Exploiting negative differential resistance in monolayer graphene FETs for high voltage gains

Roberto Grassi, Antonio Gnudi, Member, IEEE, Valerio Di Lecce, Elena Gnani, Member, IEEE, Susanna Reggiani, Member, IEEE, and Giorgio Baccarani, Fellow, IEEE

Abstract—Through self-consistent quantum transport simulations, we evaluate the RF performance of monolayer graphene FETs in the bias region of negative output differential resistance. We show that, compared to the region of quasi-saturation, a voltage gain larger than 10 can be obtained, at the cost of a decrease in the maximum oscillation frequency of about a factor of 1.5–3 and the need for a careful circuit stabilization.

Index Terms—Graphene FET, negative differential resistance, terahertz operation, voltage amplifier.

I. INTRODUCTION

GRAPHENE has been suggested as a promising material for analog and radio-frequency (RF) applications due to its exceptional electrical properties. In particular, the high mobility and large group velocity can translate to a high device transconductance $g_m$ and high cut-off frequency $f_T$, and there is no need for a band gap to switch off the device as in digital applications [1]. Fabricated graphene field-effect transistors (GFETs) exhibiting $f_T$ of hundreds of gigahertz have already been reported [2]–[5], together with the first applications [6]–[9]. However, challenges still remain. Particularly in short-channel devices, where velocity saturation does not occur, the lack of a band gap leads to poor current saturation (i.e., pronounced drain conductance $g_d$), which negatively affects the device performance as an amplifier. This is especially true at low frequency, where the ability to amplify signals is expressed by the intrinsic voltage gain $g_m/g_d$, which is limited to only few units in monolayer GFETs [8]–[11], with a record value of 5.3 for channel lengths of the order of 1 $\mu$m. While voltage gain is not strictly necessary at high frequency, a large $g_d$ also contributes to degrade to some extent the maximum oscillation frequency $f_{max}$, which is the maximum frequency at which power gain can be be obtained, and, in many applications, represents a more important figure of merit than $f_T$ [12].

To address the above issues, the use of bilayer graphene, where a band-gap can be introduced through a vertical electric field, has been suggested and values of $g_m/g_d$ as high as 35 have been experimentally demonstrated [13], [14]. In this paper, as an alternative approach, we study whether the RF performance of monolayer GFETs, in particular the voltage gain, can be improved by choosing the bias point in the region of negative differential drain resistance (NDR), i.e., of negative $g_d$. Such NDR has been observed experimentally in long-channel devices [15], [16] and predicted by numerical simulations for short-channel lengths as well [17]–[22], but an analysis of the device small-signal behavior in the NDR region has not been reported yet. Obviously, a device with negative $g_d$ cannot be used as amplifier with a high-impedance load, since the resulting circuit is unstable. However, if the load impedance $1/G_L$ is sufficiently low, the parallel of $g_d$ and $G_L$ can be made positive and, in principle, smaller than the achievable values of $g_d$ in the standard bias region of quasi-saturation, thus potentially resulting in a higher voltage gain.

The objective of the paper is to numerically investigate the feasibility of the above idea. Steady-state simulations are performed to compute the dc $I–V$ and $Q–V$ characteristics of the device, which allow the extraction of the parameters of a small-signal equivalent circuit and, ultimately, of the analog and RF figures of merit. The device structure, simulation model, and small-signal model are described in Section II. The $I–V$ characteristics are presented in Section III, where the connection between $g_m$ and the underlying device physics is also clarified. The circuit stability of the device in the common-source configuration, for a bias point in the NDR region, is analyzed in detail in Section IV. RF performance is discussed in Section V followed by conclusions in Section VI.

II. DEVICE STRUCTURE AND MODEL

We consider the dual-gate device structure represented in Fig. 1 which is similar to the experimental one in [3], although...
The simulations are performed using an in-house developed code for GFETs, based on the self-consistent solution of the 2D Poisson equation and the ballistic non-equilibrium Green’s function (NEGF) equations [24], with a $p_z$ tight-binding Hamiltonian and a mode-space solution approach. The model is the same as the one in [19] but with a different treatment of the interfaces: the source and drain self-energies are computed with the metal-graphene coupling strength $\Delta$ set to zero, and Neumann boundary conditions instead of Dirichlet. The latter are used in Poisson’s equation at the source and drain ends. Moreover, instead of assuming a finite channel width with periodic boundary conditions as in [19], the device is taken to be infinite in the transverse direction, so that sums over modes are replaced by integrals over the transverse wavevector. The latter are performed by a Gaussian quadrature with 40 $k$-points.

The small-signal frequency behavior of the device is analyzed through the usual quasi-static approximation, which consists in constructing a small-signal equivalent circuit, whose resistive and capacitive elements are extracted from the dc characteristics of charge and current at the various terminals (Fig. 2). The small-signal circuit model is the same used for silicon MOSFETs [23], with the back gate acting as bulk terminal. The source/drain charge $Q_{S/D}$ is taken equal to the charge contribution relative to injection from source/drain of the ballistic transport model. Other charge-partitioning schemes are possible: the authors of [25] have checked that different choices of $Q_S$ and $Q_D$ (with fixed sum $Q_S + Q_D$) have a negligible impact on their results. Source, drain, and gate contact resistances $R_s$, $R_d$, and $R_g$ are included in the model as additional parameters. The effective doping corresponding to $V_{GS} = 0$ is about $1.9 \times 10^{13}$ cm$^{-2}$. From a technological point of view, such electrostatic doping technique is easier and more controllable, although the dual gate structure introduces some complications due to the additional wiring and management of the high back-gate voltage.

The simulations are performed using an in-house developed code for GFETs, based on the self-consistent solution of the 2D Poisson equation and the ballistic non-equilibrium Green’s function (NEGF) equations [24], with a $p_z$ tight-binding Hamiltonian and a mode-space solution approach. The model is the same as the one in [19] but with a different treatment of the interfaces: the source and drain self-energies are computed with the metal-graphene coupling strength $\Delta$ set to zero, and Neumann boundary conditions instead of Dirichlet. The latter are used in Poisson’s equation at the source and drain ends. Moreover, instead of assuming a finite channel width with periodic boundary conditions as in [19], the device is taken to be infinite in the transverse direction, so that sums over modes are replaced by integrals over the transverse wavevector. The latter are performed by a Gaussian quadrature with 40 $k$-points.

The small-signal frequency behavior of the device is analyzed through the usual quasi-static approximation, which consists in constructing a small-signal equivalent circuit, whose resistive and capacitive elements are extracted from the dc characteristics of charge and current at the various terminals (Fig. 2). The small-signal circuit model is the same used for silicon MOSFETs [23], with the back gate acting as bulk terminal. The source/drain charge $Q_{S/D}$ is taken equal to the charge contribution relative to injection from source/drain of the ballistic transport model. Other charge-partitioning schemes are possible: the authors of [25] have checked that different choices of $Q_S$ and $Q_D$ (with fixed sum $Q_S + Q_D$) have a negligible impact on their results. Source, drain, and gate contact resistances $R_s$, $R_d$, and $R_g$ are included in the model as additional parameters. The effective doping corresponding to $V_{GS} = 0$ is about $1.9 \times 10^{13}$ cm$^{-2}$. From a technological point of view, such electrostatic doping technique is easier and more controllable, although the dual gate structure introduces some complications due to the additional wiring and management of the high back-gate voltage.
of GFETs with metal-doped [19], [20] and chemically-doped [21], [22] source and drain regions. The different behavior is due to the formation of either an n-n-n or an n-p-n double junction [20]. The agreement with experiments [15], [16] is only qualitative due to the gap between the conditions considered in the simulation (20-nm channel length and ballistic transport in ideal graphene) and the limitations of the present graphene technology (e.g., contact resistance and interface effects).

NDR is obtained at the cost of lower \( I_D \) and transconductance \( g_m \), as is evident from the trans-characteristics and the corresponding \( g_m \) vs. \( V_{GS} \) plots in Fig. 4. The peak \( g_m \) decreases by more than a factor of four. The reason can be ascribed to: (i) reduced transmission due to double band-to-band tunneling across the n-p-n junction [28]; (ii) a transport-mode bottleneck effect induced by the Dirac point at the drain side [20]. Similar asymmetric performance with respect to top gate bias in dual-gated structures is observed in experiments [29], [30]. In Fig. 4, we also show the results obtained by increasing \( t_{ox} \) to 2.4 nm (EOT = 1 nm) or by lowering \( V_{BS} \) to 5 V (effective doping of \( 1 \times 10^{13} \text{ cm}^{-2} \)). As one might expect, a larger EOT leads to significant degradation of the peak \( g_m \), in both bias regions, due to reduced electrostatic control of the top gate on the channel potential. In the bias region corresponding to NDR, a lower \( V_{BS} \) also goes in the direction of decreasing the peak \( g_m \), highlighting the importance of a heavy doping of the source and drain regions in this transport regime. Increasing further the drain degeneracy with respect to the case with \( V_{BS} = 9 \text{ V} \) would require the use of a high-\( \kappa \) substrate material, for the vertical electric field in the back dielectric is already close to the SiO2 limit of 1 V/\( \text{nm} \) at \( V_{BS} = 9 \text{ V} \). On the other hand, since a higher \( \kappa \) also implies a larger back gate capacitance, the back oxide thickness (and consequently \( V_{BS} \)) should be increased to avoid a counterproductive effect on \( g_m \).

The extracted small-signal parameters for the bias points of peak \( g_m \) of the quasi-saturation and NDR regions are provided for reference in Table I. Since the \( Q-V \) characteristics (not shown) were found to be affected by numerical noise, we used a Savitsky-Golay filter of order two [31] to compute the parameters in Table I, rather than using finite differences as for the \( g_m \) plots in Fig. 4.

### IV. Stability analysis

In this section, we study the stability of the reference device in the common-source configuration, at \( V_{DS} = -0.45 \text{ V} \) and \( V_{GS} = -1.3 \text{ V} \) (operating point of peak \( g_m \) of the NDR region). We consider a channel width \( W = 1 \mu\text{m} \). From the small-signal circuit of Fig. 2, one can derive the expressions of the \( Y \)-parameters of the extrinsic transistor \( Y_{11}, \ldots, Y_{22} \), which allow to compute the output and input admittances \( Y_{out} \) and \( Y_{in} \) as a function of the source and load admittances \( Y_A \) and \( Y_L \), respectively (see circuit in Fig. 5a for symbol definitions). Regarding the functional dependence of \( Y_A \) and \( Y_L \) on frequency \( f \), we assume the circuit models in Figs. 5(b)-(c), where \( L_A \) and \( L_L \) represent series parasitic interconnect inductances.

The stability of an RF amplifier is usually ensured by requiring that both \( Y_{out} \) and \( Y_{in} \) have a positive real part in the whole frequency range where the amplifier behaves as an active network [32].

\[
\Re\{Y_{out}\} > 0, \quad \Re\{Y_{in}\} > 0. \tag{1}
\]

The “stability circles” technique then allows to find on the Smith chart the range of values of \( Y_A \) and \( Y_L \) for which (1) are satisfied at each frequency. Such approach, however, cannot be applied in the present case, since \( \Re\{Y_{out}\} \) is potentially negative at low frequency, where the device is unilateral. The real part of \( Y_{out} \) is plotted as a function of frequency in Fig. 6 left for the case of \( |Y_A| \equiv \infty \) (short-circuit at the input port) and for values of \( R_a = R_d \) from 200 down to 50 \( \Omega \cdot \mu\text{m} \), i.e. from typical experimental values down to best achievable ones [53]. The low-frequency value of \( Y_{out} \), given by the output conductance \( g_{out} \) of the extrinsic transistor

\[
g_{out} = \frac{g_d}{1 + (R_s + R_d)g_d + R_s(g_m + g_{mb})}, \tag{2}
\]

is strongly affected by the source and drain contact resistances, as shown in the figure, but is independent of \( Y_A \). We note that stability is still possible if

\[
\Re\{Y_{out} + Y_L\} > 0, \quad \Re\{Y_{in} + Y_A\} > 0, \tag{3}
\]

which represent less restrictive requirements than (1). Acceptable values of \( Y_A \) and \( Y_L \) must satisfy both inequalities simultaneously. Let us first start by assuming \( |Y_A| \equiv \infty \). We
consider here a value of contact resistance of \(100 \Omega \cdot \mu m\), for which \(g_{\text{out}} = -37.9 \, mS\). To satisfy the first inequality in (3) at low frequency, the negative value of \(g_{\text{out}}\) must be compensated by a load conductance \(G_L > -g_{\text{out}}\), as already mentioned in Section I. However, a too large parasitic inductance \(L_L\) might cancel the effect of \(G_L\) at high frequency, as illustrated by the plot of \(\Re\{Y_{\text{out}} + Y_L\}\) in Fig. 6(right) for different values of \(L_L\) (see only the curves with \(L_A = R_A = 0\), the other ones being discussed later). We find that \(L_L\) must be limited to \(\approx 10 \, \mu H\), an upper bound which should be compatible with an integrated version of the amplifier. Having fixed the values of \(G_L\) and \(L_L\) this way, we look for values of load capacitance \(C_L\), source resistance \(R_A\), and \(L_A\) that allow to satisfy the second inequality in (3). As shown in Fig. 7(left), the real part of \(Y_{\text{in}}\) is not significantly affected by \(C_L\). Its negative plateau can be compensated by sufficiently low values of \(R_A\) and \(L_A\) (Fig. 7(right)). We find that choosing \(R_A = 50 \, \Omega\), which is the typical characteristic impedance of a transmission line, together with the same upper bound of \(10 \, \mu H\) for \(L_A\) as for \(L_L\), provides \(\Re\{Y_{\text{in}} + Y_A\} > 0\). The stability of the circuit is finally demonstrated by checking that the same values of \(R_A\) and \(L_A\) also give \(\Re\{Y_{\text{out}} + Y_L\} > 0\) (triangles down in Fig. 4(right)). Of course, instead of the procedure outlined here, one could also have tested the circuit stability using the standard pole analysis.

In summary, we have shown that is possible to ensure the stability of the circuit by canceling out the negative real part of \(Y_{\text{out}}\) and \(Y_{\text{in}}\), through a proper choice of the load conductance \(G_L\) and the source resistance \(R_A\), respectively. The procedure requires: (i) an estimate of the contact resistances and hence of \(g_{\text{out}}\); (ii) small enough parasitic interconnect inductances at the input and output port, or the effect of \(G_L\) and \(R_A\) is made void at high frequency.

### V. ANALOG AND RF METRICS

The following figures of merit are evaluated for the device in the common-source configuration and biased in either the NDR or quasi-saturation region: dc voltage gain \(A_{v0} = v_2/v_1|_{f=0}\) with load \(G_L\), cut-off frequency \(f_T\), and maximum oscillation frequency \(f_{\text{max}}\). From the small-signal circuit in Fig. 2 a simple expression for \(A_{v0}\) can be derived:

\[
A_{v0} = -\frac{g_m/g_d}{1 + G_L/g_{\text{out}}},
\]

where \(g_{\text{out}}\) is given by (2). \(f_T\) is obtained by extrapolating the low-frequency short-circuit current gain \(|H_{21}| = |Y_{21}/Y_{11}|\) to unity at \(-20 \, \text{dB/dec}\), and \(f_{\text{max}}\) by extrapolating the low-frequency maximum stable gain MSG = \(|Y_{21}/Y_{12}|\) to unity at \(-10 \, \text{dB/dec}\), \(f_{\text{max}}\) is defined here with reference to MSG rather than Mason’s unilateral gain [32], since the latter cannot be defined in the NDR region. In the case of \(R_g = C_{\text{int}} = C_{\text{ext}} = 0\), analytical expressions for \(f_T\) and \(f_{\text{max}}\) can be derived by isolating the \(1/s\) terms in the expansion of \(Y_{21}(s)/Y_{11}(s)\) and \(Y_{21}(s)/Y_{12}(s)\), respectively.
and by equating their magnitude to unity:

\[ f_T = \left( \frac{2\pi}{g_m} \right), \quad D = |C_{gg}[1 + (R_s + R_d) g_d + R_s g_m]| + \frac{C_{gd}(R_s + R_d) g_m}{1 + R_s (g_m + g_b)}, \]

\[ f_{\text{max}} = \left( \frac{2\pi}{g_m} \right) \left| \frac{|C_{gs} + C_{gd} R_s g_d + C_{gd} [1 + R_s (g_m + g_b)]}{C_{gs} + C_{gd} + C_{gd}} \right|, \]

where the total gate capacitance \( C_{gg} \) is related to the circuit elements in Fig. 7 through \( C_{gg} = C_{gs} + C_{gd} + C_{gd} \).

Let us start considering the peak-\( g_m \) bias point of the NDR region (\( V_{DS} = -0.45 \) V, \( V_{GS} = -1.3 \) V) and a value of contact resistance \( R_s = R_d = 100 \) Ω·μm. Again, we assume \( W = 1 \) μm. Fig. 8 left shows the frequency magnitude response of the voltage gain \( v_2/v_1 \) for different values of \( G_L \) (and fixed values of \( C_L \) and \( L_L \)). In accordance with [3], the low-frequency value \( |A_{\text{dc}}| \) is strongly peaked around \( G_L = -g_{\text{out}} \approx 37.9 \) mS. If the difference between \( G_L \) and \( -g_{\text{out}} \) is less than 5%, a voltage gain larger than 10 can be obtained. Furthermore, the larger the gain, the smaller the corresponding bandwidth, resulting in an approximately constant gain-bandwidth product GBW of about 200 GHz. In the same figure, we also plot the frequency response of \( |v_2/v_1| \) for \( G_L = 37.9 \) mS (see legend for values of \( R_A \) and \( L_A \)), which is nearly identical to \( |v_2/v_1| \), indicating a minor effect of the source admittance. Even when including the additional parasitics \( R_g, C_{int}, \) and \( C_{ext} \), the frequency response does not change significantly. The value of \( R_g = 4 \) Ω considered here has been calculated in a similar way to [25], by assuming a tungsten gate (resistivity of 56 Ω·m) of dimensions \( W \times L_g \times L_g = 1 \) μm × 20 nm × 60 nm, contacted on both sides [23, Eq. 6.62]. The value of \( C_{int} = C_{ext} = 0.1 \) fF is the same as in [26]. In Fig. 8 right the frequency response of the current gain \( |H_{21}| \) and of MSG is reported. For high enough frequencies, where the transistor becomes unconditionally stable, MSG is replaced by the maximum available gain MAG [32] (almost invisible in Fig. 8 right). In the case of \( R_g = C_{int} = C_{ext} = 0 \), values of \( f_T = 2.3 \) THz and \( f_{\text{max}} = 890 \) GHz are extracted. For comparison purposes, we also plot the transducer power gain \( G_T \), obtained with source and load parameters (see legend) that ensure the stability of the amplifier. It can be seen that \( G_T \) falls off to one at a frequency not too far from \( f_{\text{max}} \). Again, the inclusion of additional parasitics has only a limited impact.

Similar plots of voltage gain, current gain, and power gain, but for the peak-\( g_m \) bias point of the quasi-saturation region (\( V_{DS} = 0.45 \) V, \( V_{GS} = 0.2 \) V), are shown in Fig. 9. For reference, Mason’s unilateral gain \( U \) vs. frequency (right) for a 1-μm-wide device at \( V_{DS} = 0.45 \) V and \( V_{GS} = 0.2 \) V (quasi-saturation regime). The dashed curves are obtained with \( R_g = 4 \) Ω and \( C_{int} = C_{ext} = 0.1 \) fF. The other parameters are indicated in the legend.
shown in the inset of each figure. The discontinuity of $A_{\text{v0}}$ (Eq. 4), a strong dependence of $f_{\text{max}}$, and the respective voltage gain characteristics in the NDR region compared to the quasi-saturation region is shown in Fig. 10. The higher peak of the voltage gain to the input signal, which thus limits the use of the device for high-frequency applications. Through a small-signal analysis with parameters extracted from atomistic quantum transport simulations, the stability and RF performance of the transistor in the common-source amplifier configuration have been evaluated. Stability has been found to be a critical issue: compensation of the negative real part of the input and output admittances is required by means of a careful calibration of the source and load networks. Such compensation can be unfeasible if the series parasitic inductance is too large. Voltage gains exceeding the intrinsic gain in the quasi-saturation regime and larger than 10 can actually be achieved. However, this comes at the expenses of a voltage swing available to the input signal smaller than 1 mV and of a reduced bandwidth. Also, $f_{\text{max}}$ is found to be smaller than in the quasi-saturation regime as a result of a four-fold decrease of $g_m$, which is intrinsically related to the device physics responsible for the NDR mechanism.

**Acknowledgment**

The authors would like to thank Prof. A. Santarelli of University of Bologna and Dr. T. Low of IBM T. J. Watson Research Center for useful discussions.

**References**

[1] F. Schwierz, “Graphene transistors,” *Nature Nanotechnology*, vol. 5, pp. 487–496, 2010.

[2] 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.

[3] Y.-M. Lin, C. Dimitrikopulos, K. A. Jenkins, D. B. Farmer, H.-Y. Chiu, A. Grill, and P. Avouris, “100-GHz Transistors from Wafer-Scale Epitaxial Graphene,” *Science*, vol. 327, no. 5966, pp. 662, 2010.

[4] Y. Wu, Y.-M. Lin, A. A. Bol, K. A. Jenkins, F. Xia, D. B. Farmer, Y. Zhai, and P. Avouris, “High-frequency, scaled graphene transistors on diamond-like carbon,” *Nature*, vol. 472, no. 7341, pp. 74–78, 2011.

[5] R. Cheng, J. Bai, L. Liao, H. Zhou, Y. Chen, L. Liu, Y.-C. Lin, S. Jiang, Y. Huang, and X. Duan, “High-frequency self-aligned graphene transistors with transferred gate stacks,” *Proceedings of the National Academy of Sciences*, 2012.

[6] H. Wang, A. Hsu, J. Wu, J. Kong, and T. Palacios, “Graphene-Based Ambipolar RF Mixers,” *Electron Device Letters, IEEE*, vol. 31, no. 9, pp. 906–908, 2010.

[7] Y.-M. Lin, A. Valdes-Garcia, S.-J. Han, D. B. Farmer, I. Meric, Y. Sun, Y. Wu, C. Dimitrikopulos, A. Grill, P. Avouris, and K. A. Jenkins, “Wafer-scale graphene integrated circuit,” *Science*, vol. 332, no. 6035, pp. 1294–1297, 2011.

[8] S.-J. Han, K. A. Jenkins, A. Valdes Garcia, A. D. Franklin, A. A. Bol, and W. Haensch, “High-frequency graphene voltage amplifier,” *Nano Letters*, vol. 11, no. 9, pp. 3690–3693, 2011.

[9] E. Guerriero, L. Polloni, L. G. Rizzi, M. Bianchi, G. Mondello, and R. Sordan, “Graphene audio voltage amplifier,” *Small*, vol. 8, no. 3, pp. 357–361, 2012.

[10] Y. Wu, K. A. Jenkins, A. Valdes-Garcia, D. B. Farmer, Y. Zhu, A. A. Bol, C. Dimitrikopulos, W. Zhu, F. Xia, P. Avouris, and Y.-M. Lin, “State-of-the-art graphene high-frequency electronics,” *Nano Letters*, vol. 12, no. 6, pp. 3062–3067, 2012.

[11] L. G. Rizzi, M. Bianchi, A. Behnam, E. Carrion, E. Guerriero, L. Polloni, E. Pop, and R. Sordan, “Cascading wafer-scale integrated graphene complementary inverters under ambient conditions,” *Nano Letters*, vol. 12, no. 8, pp. 3948–3953, 2012.

[12] F. Schwierz, “Electronics: Industry-compatible graphene transistors,” *Nature*, vol. 472, no. 7341, pp. 41–42, 2011.
