

Graphene Tunneling FET and its Applications in Low-power Circuit Design

Xuebei Yang[†] Jyotsna Chauhan[‡] Jing Guo[‡] Kartik Mohanram[†]

[†]Department of Electrical and Computer Engineering, Rice University, Houston

[‡]Department of Electrical and Computer Engineering, University of Florida, Gainesville

xbyang@rice.edu jyotsna.chauhan@ufl.edu guoj@ufl.edu kmram@rice.edu

Abstract

Graphene nanoribbon tunneling FETs (GNR TFETs) are promising devices for post-CMOS low-power applications because of the low subthreshold swing, high I_{on}/I_{off} , and potential for large scale processing and fabrication. This paper combines atomistic quantum transport modeling with circuit simulation to explore GNR TFET circuits for low-power applications. A quantitative study of the effects of variations on the performance and reliability of GNR TFET circuits is also presented. Simulation results indicate that GNR TFET circuits are extremely competitive in performance in comparison to conventional CMOS circuits at comparable operating points, with static power consumption that is lower by 8–9 orders of magnitude. It is also observed that GNR TFET circuits are susceptible to parameter variations, motivating engineering and design challenges to be addressed by the device and CAD communities.

Categories and Subject Descriptors: B.7.1 [Integrated circuits]: Types and Design Styles—Advanced technologies

General Terms: Design, Performance, Reliability

Keywords: Graphene nanoribbons, tunneling FETs, low-power

1. Introduction

Since its discovery in 2004, graphene, which is a monolayer of carbon atoms packed into a two-dimensional honeycomb lattice, has attracted strong interest as an alternative device technology for future nanoelectronics [1–3]. Graphene’s high electron and hole mobility, combined with high mechanical and thermal stability and high resistance to electro-migration make it an excellent candidate material for post-silicon electronics. The potential to produce wafer-scale graphene films with full planar processing for devices promises high integration potential with CMOS processes, which is a significant advantage over carbon nanotubes [4].

Although two-dimensional graphene is a zero band-gap semiconductor and not preferred for logic design, it has been demonstrated that a band-gap can be opened up by using graphene nanoribbons (GNRs) due to quantum confinement in the width direction [4]. In the past, most graphene FETs have been fabricated using intrinsic nanometer-wide GNRs as channel material with metallic contacts [5–7]. More recently, however, FETs based on the principles of tunneling (TFETs henceforth) have been studied and attracted

strong interest [8–12]. The commonly proposed graphene TFET structure uses an intrinsic GNR as the channel, with p-doped GNR as the source and n-doped GNR as the drain. The operation of these p-i-n GNR TFETs is mainly based on tunneling current, which is more sensitive to terminal biases, instead of thermionic current. As a result, GNR TFETs are projected to exhibit extremely low subthreshold swing as compared to conventional CMOS, without a significant reduction in speed of operation. For example, in [8], it is shown that a sub-threshold swing of 0.19 mV/dec that is over 2 orders of magnitude lower than conventional CMOS can be achieved using GNR TFETs. A lower sub-threshold swing results in a higher I_{on}/I_{off} ratio and lower leakage power, motivating strong interest in GNR TFETs for low-power applications.

This paper studies and presents results on the suitability of GNR TFETs for low-power applications. It combines atomistic quantum transport modeling in intrinsic p-i-n GNR TFETs with a circuit simulator that includes parasitics and non-idealities that are necessary to capture extrinsic effects in fabricated GNR TFETs. Intrinsic TFETs are simulated by self-consistently solving an atomistic quantum transport equation based on the non-equilibrium Green’s function (NEGF) formalism with the 3D Poisson’s equation. These rigorous simulations provide $I-V$ and $Q-V$ data for intrinsic GNR TFETs, which are integrated into a circuit-level simulation framework based on lookup tables for technology exploration of GNR TFET circuits. The simulator is used to study the delay, power, and noise margins of representative GNR TFET circuits including inverters and ring oscillators. Results are compared to scaled CMOS at the 22, 32, and 45 nm nodes using the predictive technology model [13]. At comparable operating points, the frequency of GNR TFETs is 0.001–0.1X the frequency of these CMOS nodes, with an 8–9 orders of magnitude improvement in static power. This demonstrates the huge advantage and potential of GNR TFETs over scaled CMOS in low-power applications.

We further evaluate the effect of parametric variations in GNR TFET fabrication on performance, power, and reliability in this paper. Due to the atomically thin and nanometer-wide geometries of GNRs, variabilities in device parameters are projected to impact circuit performance and reliability. Our atomistic NEGF simulation of variations in device parameters has identified two important sources of parametric variations in GNR TFETs: variations in GNR width and variations in drain/source doping level. Variation in GNR width affects the energy band-gap of GNRs, while variation in source/drain doping level affects the energy band structure near the contacts. Both variations result in a change of the I_{on}/I_{off} ratio, which in turn impacts circuit delay, power, and noise margins. Our simulations of basic logic structures indicates that GNR width variations have the largest impact on circuit delay, static power, and noise margin. However, even for the worst case GNR width variations studied in this paper, the static power is still significantly lower than the scaled CMOS nodes. On the other hand, variation in

This research was supported by NSF grants CCF-0916636 and CCF-0916683.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GLSVLSI’10, May 16–18, 2010, Providence, Rhode Island, USA.

Copyright 2010 ACM 978-1-4503-0012-4/10/06 ...\$10.00.

source doping level has a very small impact on circuit power, but it has a comparably larger influence on the delay and noise margin. Note that variation in drain doping level has a negligible overall impact on circuit behavior, because n-type and p-type TFETs are based on electron and hole current respectively, which drain doping level has little control over. These results show that even in the worst-case, graphene-based TFETs exhibit significant potential for low-power applications in comparison to scaled CMOS.

This paper is organized as follows. Section 2 describes the background and the operating principles of GNR TFETs. Section 3 describes technology exploration for GNR TFET circuits, with comparisons to scaled CMOS. Sections 4 and 5 discuss the effects of parametric variations on GNR TFETs and GNR TFET circuits, respectively. Section 6 provides background on the quantum device simulator and the circuit simulator. Section 7 is a conclusion.

2. TFET background and operating principles

In this section, we describe the device structure of the GNR TFET considered in this paper and its basic operating principles. The details of our quantum atomistic device simulator are provided in Section 6.1 at the end of this paper.

Figure 1(a) presents the structure of the GNR TFET considered in this paper. A 50nm-long armchair-edge GNR (A-GNR) is used as the channel material, and the GNR width is assumed to be $N=13$, where N denotes the GNR index [14]. Double gate geometry is implemented using HfO_2 as the gate insulator, with a thickness of 1.5nm and dielectric constant of 16. The doping level of source and drain is set to be 0.01 dopant per atom, or 3.81×10^{15} dopant/cm². Notice that these values are nominal values, and we will explore the effects of variations in these parameters.

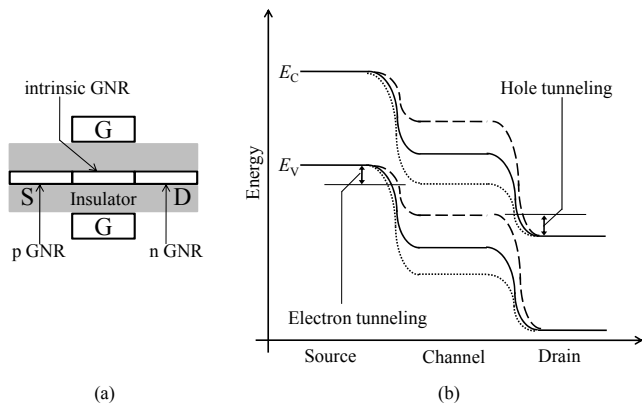


Figure 1: (a) The device structure of the GNR TFET (b) The energy band diagram of the TFET under $V_G = 0$ (solid line), high V_G (dot line), and low V_G (dashed line)

The operating principles of the GNR TFET can be understood by looking at its energy band diagram, presented in Figure 1(b). Since source and drain are doped p-type and n-type respectively, the thermionic current is negligible in GNR TFETs. Instead, the operation of GNR TFETs depends mainly on band-to-band tunneling current of both electrons and holes. When drain voltage is added but gate voltage remains zero, the conduction band E_C in the channel is higher than the valence band E_V at the source side, and E_V in the channel is lower than E_C at the drain side, as shown in Figure 1(b) for the solid line. Therefore, both electrons and holes find it difficult to tunnel into the channel. If the gate voltage is increased, however, to a level where E_C in the channel becomes lower than E_V at the source side, as shown in Figure 1(b) with the dotted line,

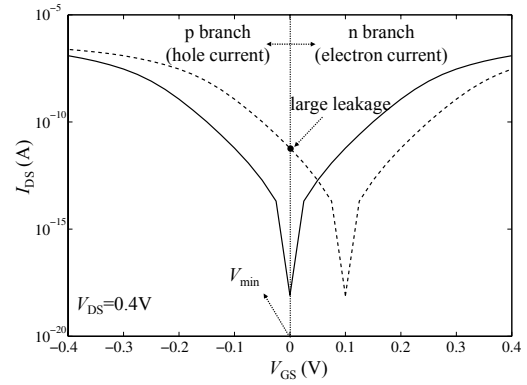


Figure 2: I - V characteristics for $N = 13$ GNR TFET. V_{\min} is 0 and 0.1V for solid and dashed line, respectively.

electrons can tunnel from E_V at source to E_C inside the channel, resulting in band-to-band tunneling and a significant increase in the current. Similarly, if gate voltage is reduced to a level where E_V in the channel is higher than E_C in the drain, illustrated by the dashed line in Figure 1(b), hole-based band-to-band tunneling will occur at the drain side, significantly increasing the current.

2.1 Reducing ambipolarity

The I - V characteristics for the $N=13$ GNR TFET under $V_{DS} = 0.4$ is shown in Figure 2. Notice that the current is exponentially and linearly proportional to the gate voltage at low $|V_G|$ and high $|V_G|$, respectively. Ambipolar conduction is clearly shown, where hole current dominates the left branch of the current curve while electron current dominates the right branch. The minimum conduction point V_{\min} of the ambipolar curve can be shifted by adjusting the work-function of the gate electrode, as shown in Figure 2 for $V_{\min} = 0$ (solid line) and $V_{\min} > 0$ (dashed line). Notice that if $V_{\min} \neq 0$, i.e., if $V_{\min} > 0$ for example, the ‘off’ current of the GNR TFET at $V_G = 0$ increases drastically, leading to large leakage current. Techniques based on gate underlap and asymmetric doping explained below have been proposed to reduce ambipolarity.

Gate underlap: Gate underlap at either end of the channel will affect the energy band-gap and hence the current behavior. For example, if the gate underlap occurs at drain end of the channel, as shown in Figure 3, it results in a linear drop of the potential in the ungated part. This makes the tunneling barrier at the drain contact larger, significantly suppressing the hole tunneling current. However, since the energy band structure at source side is not influenced, the electron tunneling current is unaffected. Similarly, if gate underlap occurs at the source side, n-type conduction decreases while p-type conduction remains unaffected. Note that in Figure 3, the current remains nearly constant near $V_{GS} = 0$ for underlap of 10nm and 15nm, respectively. This is due to the direct source-drain tunneling current, which is injected from the source and tunnels through the band-gap region in the gated portion of the channel, and finally reaches the drain contact. This can be regarded as the minimum possible current of the GNR TFETs.

Asymmetric doping: The doping level influences the energy band structure near the contacts, which determines the tunneling probability, and therefore the tunneling current for electrons and holes. If the doping levels at both source and drain ends of the channel are the same, the p-type and n-type current will be symmetric. However, if the drain doping level is reduced, the width of the band-to-band tunneling barrier for holes at the drain contact will increase due to a larger electrostatic screening length for a lower doping density. Therefore, the p-type conduction branch is significantly

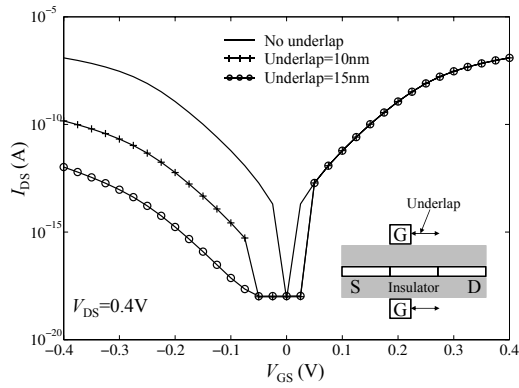


Figure 3: I - V characteristics for drain-side gate underlap.

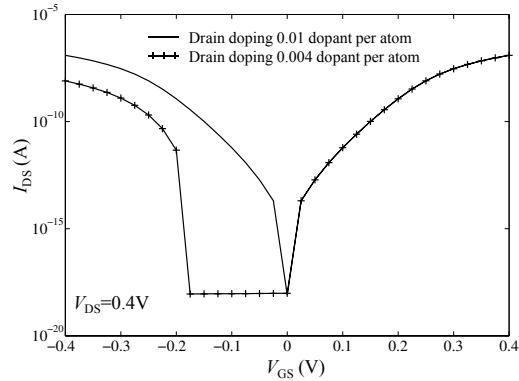


Figure 4: I - V characteristics for drain doping variations. Doping level at source is fixed at 0.01 dopant per atom.

suppressed because tunneling current decreases drastically with an increasing barrier width. In contrast, because the n-type conduction is controlled by the electron band-to-band tunneling from the source to the channel which remains unaffected, it is insensitive to the drain-doping density, as presented in Figure 4. Along the same lines, reduced source doping level serves to suppress n-type conduction while leaving p-type conduction unaffected.

These techniques have been demonstrated to be very effective in reducing ambipolarity [12]. In this paper, however, we assume for simplicity that the work-function is chosen such that the minimal conduction point $V_{\min} = 0$, as shown in Figure 2 for the solid line. Therefore n-type and p-type FETs utilize the right and left branch, respectively, and the leakage can be reduced without the help of ambipolarity-reduction techniques. Although variations in work-function of gate electrode will change the minimum conduction point and necessitate the usage of the above techniques, it is beyond the scope of this paper and not discussed further.

3. GNR TFET circuits

GNR TFETs have low subthreshold swing and low ‘off’ current, making them very suitable for subthreshold low-power applications. In this section, we perform an extensive study on the circuit performance of GNR TFETs for low supply voltage. A simulator based on table lookup techniques is implemented to simulate circuits built with GNR TFETs, with details of the circuit simulation model described in Section 6.2. We choose a 15-stage ring oscillator where each inverter drives a fanout-of-four load as the representative circuit for this study. Figures of merit considered include frequency, static power, dynamic switching energy, and noise margin, and the trade-offs between these metrics are explored for

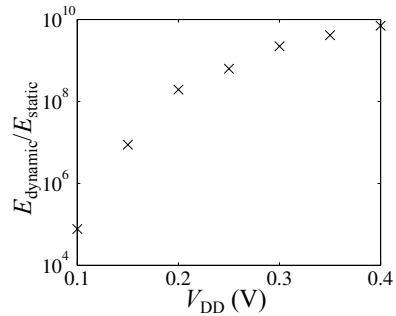


Figure 5: $E_{\text{dynamic}}/E_{\text{static}}$ for different V_{DD} .

different values of V_{DD} for GNR TFET circuits. For comparison, we also perform simulations on the 22nm, 32nm, and 45nm CMOS nodes using the predictive technology model [13] under comparable operating conditions, and report the results in Table 1.

3.1 TFET circuit performance

In the subthreshold domain, the total energy consumption per operation is an important metric because it captures the energy-saving performance of a circuit. The total energy consumption per operation $E_{\text{total}} \approx \frac{1}{\alpha} E_{\text{static}} + E_{\text{dynamic}}$, where α is the activity factor, E_{static} is the static leakage energy, and E_{dynamic} is the switching energy. E_{dynamic} can be roughly expressed as $C_L V_{\text{DD}}^2$, where C_L is the load capacitance. As shown in Table 1, E_{dynamic} continuously decreases with supply voltage V_{DD} , from 0.882fJ at $V_{\text{DD}} = 0.4\text{V}$ to 0.052fJ at $V_{\text{DD}} = 0.1\text{V}$. $E_{\text{static}} = P_{\text{static}} \cdot t$, where P_{static} is the static leakage power and t is the time when the gate is static. P_{static} for GNR TFETs also decreases continuously with V_{DD} , from 19.4 aW at $V_{\text{DD}} = 0.4\text{V}$ to 3.49aW at $V_{\text{DD}} = 0.1\text{V}$. However, since the frequency of the circuit decreases much faster with V_{DD} , as shown later, the leakage time, which is proportional to the delay of the circuit increases and results in an increasing E_{static} . In Figure 5, we plot the $E_{\text{dynamic}}/E_{\text{static}}$ ratio for different V_{DD} . It is clear that although the ratio continuously decreases as V_{DD} scales down, it is still as high as nearly 10^5 at $V_{\text{DD}} = 0.1\text{V}$. This means that as long as the activity factor is not less than 10^{-5} , which is a very low value in practice, dynamic energy will always dominate the total energy consumption in GNR TFETs.

While downscaling V_{DD} serves to lower the energy consumption, however, it comes at the cost of reduction in frequency. As shown in Table 1, the frequency of the 15-stage ring oscillator continues to decrease as the supply voltage is reduced, from 154.6MHz at $V_{\text{DD}} = 0.4\text{V}$ to 5.2kHz at $V_{\text{DD}} = 0.1\text{V}$. Further, owing to the exponential relationship between current and V_G at low V_G , the frequency drops much faster as V_{DD} is scaled down.

The other shortcoming of downscaling supply voltage is lower reliability, which can be measured by calculating the static noise margin (SNM) for an inverter. As shown in Table 1, the SNM of the inverter decreases as supply voltage is down-scaled, from 0.18V, or 45% V_{DD} at $V_{\text{DD}} = 0.4\text{V}$ to 0.036V, or 36% V_{DD} at $V_{\text{DD}} = 0.1\text{V}$.

3.2 Comparison to scaled CMOS

Table 1 compares the operating frequency, energy, and SNM of the 15-stage ring oscillator for GNR TFETs and scaled CMOS nodes (simulated using the PTM model [13]).

The most striking difference between the scaled CMOS nodes and GNR TFETs is the extremely low static power consumption of GNR TFETs, which is about nine orders of magnitude lower than scaled CMOS at all voltages. Under high supply voltage, $V_{\text{DD}} = 0.4\text{V}$ for example, low static power of GNR TFET circuits does not result in a significant reduction in total energy con-

Table 1: Delay, static power, dynamic energy and static noise margin (SNM) for GNR TFETs and scaled CMOS.

V_{DD} (V)	Frequency (MHz)				Static power P_{static} (W)				Dynamic energy $E_{dynamic}$ (fJ)				SNM (V)			
	TFET	45nm	32nm	22nm	TFET	45nm	32nm	22nm	TFET	45nm	32nm	22nm	TFET	45nm	32nm	22nm
0.4	154.6	1203	1435	1744	1.94×10^{-17}	2.61×10^{-8}	2.33×10^{-8}	2.03×10^{-8}	0.882	3.18	2.08	1.15	0.18	0.177	0.165	0.156
0.35	98.18	775.2	862.2	1030	1.56×10^{-17}	1.92×10^{-8}	1.62×10^{-8}	1.36×10^{-8}	0.657	2.34	1.58	0.846	0.154	0.153	0.142	0.135
0.3	50.63	412.3	443.8	513.1	1.08×10^{-17}	1.38×10^{-8}	1.09×10^{-8}	8.72×10^{-9}	0.474	1.64	1.05	0.612	0.134	0.128	0.121	0.112
0.25	17.58	188	194.1	219.7	8.97×10^{-18}	9.69×10^{-9}	7.16×10^{-9}	5.45×10^{-9}	0.321	0.989	0.743	0.414	0.107	0.104	0.098	0.091
0.2	6.86	76.6	79.7	85.5	7.14×10^{-18}	6.63×10^{-9}	4.79×10^{-9}	3.33×10^{-9}	0.203	0.795	0.503	0.279	0.086	0.081	0.076	0.070
0.15	0.4156	28.6	30.3	31.0	5.33×10^{-18}	2.99×10^{-9}	2.63×10^{-9}	1.92×10^{-9}	0.113	0.520	0.309	0.181	0.06	0.055	0.053	0.049
0.1	0.0052	11.2	11.7	12.2	3.49×10^{-18}	2.37×10^{-9}	1.55×10^{-9}	1.01×10^{-9}	0.0519	0.341	0.167	0.122	0.036	0.032	0.030	0.027

¹Note that static energy can be derived by multiplying static power with delay.

sumption per operation compared to CMOS nodes since dynamic energy dominates at these voltage nodes for both CMOS and GNR TFETs, and GNR TFETs and CMOS circuits consume comparable dynamic energy. However, as the supply voltage is scaled down, whereas static energy is still negligible for GNR TFETs at moderate activity factor, it becomes comparable to dynamic energy for scaled CMOS nodes, such that the benefit brought by low static energy of GNR TFETs becomes more significant. For example, at $V_{DD} = 0.4V$, the GNR TFET circuit has 23% less energy consumption per operation compared to 22nm CMOS. However, at $V_{DD} = 0.1V$ and activity factor $\alpha = 0.1$, 22nm CMOS will consume $18.3 \times$ more energy than TFET circuits for each operation. This difference will be even higher with a lower activity factor. Besides the low energy consumption at low supply voltage, TFET circuits, compared to scaled CMOS nodes, also provide a better noise margin by 1–25%, which is crucial in subthreshold circuits. In terms of operating frequency, CMOS circuits outperform GNR TFET circuits. However, in subthreshold low-power applications, the frequency of operation is usually low and of the order of 1–100kHz for environmental and biological applications. In such domains, these results indicate that GNR TFET circuits are a very strong candidate for low-power applications over scaled CMOS.

4. Variability in GNR TFETs

Variability is expected to play an important role in graphene electronics in practice. Variability can come from the difficulty to control GNR width, doping levels in source and drain, insulator thickness and so on. Our atomistic NEGF simulation of a wide variety of variability mechanisms have identified the important role of the GNR width variation and contact doping level variation of GNR TFETs [12], which are the subject of this study. Other defect and variability mechanisms exist and should be explored in future studies, but we expect the effects are qualitatively similar and can be explored by readily extending the current simulation framework.

4.1 GNR width variation

The band-gap of the GNR is determined by the GNR width. Because device characteristics are very sensitive to the band-gap of channel material, GNR width is critical to GNR TFET performance. GNR width is proportional to the GNR index. GNRs with an index of $N=3q$ and $N=(3q+1)$ are semiconducting GNRs and may be used as FET channels. Our atomistic NEGF simulation has indicated that $N=3q$ GNRs have a much smaller subthreshold swing than $N=(3q+1)$ GNRs, so only GNRs with index values of 10, 13 and 16 were selected to study the effect of variations in width in this paper. Starting with the minimum GNR index of $N=10$, which has a width of 1.6nm, the index is increased in steps of 3, or equivalently, by an incremental width of 3.7\AA .

Figure 6 illustrates how variability in GNR width affects device

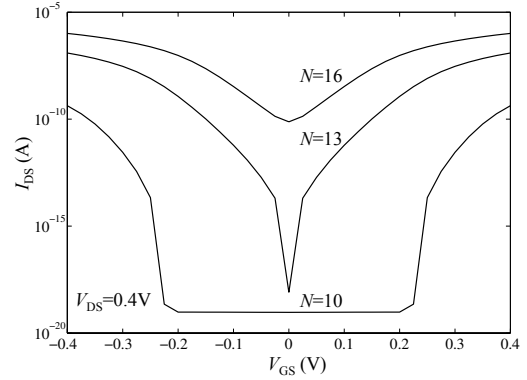


Figure 6: I - V characteristics for different GNR widths

I_{DS} - V_{GS} characteristics. For $N=(3q+1)$ GNRs, the band-gap is in general inversely proportional to the width of the GNR. Hence, as the width of the GNR increases, the band-gap decreases, allowing more electrons to enter the conduction band and more holes to enter the valence band, increasing both the ‘on’ and ‘off’ current. Note that the current almost remains the same from $V_{GS} = -0.2V$ to $V_{GS} = 0.2V$ for the $N=10$ GNR, which is due to the direct source-drain tunneling current discussed before.

4.2 Doping variation

So far we have assumed that the doping levels at both source and drain can be precisely controlled, and in Section 2 we have described the usage of asymmetric doping as a method to suppress ambipolar conduction. However, under process variations, it is very difficult to have both contacts accurately doped to a desired level, and the variation of doping levels may result in performance degradation. As presented in Section 2, the doping level of source and drain have significant influences on electron and hole current respectively, so if a TFET used as an n-type transistor suffers from source doping variations, or a TFET used as a p-type transistor suffers from drain doping variations, the current behavior will be severely impacted. Note that for GNR TFETs used as p-type transistors, the drain is usually considered to be the p-doped contact, i.e., the source of the n-type transistors. Based upon this convention, we determine that doping variation at source side will influence both the p-type and n-type transistors’ behavior, whereas drain side doping variations have negligible effect for the TFETs considered in this paper with minimum conduction point $V_{min} = 0$. In our simulations, we fix the doping level at the drain to be 0.01 dopant per atom and vary that at the source to be 0.01 or 0.004 dopant per atom, and the results for the doping variations in n-doped region is shown in Figure 4.

5. Variability in GNR TFET circuits

In this section we study the effect of variations in GNR width and doping levels on the inverter delay, power, and noise robustness. The operating supply voltage $V_{DD} = 0.4V$ is assumed to simulate all the GNR TFET circuits in this section, and the simulations were performed for an inverter with a fanout-of-4 load. Details on how we incorporate variations into our circuit simulation model can be found in Section 6.2.

5.1 Width variation

Table 2 shows the delay, power, and noise margin due to independent variations in GNR widths in both the n-type and p-type TFETs of an inverter. The delay of an inverter driving a fanout-of-4 load, with nominal n-type and p-type TFETs ($N=13$ GNR channel) is 216ps. As GNR width of both n-type and p-type TFETs decreases to $N=10$, the delay increases to 19.6ns, showing 2 orders of magnitude increase. On the other hand, if GNR width of all TFETs increases to $N=16$, the delay decreases to 62.4ps. Note in Figure 6 that as the GNR gets wider, the rate of increase of the ‘on’ current becomes slower. Therefore, downscaling the width of GNR will have a larger impact on the circuit delay.

The static power of the inverter is also significantly impacted by the variations in GNR width, while the dynamic power is relatively less affected. For a nominal inverter where all the GNRs are $N=13$, the static power is 1.29aW. When GNR width decreases to $N=10$ and increases to $N=16$, the static power decreases by 1 order of magnitude and increases by 8 orders of magnitude, respectively. This is due to the significant impact of GNR width on the ‘off’ current of the TFET. Within this table, the static power exhibits the largest influence due to the variation of GNR width.

Finally, it is observed that the noise margin of the inverter is also impacted by an order of magnitude in the worst case. In contrast to the delay and power figures of merit, the noise margin is not greatly affected when the n-type and p-type TFETs of an inverter have the same widths. For instance, the static noise margin almost remains the same for $N=10$, $N=13$ and $N=16$ GNR TFETs, with a difference of only 4.5%. However, when n-type and p-type TFETs have different widths, the influence becomes significantly larger and reaches the worst-case deviation when there is maximum mismatch of $N=10$ and $N=16$. In this case, the ‘off’ current of $N=16$ TFET is only one order of magnitude lower than the ‘on’ current of $N=10$ TFET, therefore the static noise margin is significantly lowered to 8mV from the nominal value of 180mV.

5.2 Doping variation

Table 3 shows the delay, power, and noise margin due to doping variations that affect the GNRs in either the n-type and p-type TFET of an inverter. Note that as discussed in Section 4, doping variations at drain side will not affect the performance of p-type and n-type TFETs, therefore only variations at source side is considered here. In our simulation, the drain doping level is fixed at 0.01 dopant per atom, and the source doping level is varied between 0.004 and 0.01 dopant per atom.

Inverter delay is degraded by one order of magnitude when the source doping level of all TFETs is changed from 0.01 dopant per atom to 0.004 dopant per atom, due to the altered current behavior resulting from the decrease in doping. For the same reason, the static power increases by 20%, and the dynamic energy increases by 6.9%.

The noise margin is significantly affected by the doping variations. Even for the same doping variations for n-type and p-type TFETs, the SNM is reduced to 21 mV, as compared to the nominal

case of 180mV. If only p-type or n-type TFETs suffer from doping variations, the SNM is further reduced to 18mV.

6. Quantum and circuit simulation

In this section, we first provide the details of our quantum-based device simulator. Next the circuit simulation model for the TFETs and the method to incorporate variations is summarized.

6.1 Quantum simulation

As devices scale with technology, the widely used compact modeling and simulation approach is unable to capture several important features of device physics including quantum effects. In this paper we implement a powerful quantum transport simulation framework based on the non-equilibrium Green’s function (NEGF) formalism, which provides an ideal approach for bottom-up device modeling and simulation [15] for the following reasons: (1) atomistic descriptions of devices can be readily implemented, (2) open boundaries can be rigorously treated, and (3) multi-phenomena (e.g., inelastic scattering and light emission) can be modeled.

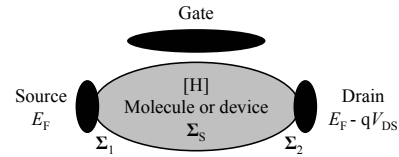


Figure 7: NEGF formalism for a generic transistor

Figure 7 summarizes the procedure to apply the NEGF approach to a generic transistor. The transistor channel, which can be a piece of silicon, a GNR, a nanowire, or a single molecule, is connected to the source and drain contacts. The gate modulates the conductance of the channel. One first identifies a suitable basis set and derives the Hamiltonian matrix \mathcal{H} for the isolated channel. Then, the self-energy matrices Σ_1 , Σ_2 , and Σ_S are computed. The self-energy matrices describe how the channel couples to the source contact, the drain contact, and the dissipative processes. Next, the retarded Green’s function, including self-consistent electrostatic potential \mathcal{U} , is computed as

$$G^r(E) = [(E + i0^+)\mathcal{I} - \mathcal{H} - \mathcal{U} - \Sigma_1 - \Sigma_2 - \Sigma_S]^{-1}. \quad (1)$$

Finally, the physical quantities of interest, such as the charge density and current, are computed from the Green’s function.

In this paper, the DC characteristics of ballistic TFETs are simulated by solving open-boundary Schrodinger equation in an atomistic p_z orbital basis set using the NEGF formalism. A nearest neighbor TB parameter of $t_0 = -2.7eV$ is used, and we further consider edge bond relaxation by using $t'_0 = C_{edge}t_0$ for the edge bonds, where $C_{edge} = 1.12$ as parameterized to the *ab initio* band structure simulations. The atomistic transport equation is self-consistently solved with a three-dimensional (3D) Poisson equation using the finite element method, which is efficient to treat a device with multiple gates because it can easily handle an arbitrary grid for complex geometry.

6.2 Circuit simulation

A simulator based on table lookup techniques is implemented to simulate circuits built with TFETs in this paper. The simulator uses the drain current $I_D(V_G, V_D)$ and channel charge $Q(V_G, V_D)$ computed for the intrinsic TFET using the quantum transport simulations described in Section 6.1. These values were used to populate a lookup table at discrete voltage steps of V_{GS} and V_{DS} ranging from 0V to 0.4V. The intrinsic gate and drain capacitances

Table 2: Effects of variations in GNR width on inverter delay, power/energy consumption, and SNM

N	nTFET												
	Delay (s)			Static power (W)			Dynamic energy (J)			SNM (V)			
	10	13	16	10	13	16	10	13	16	10	13	16	
pTFET	10	1.96×10^{-8}	1.11×10^{-8}	1.35×10^{-8}	1.49×10^{-19}	7.39×10^{-19}	4.93×10^{-11}	2.14×10^{-17}	2.52×10^{-17}	2.96×10^{-17}	0.0179	0.08	0.008
	13	1.11×10^{-8}	2.16×10^{-10}	1.41×10^{-10}	7.39×10^{-19}	1.29×10^{-18}	6.01×10^{-11}	2.52×10^{-17}	2.90×10^{-17}	3.20×10^{-17}	0.08	0.18	0.108
	16	1.35×10^{-8}	1.41×10^{-10}	6.24×10^{-11}	4.93×10^{-11}	6.01×10^{-11}	1.20×10^{-10}	2.96×10^{-17}	3.20×10^{-17}	3.65×10^{-17}	0.008	0.108	0.172

Table 3: Effects of source doping variation on inverter delay, power/energy consumption, and SNM

Doping level (Dopant per atom)	nTFET								
	Delay (s)		Static power (W)		Dynamic energy (J)		SNM (V)		
	0.01	0.004	0.01	0.004	0.01	0.004	0.01	0.004	
pTFET	0.01	2.16×10^{-10}	1.91×10^{-9}	1.28×10^{-18}	1.56×10^{-18}	2.90×10^{-17}	2.98×10^{-17}	0.18	0.016
	0.004	1.91×10^{-9}	2.61×10^{-9}	1.56×10^{-18}	1.54×10^{-18}	2.98×10^{-17}	3.10×10^{-17}	0.016	0.021

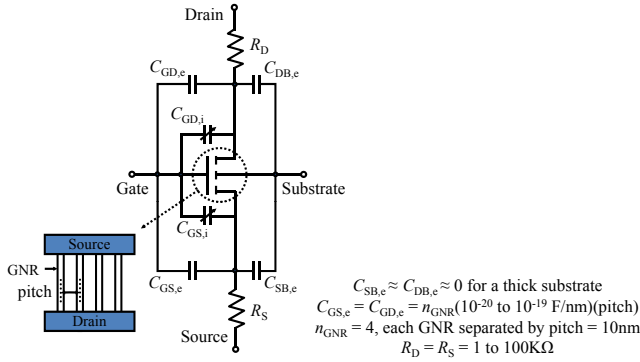


Figure 8: Device model for circuit simulation

$C_{GD,i}$ and $C_{GS,i}$ vary with the gate and drain voltages. These values can be computed and stored in the lookup table by differentiating the channel charge w.r.t V_{GS} and V_{DS} respectively. Thus, $C_{GD,i} = |\partial Q/\partial V_{DS}|$ and $C_{G,i} = C_{GS,i} + C_{GD,i} = |\partial Q/\partial V_{GS}|$, which yields $C_{GS,i} = |\partial Q/\partial V_{GS}| - |\partial Q/\partial V_{DS}|$.

The extrinsic n-type and p-type TFETs were modeled by adding the parasitic capacitances and contact resistances around the intrinsic TFET as shown in Figure 8. In practice multiple GNRs are often fabricated in an array for a wide contact. The TFET considered in this paper consists of 4 equi-distant GNRs that form the channel. The pitch refers to the spacing between the neighboring GNRs in the TFET channel, and in this paper, we assume that an individual GNR channel has a contact width of 10nm, and that this equals the pitch in the GNR array. Thus, a TFET with 4 GNRs has a total contact width of 40nm. The current in the TFET is 4 times the current in the individual GNR channel, and the parasitics are also 4 times that for an individual GNR. Thus, the parasitic junction capacitances $C_{GD,e}$ and $C_{GS,e}$ are given by 0.01–0.1aF/nm times the total TFET contact width of 40nm. It is assumed that the substrate is thick enough that the extrinsic parasitic capacitances $C_{DB,e}$ and $C_{SB,e}$ are negligible. The contact resistances were assumed to range from 1KΩ to 100KΩ, with a nominal value of 10KΩ. The contact capacitance at the device terminals and interconnect capacitance are assumed negligible and are not considered in this paper.

The variations in GNR TFETs can occur and be modeled in two ways. In the first case, only one GNR in the array is subject to a variation, for example, width variation, and all the other GNRs remain nominal. The total current is given by the sum of the currents in the GNRs, nominal or otherwise. However, in practice multiple GNRs in the array may suffer from variation and this is handled in

the second case by assuming that each GNR in the channel experiences the same variation. In this paper, we choose this approach to investigate the effect of variations. The results in this work, therefore, establish lower and upper limits for the effects of variations on GNR TFET circuits.

7. Conclusions

This paper assessed the viability of graphene-based tunneling FETs for low-power applications based on a bottom-up multi-scale framework that treats atomistic scale features in circuit simulations. It was demonstrated that GNR TFET circuits promise extremely low power consumption compared to conventional CMOS, and also offer comparable speed and high noise margins. We also motivated the need to consider the effects of parameter variations in GNR width and contact doping that affect circuit delay, static power, and noise margin. This assessment of the effects of variability and parasitics indicate their important role in circuit performance and design optimization for future graphene-based TFETs.

References

- [1] K. S. Novoselov *et al.*, "Electric field effect in atomically thin carbon films," *Science*, vol. 306, pp. 666–669, 2004.
- [2] R. V. Noorden, "Moving towards a graphene world," *Nature*, vol. 442, pp. 228–229, 2006.
- [3] A. K. Geim and K. S. Novoselov, "The rise of graphene," *Nature Materials*, vol. 6, pp. 183–191, 2007.
- [4] W. A. de Heer *et al.*, "Pionics: The emerging science and technology of graphene-based nanoelectronics," in *Intl. Electron Devices Meeting Technical Digest*, pp. 199–202, 2007.
- [5] M. Y. Han *et al.*, "Energy band-gap engineering of graphene nanoribbons," *Physical Review Letters*, vol. 98, p. 206805, 2007.
- [6] J. Bai *et al.*, "Rational fabrication of graphene nanoribbons using a nanowire etch mask," *Nano Letters*, vol. 9, pp. 2083–2087, 2009.
- [7] L. Jiao *et al.*, "Narrow graphene nanoribbons from carbon nanotubes," *Nature*, vol. 458, pp. 877–880, 2009.
- [8] Q. Zhang *et al.*, "Graphene nanoribbon tunnel transistors," *IEEE Electron Device Letters*, vol. 29, pp. 1344–1346, 2008.
- [9] S. Poli *et al.*, "Computational study of the ultimate scaling limits of cnt tunneling devices," *IEEE Trans. Electron Devices*, vol. 55, pp. 313–321, 2008.
- [10] S. Koswatta and M. Lundstrom, "Influence of phonon scattering on the performance of p-i-n band-to-band tunneling transistors," *Applied Physics Letters*, vol. 92, p. 043125, 2008.
- [11] G. Fiori and G. Iannaccone, "On the possibility of tunable-gap bilayer graphene fet," *IEEE Electron Device Letters*, vol. 30, pp. 261–264, 2009.
- [12] P. Zhao *et al.*, "Computational study of tunneling transistor based on graphene nanoribbon," *Nano Letters*, vol. 9, pp. 684–688, 2009.
- [13] Y. Cao *et al.*, "New paradigm of predictive MOSFET and interconnect modeling for early circuit design," in *Proc. Custom Integrated Circuits Conference*, pp. 201–204, 2000.
- [14] K. Nakada *et al.*, "Edge state in graphene ribbons: Nanometer size effect and edge shape dependence," *Physical Review B*, vol. 54, pp. 17954–17961, 1996.
- [15] S. Datta, *Quantum transport: Atom to transistor*. Cambridge, UK: Cambridge University Press, 2005.