Abstract—In this paper, we investigate the electrical behavior of transistors based on a vertical graphene-hexagonal boron nitride (hBN) heterostructure, using atomistic multiphysics simulations based on density-functional theory and non-equilibrium Green’s function formalism. We show that the hBN current-blocking layer is effective and allows modulation of the current by five orders of magnitude, confirming experimental results. We also highlight—through accurate numerical calculations and simplified analytical modeling—some intrinsic limitations of vertical heterostructure transistors. We show that the overlap between gate contacts and source/drain leads screens the electric field induced by the gates and is responsible for the excessive degradation of the sub-threshold swing, the $I_{ON}/I_{OFF}$ ratio, and the cut-off frequency.

Index Terms—Computational electronics, electron devices, graphene, nanoelectronics.

I. INTRODUCTION

A very abstract level of physical description, all transistors share the same operating principle: the voltage applied to the control (gate or base) electrode modulates a potential energy barrier, which blocks the current between the other two electrodes. This potential energy barrier can be formed in a region of the same material as the other device regions traversed by charge carriers, as in a typical MOSFET, or it can be obtained with heterostructures, i.e., by properly juxtaposing regions of different materials.

The latter situation occurs when one wants to optimize and “engineer” the blocking properties of the energy barrier, as for example in a heterojunction bipolar transistor, a device proposed already in Shockley’s junction transistor patent of 1951 [1], and commonly used in ultrafast electronics since the late 80s. Related to the latter option, the concept of using heterostructures has been proposed by Carver Mead in 1961 for tunneling emission devices [2] and demonstrated in ballistic hot electron transistors [3] and resonant tunneling heterostructure transistors [4].

Heterostructures are very promising for graphene, which has a zero or very small energy gap. Indeed, potential energy barriers in a graphene channel have poor current-blocking quality, causing the small current modulation experimentally observed in graphene transistors [5]. Measured current modulation can be much larger if the gap is increased by means of lateral confinement, as in nanoribbons [6], but the energy gap has large variation if the number of dimer lines in the transversal direction changes by only one, [7] and therefore such approach would require prohibitive fabrication tolerances.

Recently, fully 2-D graphene transistors based on lateral heterostructures have been proposed and investigated from the theoretical point of view [8], [9]. They are inspired by the first experimental success in realizing graphene-boron nitride (BN) lateral heterostructures [10].

Vertical heterobarrier graphene transistors have also been proposed in simulation studies [11], [12], and only two experiments as now demonstrate transistor action [13], [26].

Hexagonal BN and BCN (hBN and hBCN, respectively) layers are particularly well suited as dielectrics, since they are almost lattice matched to graphene [14], and can be similarly obtained by mechanical exfoliation. Other planar dielectrics, such as MoS₂ or graphene, can be of interest in the quest for barrier optimization.

In this paper, we investigate a vertical heterostructure graphene-hBN transistor with atomistic simulations, in order to analyze the advantages and disadvantages of the vertical structure. Indeed, whereas it is probably easier to fabricate as compared to the lateral heterostructure transistor [9], it has some specific disadvantages that hinder device operation and performance. The main issue is related to the position of the gate above the source and drain extensions. This is a real problem, since mobile charge is present in high concentration in the extensions and screens the electric field induced by the gate. Luckily, such screening is only partial, due to the low density of states of graphene sheets.

Whereas current modulation is obtained and is indeed important, a qualified definition of current modulation matters for digital circuit operation: the one obtained by varying the gate voltage $V_{GS}$ of a quantity equal to the supply voltage $V_{DD}$, when the drain voltage $V_{DS} = V_{DD}$.

This is quantitatively expressed by the so-called $I_{ON}/I_{OFF}$ ratio, i.e., the ratio of the drain current $I_{ON}$ in the ON state ($V_{GS} = V_{DS} = V_{DD}$) to the current $I_{OFF}$ in the OFF state ($V_{GS} = 0$ and $V_{DS} = V_{DD}$). If the gate voltage is screened by source and drain leads, the transconductance is degraded,
and therefore also the $I_{ON}/I_{OFF}$ ratio is smaller than $10^4$, the minimum acceptable level for digital circuit operation [15].

II. Physical Models and Numerical Methods

The adopted method is a multi-scale approach based on accurate ab-initio simulations in order to compute the transmission coefficient, which has been fitted by means of a semi-empirical $p_z$ tight-binding (TB) Hamiltonian, to be included in our open-source NanoTCAD ViDES code [16].

In particular, we have performed density-functional theory calculation of the transmission probability of states in semi-infinite graphite leads through the scattering region containing the thin film, according to the formalism by Choi et al. [17]. The film is formed by one, three, or five hBN layers arranged in Bernal stacking as in a recent paper by Ribeiro [18]. The scattering region also includes at least four atomic layers of graphite on each side of the barrier, and it has been optimized on purpose.

Ab-initio calculations have been performed by means of Quantum Espresso [19], using a plane wave basis set in the local density approximation [20] for the exchange and correlation. A 35 Ry wave function cutoff has been considered, while the Brillouin zone has been sampled using a $30 \times 30 \times 30$ Monkhorst-Pack grid in order to obtain convergence of total energy. All atoms in the scattering region are fully relaxed with a force between two adjacent atoms smaller than 0.01 eV/Å.

The ballistic conductance has been calculated by means of the PWCOND [21] module of Quantum Espresso. The transport properties are studied in the framework of Landauer-Büttiker formalism [22], where the total transmission at energy $E$ is obtained as the sum over the transmission probability of the conducting eigenchannels, averaged over the 2-D Brillouin zone normal to the transport direction.

The parameters of a $p_z$ TB Hamiltonian have then been chosen with a fitting procedure, as the ones providing the best fit of transmission coefficients with results obtained from DFT calculations.

The extracted parameters have been then included in the open-source NanoTCAD ViDES simulator, able to solve the 2-D Poisson equation self-consistently with the Schrödinger equation with open boundary conditions, within the non-equilibrium Green’s functions formalism [22].

In particular, the Poisson equation in the 3-D domain reads

$$\nabla \left( \epsilon \nabla (\phi(r)) \right) = -q [p(r) - n(r) + \epsilon_{ix}] \quad (1)$$

where $(\phi(r))$ is the electrostatic potential, $q$ is the electron charge, $(\epsilon)$ is the dielectric constant, and $\epsilon_{ix}$ is the fixed charge in correspondence of the doped reservoirs. The electron and hole concentrations ($n$ and $p$, respectively) are computed by means of the NEGF formalism.

The Green’s function can then be expressed as

$$G(E) = [E I - \Sigma_S - \Sigma_D]^{-1} \quad (2)$$

where $E$ is the energy, $I$ the identity matrix, $H$ the Hamiltonian of the stacked graphene/hBN/graphene structure, and $\Sigma_S$ and $\Sigma_D$ are the self-energies of the source and drain, respectively.

The considered channel material is sketched in Fig. 1. Two graphene layers are placed at the top and the bottom, separated by $n$ layers of hBN, all arranged in the Bernal stacking. Semi-infinite leads are considered at the right and at the left ends of the top and bottom graphene flakes, by means of self-energy definition ($\Sigma_D$ and $\Sigma_S$ respectively).

Let us focus our attention on $H$, which is expressed by means of an atomistic ($p_z$ orbitals) real space basis.

Bloch periodic boundary conditions are imposed along the $x$ direction with period equal to $\Delta = \sqrt{3}a$, where $a = 0.144$ nm is the in-plane atom-to-atom distance. In this way, the $k_x$ wave vector appears in the Hamiltonian.

As an example, let us consider the channel shown in Fig. 1(a), composed by two graphene flakes and $n$ hBN layers (in this particular case, $n = 4$). Each layer is composed by $N_C$ atoms, which are periodically repeated along the $x$ direction as in Fig. 1(b) (In Fig. 1(a), $N_C = 20$). $H$ is a tridiagonal block matrix, with $N_C$ block matrices on the diagonal, whose order is equal to $(n + 2)$.
In particular, $H$ reads

$$H = \begin{pmatrix} D_1 & D_2 & D_3 \\ D_2 & D_4 & D_5 \\ D_3 & D_5 & D_{NC} \end{pmatrix}$$

(3)

where

$$D_i = \begin{pmatrix} \Phi_1 & \Phi_2 & \Phi_3 \\ \Phi_2 & \Phi_4 & \Phi_5 \\ \Phi_3 & \Phi_5 & \Phi_{\theta+2} \end{pmatrix}$$

(4)

and $\Phi_i$ are the potentials of the $i$-th atom as ordered as in Fig. 1(a) and equal to the sum of the on-site energy ($E_{\text{ onsite}}$) obtained through the fitting procedure and the energy potential ($U_i$) provided by the Poisson equation.

If we now define $t$ as

$$t = t e^{i k_x \Delta}$$

(5)

where $i$ is the imaginary number, $t$ the hopping parameter between two nearest-neighbor atoms on the same plane, and $t_\theta$ the hopping parameter between two overlaying atoms belonging to different planes, the off-diagonal block matrices read

$$1 = \begin{pmatrix} t & t & t \\ t & t + t_p & t \end{pmatrix}$$

(6)

$$2 = \begin{pmatrix} t + t^* & t_p & t_p \\ t_p & t + t & t_p \\ t_p & t_p & t + t \end{pmatrix}$$

(7)

Referring to Fig. 1(a), we can divide the channel in three regions: region A, where only the top graphene flake is considered, region B, where the two graphene flakes as well as the $n$ hBN layers are taken into account, and region C, where only the bottom graphene flake is considered.

From a numerical point of view, this structure has been modeled through the atomistic semi-empirical TB Hamiltonian, setting to zero the hopping parameters between black and gray atoms in Fig. 1(a), and considering different parameters in the three different regions. In particular, $E_{\text{ onsite}}$ has been taken equal to zero in correspondence of the C atoms, while $B_{\text{ onsite}}$ and $N_{\text{ onsite}}$ are the on-site energies for the B and N atoms. $t_{\text{CC}}$ and $t_{\text{BN}}$ are the in-plane C-C and B-N hopping parameters, respectively.

When solving the integrals along the $k_x$ axis, we have verified that 64 $k$ values are sufficient to obtain accurate results. As far as the computation of the self-energy $\Sigma$ is concerned, we have adopted a closed-form expression based on the algebra derived in [23] and [24], which provides faster results as compared to the transfer Hamiltonian formalism proposed by Sancho et al. [25]. Transport computation can be still too computationally demanding, so that integration over the $k$ axis has been parallelized by means of MPI subroutines.

The nonlinear system has been solved with the Newton/Raphson (NR) method within a Gummel iterative scheme. In particular, the Schrödinger equation is solved at the beginning of each NR cycle of the Poisson equation, and the charge density is kept constant until the NR cycle converges (i.e., the correction on the potential is smaller than a predetermined value). The algorithm is then repeated cyclically until the norm of the difference between the potential computed at the end of two subsequent NR cycles is smaller than a predetermined value.

III. RESULTS AND DISCUSSION

The simulated device is shown in Fig. 2. In particular, the top and bottom graphene layers are electrically separated by a hBN dielectric layer and connected with the source and drain contacts, respectively. An interlayer distance of 0.35 nm has been assumed. Carrier transport occurs along the vertical direction. The simulated device is a p-MOS field effect transistor, with top and bottom high-k gate dielectrics (i.e., HfO$_2$ with relative dielectric constant equal to 25) with thickness of 4 nm. Top and bottom metallic gates are driven by the same voltage $V_{GS}$. Different numbers of hBN layers have been taken into account, in order to investigate the effect of barrier thickness on device electrical properties.
Fig. 3. Transmission coefficient computed by means of DFT calculations (solid lines) and tight-binding simulations (dashed lines) for different numbers of BN layers.

**TABLE I**

<table>
<thead>
<tr>
<th># hBN layers</th>
<th>(B_{\text{ onsite}}) (eV)</th>
<th>(N_{\text{ onsite}}) (eV)</th>
<th>(t_{CC}) (eV)</th>
<th>(t_p) (eV)</th>
<th>(t_{BN}) (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.5</td>
<td>-1.7</td>
<td>-2.54</td>
<td>-0.072</td>
<td>-1.8</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>-1.9</td>
<td>-2.7</td>
<td>-0.38</td>
<td>-5.5</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
<td>-1.9</td>
<td>-2.7</td>
<td>-0.45</td>
<td>-5.5</td>
</tr>
</tbody>
</table>

In Fig. 3, we show the transmission coefficients of the barrier seen by holes, as a function of the energy and for varying thickness of the hBN layer, when considering a constant potential profile along the heterostructure. As shown in Fig. 3, these curves have been used in order to extract the TB parameters, reported in Table I. Fitting has been optimized for the valence band, since the considered device is a p-type MOSFET. The valence band edge \(BV\) is approximately at an energy of 1.9 eV, as it is clear from the sharp drop of the transmission coefficient in Fig. 3.

The large difference between the parameters obtained for the one-layer and the three- and five-layer hBN barriers reflects the different electrical behavior. While the hBN monolayer device almost behaves as the gapless graphene, three- and five-layer hBN show a band gap larger than 4 eV.

We note that for energy larger than \(BV\), the transmission coefficient decreases exponentially as the number of layers is increased, since the main contribution is due to tunneling, i.e., evanescent modes. For lower energy, the transmission coefficient is independent of the number of layers, since transport is thermionic, i.e., due to propagating modes.

In Fig. 4, we show the drain current \(I_{DS}\) as a function of the gate voltage, for a drain voltage of 0.5 V (source is grounded) and for a different number of hBN layers (i.e., one, three, and five). As can be seen, the one-layer hBN barrier has basically no current-blocking properties, whereas the three- and five-layer hBN allow a current modulation of five orders of magnitude. For \(V_{GS} < -4.5\) V, current is independent of the number of layers, which means it is essentially of thermionic nature.

We note that in the considered device a threshold voltage cannot be determined in terms of charge populating the channel, since graphene is a semi-metal and the channel is inverted for all the considered gate voltages. However, we can still define a threshold voltage in terms of current in the channel: in our case, we assume as a threshold \(I_{TH} = 10^{-5} \mu \text{A/\mu m}\).

The serious problem, from the application viewpoint, is that large current modulation occurs over several volts. To achieve an \(I_{ON}/I_{OFF}\) ratio of 10^4, the minimum required by the International Technology Roadmap for Semiconductors [15] for digital applications, one would need a \(V_{GS}\) sweep of 2 V, which is much larger than the supply voltage \(V_{DD}\) required in next-generation devices (\(V_{DD} < 0.7\) V). On the other hand, for \(V_{DD} = 0.5\) V, the \(I_{ON}/I_{OFF}\) ratio would be at most 30, corresponding to a sub-threshold swing (SS) of \(\approx 330\) mV/dec, to be compared with \(SS = 90\) mV of modern MOSFETs, and to the ideal value of 60 mV/dec.

This is an intrinsic problem, that cannot be solved with improvements in device fabrication. As we already mentioned, it is due to the fact the top and bottom gates are partially screened by the underlying graphene channel, which, in turns translates into a poor electrostatic control of the channel barrier through the gate voltage.

A simple capacitive model of the device is shown in Fig. 5, on the right: \(C_1\) and \(C_3\) represent the capacitance between the metallic gates and the graphene sheets, \(C_2\) the capacitance...
between the two graphene sheets, while $C_4$ and $C_5$ the quantum capacitances of the top and bottom graphene sheets. In particular, the quantum capacitance $C_Q$ has been obtained as

$$C_Q = C_{ox} \frac{\partial V_G}{\partial \epsilon} - 1$$

where $\epsilon$ is the potential in the channel, $V_G$ is the applied gate voltage, and $C_{ox}$ is the oxide capacitance. We consider capacitances per unit area.

Average capacitances can be extracted from device simulations and are shown in Fig. 5 for the bias point $V_{DS} = 0.5$ V, $V_{GS}$ ranging from 0 to 0.5 V. $C_1$ and $C_3$ have been computed as $\lambda_{ox}$, where $\lambda$ is the dielectric constant of HfO$_2$ and $t_{ox}$ is the dielectric thickness.

As can be seen, $C_4$ and $C_5$ are larger than the electrostatic capacitances $C_1$ and $C_3$, which explains the poor effectiveness of the gate voltage in modulating the current. Graphene somewhat helps because it has a relatively small density of states and therefore smaller quantum capacitances $C_4$ and $C_5$. Roughly, we have $SS \approx ((C_3 + C_5)/C_1^2)k_BT \ln 10/q = 386$ mV/dec, which can be improved only slightly by further thinning of the dielectric layer, considering that we still have to avoid dielectric breakdown (for an effective oxide thickness of the HfO$_2$ layer of just 2 nm, we would have $C_d = 0.11$ F/m$^2$ and $SS = 220$ mV/dec).

The presence of parasitic capacitance strongly limits the expected performance of the vertical graphene-hBN transistor at high frequency. In order to assess its performance, we have evaluated the cut-off frequency $f_T$, which in the quasi-static model is obtained as

$$f_T = \frac{\partial I_{DS}/\partial V_{GS}}{2\pi Q/\partial V_{GS}}$$

where $Q$ is the total mobile charge in the device channel. In Fig. 6, we show the cut-off frequency for the devices with the best performance in terms of current modulation, i.e., the devices with three and five layers hBN barrier. As can be seen, the large capacitances lead to sub-GHz $f_T$, whereas $f_T$ in state-of-the-art CMOS processes can reach few hundred GHz.

\section{IV. Conclusion}

We have shown that a vertical graphene-hBN heterostructure can provide the current-blocking properties that are missing in graphene-only devices and can allow to obtain a large current modulation of five orders of magnitude, confirming recent experimental results \cite{13}.

Unfortunately, the screening of the field induced by the gate voltage due to the overlap between the gates and the leads of source and drain terminals strongly degrades the sub-threshold slope and therefore the transconductance, the $I_{ON}/I_{OFF}$ ratio, and the cut-off frequency. Such results make the device unfit for digital electronics, where $I_{ON}/I_{OFF}$ ratio of the order to $10^4$ are required for supply voltages $V_{DD} < 0.7$ V \cite{15}.

In comparison, the lateral heterostructure transistor proposed in \cite{9} has an intrinsic advantage, in that the screening of the gate potential only occurs in the small overlap regions between the gate and source and drain extensions. For this reason, in that case, a more promising $SS = 110$ mV/dec was obtained.

\begin{thebibliography}{10}
\end{thebibliography}


**Gianluca Fiori** received the Ph.D. degree from the University of Pisa, Pisa, Italy, in 2005. Since 2007, he has been an Assistant Professor with the University of Pisa.

**Samantha Bruzzone** received the Ph.D. degree from the University of Pisa, Pisa, Italy, in 2003. Since 2010, she has been with the Department of Information Engineering, University of Pisa.

**Giuseppe Iannaccone** (M’98–SM’10) is currently a Professor of electronics with the University of Pisa, Pisa, Italy. His research interests include electron device modeling, nanoelectronics, analog design, and smart systems.