Y. Goharrizi, M. Pourfath, M. Fathipour, and H. Kosina, "Device performance of graphene nanoribbon field-effect transistors in the presence of line-edge roughness," *Electron Devices, IEEE Transactions on*, vol. 59, no. 12, pp. 3527–3532, 2012.

[78] S. Heinze, J. Tersoff, and P. Avouris, "Electrostatic engineering of nanotube transistors for improved performance," *Applied Physics Letters*, vol. 83, no. 24, pp. 5038–5040, 2003.

[79] X. Song, J. Hu, and H. Zeng, "Two-dimensional semiconductors: Recent progress and future perspectives," *Journal of Materials Chemistry C*, vol. 1, no. 17, pp. 2952–2969, 2013.

[80] T. Fang, A. Konar, H. Xing, and D. Jena, "Mobility in semiconducting graphene nanoribbons: Phonon, impurity, and edge roughness scattering," *Physical Review B*, vol. 78, no. 20, p. 205403, 2008.

[81] K. Alam and R. K. Lake, "Monolayer MoS_2 transistors beyond the technology road map," *Electron Devices, IEEE Transactions on*, vol. 59, no. 12, pp. 3250–3254, 2012.

[82] J. Kang, D. Sarkar, W. Liu, D. Jena, and K. Banerjee, "A computational study of metal-contacts to beyond-graphene 2D semiconductor materials," in *Electron Devices Meeting (IEDM), 2012 IEEE International*. IEEE, 2012, pp. 17.4.1–17.4.4.

[83] Y. Yoon, K. Ganapathi, and S. Salahuddin, "How good can monolayer MoS_2 transistors be?" *Nano Letters*, vol. 11, no. 9, pp. 3768–3773, 2011.

[84] H.-Y. Chang, S. Yang, J. Lee, L. Tao, W.-S. Hwang, D. Jena, N. Lu, and D. Akinwande, "High-performance, highly bendable MoS_2 transistors with high-k dielectrics for flexible low-power systems," *ACS Nano*, vol. 7, no. 6, pp. 5446–5452, 2013.

[85] J. Pu, Y. Yomogida, K.-K. Liu, L.-J. Li, Y. Iwasa, and T. Takenobu, "Highly flexible MoS_2 thin-film transistors with ion gel dielectrics," *Nano Letters*, vol. 12, no. 8, pp. 4013–4017, 2012.

[86] J. Yoon, W. Park, G.-Y. Bae, Y. Kim, H. S. Jang, Y. Hyun, S. K. Lim, Y. H. Kahng, W.-K. Hong, B. H. Lee et al., "Highly flexible and transparent multilayer MoS_2 transistors with graphene electrodes," *Small*, vol. 9, no. 19, pp. 3295–3300, 2013.

[87] G.-H. Lee, Y.-J. Yu, X. Cui, N. Petrone, C.-H. Lee, M. S. Choi, D.-Y. Lee, C. Lee, W. J. Yoo, K. Watanabe et al., "Flexible and transparent MoS_2 field-effect transistors on hexagonal boron nitride-graphene heterostructures," *ACS Nano*, vol. 7, no. 9, pp. 7931–7936, 2013.

[88] Q. He, Z. Zeng, Z. Yin, H. Li, S. Wu, X. Huang, and H. Zhang, "Fabrication of flexible MoS_2 thin-film transistor arrays for practical gas-sensing applications," *Small*, vol. 8, no. 19, pp. 2994–2999, 2012.

[89] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis, "Single-layer MoS_2 transistors," *Nature Nanotechnology*, vol. 6, no. 3, pp. 147–150, 2011.

[90] S. Das, H.-Y. Chen, A. V. Penumatcha, and J. Appenzeller, "High performance multilayer MoS_2 transistors with scandium contacts," *Nano Letters*, vol. 13, no. 1, pp. 100–105, 2012.

[91] W. Zhu, T. Low, Y.-H. Lee, H. Wang, D. B. Farmer, J. Kong, F. Xia, and P. Avouris, "Electronic transport and device prospects of monolayer molybdenum disulphide grown by chemical vapour deposition," *Nature Communications*, vol. 5, 2014.

[92] H. Fang, S. Chuang, T. C. Chang, K. Takei, T. Takahashi, and A. Javey, "High-performance single layered WSe_2 p-fets with chemically doped contacts," *Nano Letters*, vol. 12, no. 7, pp. 3788–3792, 2012.

[93] H. Wang, L. Yu, Y.-H. Lee, Y. Shi, A. Hsu, M. L. Chin, L.-J. Li, M. Dubey, J. Kong, and T. Palacios, "Integrated circuits based on bilayer MoS_2 transistors," *Nano Letters*, vol. 12, no. 9, pp. 4674–4680, 2012.

[94] B. Radisavljevic, M. B. Whitwick, and A. Kis, "Integrated circuits and logic operations based on single-layer MoS_2," *ACS Nano*, vol. 5, no. 12, pp. 9934–9938, 2011.

[95] J. Chang, L. F. Register, and S. K. Banerjee, "Atomistic full-band simulations of monolayer MoS_2 transistors," *Applied Physics Letters*, vol. 103, no. 22, p. 223509, 2013.

[96] J. Chang, L. F. Register, and S. K. Banerjee, "Ballistic performance comparison of monolayer transition metal dichalcogenide MoS_2 (M= Mo, W; X= S, Se, Te) metal-oxide-semiconductor field effect transistors," *Journal of Applied Physics*, vol. 115, no. 8, p. 084506, 2014.

[97] M. Salmani-Jelodar, Y. Tan, and G. Klimeck, "Single layer MoS_2 band structure and transport," in *Semiconductor Device Research Symposium (ISDRS), 2011 International*. IEEE, 2011, pp. 1–2.

[98] D. Jiménez, "Drift-diffusion model for single layer transition metal dichalcogenide field-effect transistors," *Applied Physics Letters*, vol. 101, no. 24, p. 243501, 2012.

[99] W. S. Yun, S. Han, S. C. Hong, I. G. Kim, and J. Lee, "Thickness and strain effects on electronic structures of transition metal dichalcogenides: 2H-MoS_2 semiconductors (M= Mo, W; X= S, Se, Te)," *Physical Review B*, vol. 85, no. 3, p. 033305, 2012.

[100] K. Novoselov, D. Jiang, F. Schedin, T. Booth, V. Khotkevich, S. Morozov, and A. Geim, "Two-dimensional atomic crystals," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 30, pp. 10 451–10 453, 2005.

[101] H. J. Conley, B. Wang, J. I. Ziegler, R. F. Haglund Jr, S. T. Pantelides, and K. I. Bolotin, "Bandgap engineering of strained monolayer and bilayer MoS_2," *Nano Letters*, vol. 13, no. 8, pp. 3626–3630, 2013.

[102] P. Johari and V. B. Shenoy, "Tuning the electronic properties of semiconducting transition metal dichalcogenides by applying mechanical strains," *ACS Nano*, vol. 6, no. 6, pp. 5449–5456, 2012.

[103] K. He, C. Poole, K. F. Mak, and J. Shan, "Experimental demonstration of continuous electronic structure tuning via strain in atomically thin MoS_2," *Nano Letters*, vol. 13, no. 6, pp. 2931–2936, 2013.

[104] H. Liu, A. T. Neal, and P. D. Ye, "Channel length scaling of MoS_2 MOSFETs," *ACS Nano*, vol. 6, no. 10, pp. 8563–8569, 2012.

[105] H. Liu, M. Si, S. Najmaei, A. T. Neal, Y. Du, P. M. Ajayan, J. Lou, and P. D. Ye, "Statistical study of deep submicron dual-gated field-effect transistors on monolayer chemical vapor deposition molybdenum disulfide films," *Nano Letters*, vol. 13, no. 6, pp. 2640–2646, 2013.

[106] Y. Cao, *Predictive Technology Model for Robust Nanoelectronic Design*. Springer Science & Business Media, 2011.

[107] T. Sekitani, U. Zschieschang, H. Klauk, and T. Someya, "Flexible organic transistors and circuits with extreme bending stability," *Nature Materials*, vol. 9, no. 12, pp. 1015–1022, 2010.

[108] W. Cao, J. Kang, W. Liu, and K. Banerjee, "A compact current-voltage model for 2D semiconductor based field-effect transistors considering interface traps, mobility degradation, and inefficient doping effect," *Electron Devices, IEEE Transactions on*, vol. 61, no. 12, pp. 4282–4290, Dec 2014.

[109] W. Liu, J. Kang, W. Cao, D. Sarkar, Y. Khatami, D. Jena, and K. Banerjee, "High-performance few-layer-MoS_2 field-effect-transistor with record low contact-resistance," in *Electron Devices Meeting (IEDM), 2013 IEEE International*. IEEE, 2013, pp. 19–4.

[110] M. S. Fuhrer and J. Hone, "Measurement of mobility in dual-gated MoS_2 transistors," *Nature Nanotechnology*, vol. 8, no. 3, pp. 146–147, 2013.

[111] R. Kim and M. Lundstrom, "Notes on Fermi-Dirac integrals," *arXiv preprint arXiv:0811.0116*, 2008.

[112] T. N. Nguyen, "Small-geometry MOS transistors: Physics and modeling of surface- and buried-channel MOSFETs," Ph.D. dissertation, Stanford University, Palo Alto, CA.

[113] K. Natori, "Ballistic/quasi-ballistic transport in nanoscale transistor," *Applied Surface Science*, vol. 254, no. 19, pp. 6194–6198, 2008.

[114] R. Clerc and G. Ghibaudo, "Analytical models and electrical characterisation of advanced MOSFETs in the quasi-ballistic regime," *International Journal of High Speed Electronics and Systems*, vol. 22, no. 01, p. 1350002, 2013.

[115] P. J. Price, "Monte Carlo calculation of electron transport in solids," *Semiconductors and Semimetals*, vol. 14, pp. 249–308, 1979.

[116] M. Lundstrom and Z. Ren, "Essential physics of carrier transport in nanoscale MOSFETs," *IEEE Transactions on Electron Devices*, vol. 49, no. 1, pp. 133–141, 2002.

[117] V. Litovski and M. Zwolinski, *VLSI Circuit Simulation and Optimization*. Springer Science & Business Media, 1997.