Ultimate limits for highest modulation frequency and shortest response time of field effect transistor

M. Shur[a], G. Rupper[b], and S. Rudin[b]
[a] a Department Rensselaer Polytechnic Institute, Troy, New York, USA; b U.S. Army Research Laboratory, Adelphi, Maryland, USA

ABSTRACT

For high speed high mobility devices, the conventional notion of the electron mobility determining the device speed no longer applies. The ballistic transport plays the dominant role. The electron response becomes faster with the mobility increase in a limited range of relatively low mobility values. With a further increase in the electron mobility, first the plasmonic ringing determines the characteristic response time and then the viscous transport becomes dominant for small feature sizes. The minimum response time and the maximum device modulation frequency correspond to the subpicosecond and terahertz ranges, respectively. The recent experiments of the FET switching using femtosecond optical laser pulses are in good agreement with the predicted sub picosecond switching times and demonstrate a larger sensitivity enhancement due to the constructive interference of the impinging THz pulse and the optical pulse field rectified by the device nonlinearity.

Keywords: Modulation frequency, response time, field effect transistor, two-dimensional electron gas, viscosity

1. INTRODUCTION

The ultimate speed of electronic devices is the key issue for future communications and data processing systems. The 300 GHz communication links are already being actively developed. The pursuit of higher operating frequencies and shorter switching times focused attention on higher mobility semiconductors, such as Ge, InGaAs, InAs and, more recently, graphene. However, at short device features sizes typical for high speed devices, the conventional notions of the mobility determining the device speed and of the transit time dominated cutoff frequencies no longer apply. The quasi-ballistic and ballistic transport [1] starts playing the dominant role. The electron inertia and the viscosity of the two-dimensional electron fluid in the field effect transistor (FET) channels become very important. Our analytical estimates and detailed hydrodynamic simulations [2-4] reveal that the electron response becomes faster with the mobility increase only in a limited range of relatively low mobility values. With a further increase in the electron mobility, the plasmonic ringing determines the characteristic response time that becomes of the order of the momentum relaxation time. Therefore, the response time actually starts increasing proportionally to the mobility values up to the point, where this dependence saturates due to the dominant effect of the electron viscosity. The minimum response time and the maximum device modulation frequency correspond to the subpicosecond and terahertz ranges, respectively. The recent experiments of the FET switching using femtosecond optical laser pulses are in good agreement with the predicted sub picosecond switching times and demonstrate a larger sensitivity enhancement due to the constructive interference of the impinging THz pulse and the optical pulse field rectified by the device nonlinearity. We introduce two related measures of the transistor response. The first one is the characteristic time scale of the drain voltage induced by a quick step in the gate-to-source bias Figure 1 a. In our experiments, the application of a fast-varying gate-to-source bias is done by illuminating the device by a focused beam as schematically shown in Fig. 1 b. Figure 2 shows the expected response for low-mobility, high damping rate devices (the top panel) and the for high-mobility, low damping rate devices (the bottom panel.

2. RESPONSE SPEED (DRIFT MODEL)

The drift model equations for the two-dimensional electron gas (2DEG) are given by

$$\frac{q}{m^*}\frac{\partial U}{\partial x}+\frac{v}{\tau_m}=0 \tag{1}$$

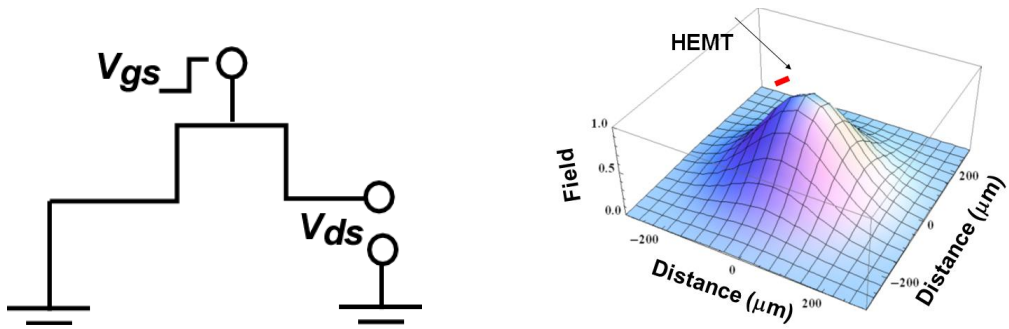
Micro- and Nanotechnology Sensors, Systems, and Applications IX, edited by Thomas George, Achyut K. Dutta, M. Saif Islam, Proc. of SPIE Vol. 10194, 101942M · © 2017 SPIE
CCC code: 0277-786X/17/$18 · doi: 10.1117/12.2261105

$$\frac{\partial n_s}{\partial t} + \frac{\partial n_s v}{\partial x} = 0 . \tag{2}$$

Here m^* is the effective mass, q is the electron charge, τ_m is the momentum relaxation time, U is the channel potential, and v is the electron drift velocity. These equations are valid when $2\pi f\tau_m \ll 1$ and $2\pi f_p\tau_m \ll 1$, where f is frequency, f_m is the fundamental plasma frequency,

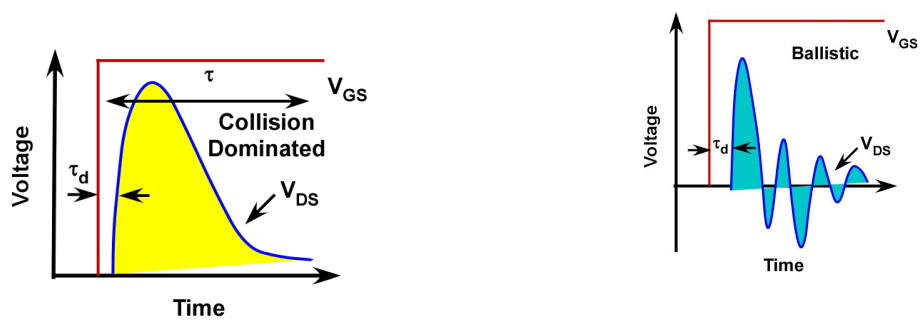
$$n_s = n_o \ln\left(1 + \exp\frac{qU_{gt}}{\eta k_B T}\right). \tag{3}$$

is the 2DEG concentration [5], U_{gt} is the gate voltage swing, k_B is the Boltzmann constant , η is the ideality factor, T is temperature, $n_0 = C\eta k_B T/q$, $C = \varepsilon_0\varepsilon/d = C_{ch}/WL$ is the gate-to-channel capacitance per unit area, C_{ch} is the gate-to-channel capacitance, ε_0 is the vacuum dielectric permittivity, ε is the dielectric constant. W is the gate width, L is the gate length, and d is the channel-to-gate separation.



(a) (b)

Figure 1. (a) Equivalent circuit diagram; (b) Space distribution of the electric field in the focused beam. Frequency is 1 THz. The size of a 10 μm wide HEMT is shown for comparison.



(a) (b)

Figure 2. The expected response for low-mobility, high damping rate devices (the top panel) and for the high-mobility, low damping rate devices (the bottom panel.

(a) (b)

Figure 3. Modulated output (a) and the definition of the maximum modulation frequency f_m (b).

The analysis of Eqs. (1-2) using the boundary conditions [6]

$$U(x=0,t)=U_{gt}+U_a(t)\cos(\omega t)\,. \tag{4}$$

$$\partial U^2/\partial x\Big|_{x=L}=2j_d/(\mu C)\,, \tag{5}$$

where j_d is the drain current density and $\mu = q\tau_m/m^*$ is the low field mobility, yields the following estimate for the ultimate response time for the above threshold regime ($U_{gt} >> \eta k_B T/q$):

$$\tau = L^2/\left(\mu U_{gt}\right). \tag{6}$$

(see reference [7]). As seen, the ultimate response time is limited by the transit time of electrons filling or emptying the channel propagating with the velocity $v = \mu U_{gt}/L$. This estimate for τ is applicable if $\mu U_{gt}/L << v_s$, where v_s is the electron saturation velocity. Interpolating this equation beyond the range of its applicability leads to the following estimate at large gate voltage swings:

$$\tau_s \approx L/v_s\,. \tag{7}$$

Figure 4 compares the values of the response time for the voltage swing U_{gt} =0.1 V

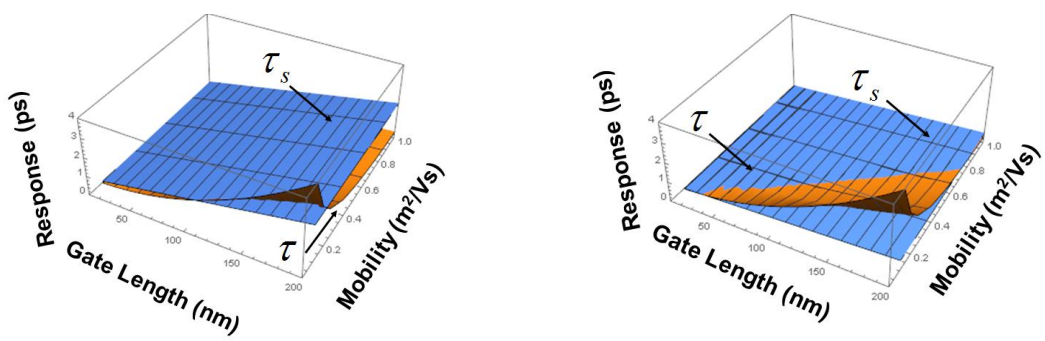


Figure 4. Comparison of response times τ and τ_s for U_{gt} = 0.1 V and $v_s = 10^5$ m/s (a) and 3.5×10^5 m/s (b).

The larger of these two response times τ and τ_s dominates. As seen, the response time could be in the sub picosecond range corresponding to the modulation frequency $f_m = 1/2\pi\tau$ in the sub-THz range. The major drawback of this model is that it does not account for the electron inertia and for the viscosity of the electronic fluid (traditionally referred to as 2DEG). These important effects could be accounted for in the frame of the hydrodynamic model.

3. RESPONSE SPEED (HYDRODYNAMIC MODEL)

The equations of the hydrodynamic model, see, for example [2-4], are the continuity equation

$$\partial n_s/\partial t+\nabla\cdot(n_s\vec{v})=0\,, \tag{8}$$

the Naiver-Stokes equation

$$\partial\vec{v}/\partial t+\left(\vec{v}\cdot\nabla\right)\vec{v}+q\nabla U/m+\nabla P/n+\vec{v}/\tau_m-\gamma\nabla^2\vec{v}=0\,, \tag{9}$$

and the energy balance equation

$$k_B\frac{\partial T}{\partial t}+\nabla\cdot\left(\theta\vec{v}\right)-\frac{k_B\chi}{c_v}\nabla^2T-\frac{m^*\gamma}{2c_v}\left(\frac{\partial v_i}{\partial x_j}+\frac{\partial v_j}{\partial x_i}-\delta_{ij}\frac{\partial v_k}{\partial x_k}\right)^2=\frac{1}{c_v}\left(\frac{\partial W}{\partial t}\right)_c+\frac{mv^2}{c_v\tau}\,. \tag{10}$$

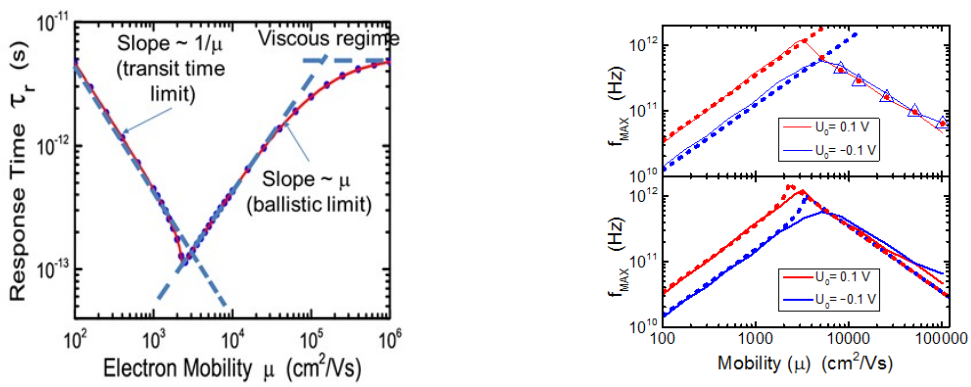
Here c_v is the heat capacity, χ/c_v is the thermometric conductivity, and the electron energy is the sum of the internal energy and the drift energy $m^*v^2/2$, and γ is the viscosity of the 2DEG. The linear analysis of these equations using the unified charge control model (UCCM) leads to the following equations for the ultimate response time τ and plasma wave velocity $s = (s_p^2 + s_{ac}^2)^{1/2}$:

$$\frac{1}{\tau} = Re\frac{1}{2}\left[-\left(\frac{1}{\tau_m}+\frac{\pi^2\gamma}{4L^2}\right)\pm\sqrt{\left(\frac{1}{\tau_m}+\frac{\pi^2\gamma}{4L^2}\right)^2-\frac{\pi^2 s^2}{L^2}}\right] \quad , \tag{11}$$

$$s_p^2 = \frac{\eta k_B T}{m^*}(1+\exp\left(-\frac{qU_{gt}}{\eta k_B T}\right)\ln\left[1+\exp\left(-\frac{qU_{gt}}{\eta k_B T}\right)\right], \tag{12}$$

$$s_{ac}^2 = \frac{E_F}{m^*(1-\exp(-E_F/k_B T))}. \tag{13}$$

Here s_p and s_{ac} are the electric and acoustic components of the plasma wave velocity. Above threshold (when the gate voltage swing $U_{gt} >> k_B T/q$), the expression for the plasma velocity simplifies to become $s = qU_{gt}/m^*$. The estimated viscosity of the 2DEG is $\gamma \sim 15$ cm^2/s (comparable to that of castor oil or glycerin at room temperature) [8]. The analysis of eq. (11) reveals three distinguished regimes: (1) collision dominated transport; (2) ballistic transport, and (3) viscous transport (see Table 1). A more detailed analyses accounting on the differences between the momentum relaxation time and energy relaxation time and their dependence on energy will be presented elsewhere. Figure 5 showing the computed response time and the corresponding maximum modulation frequency illustrates the most important conclusions of the hydrodynamic model.



(a) (b)

Figure 5. The computed response time (a) [9] ©IEEE (2016) and the corresponding maximum modulation frequency (b) (from [2]

TABLE I. Transport regimes of 2DEG

Regime	Takes place when	Predicted response time
Collision dominated transport	$\frac{1}{\tau_m^2} >> \frac{\pi^2 s^2}{L^2}$	$\frac{1}{\tau} = \frac{\pi^2 qU_{gt}\tau_m}{4L^2 m^*} = \frac{\pi^2 \mu U_{gt}}{4L^2}$
Ballistic transport	$\frac{1}{\tau_m^2} << \frac{\pi^2 s^2}{L^2}$	$\frac{1}{\tau} = \frac{2}{\tau_m}$
Viscous transport	$\frac{1}{\tau_m} << \frac{\pi^2\gamma}{8L^2}$	$\frac{1}{\tau} = \frac{\pi^2\gamma}{8L^2}$

4. SUMMARY OF EXPERIMENTAL RESULTS

The experimental study of the ultrafast response was presented in [10]. It was using a new technique based on the termination of the response by flooding the transistor by the electron hole plasma generated by the band-to-band optical pulse. These measurements confirmed the results of the hydrodynamic modeling predicting the ultra-fast transistor plasmonic response at the time scale much shorter than the electron transit time and revealed a large sensitivity enhancement (more that in 7 times) due to the constructive interference of the impinging THz pulse and the optical pulse field rectified by the transistor nonlinearity.

5. CONCLUSIONS

Our analytical estimates and hydrodynamic simulations show that the electron response becomes faster with the mobility increase only for relatively low mobility values and/or relatively large feature sizes. For large mobility values, the plasmonic ringing determines the characteristic response time and, eventually, the dominant effect of the electron viscosity determine the response time. The minimum response time and the maximum device modulation frequency correspond to the sub picosecond and terahertz ranges, respectively. The recent experiments of the FET switching using femtosecond optical laser pulses are in good agreement with the predicted sub picosecond switching times and demonstrate a larger sensitivity enhancement due to the constructive interference of the impinging THz pulse and the optical pulse field rectified by the device nonlinearity.

6. ACKNOWLEDGMENTS

The work at RPI (M. Shur) was supported in part by the U.S. Army Research Laboratory through the Collaborative Research Alliance (CRA) for Multi-Scale Modeling of Electronic Materials (MSME).

REFERENCES

[1] M. S. Shur and L. F. Eastman, Ballistic Transport in Semiconductors at Low-Temperatures for Low Power High Speed Logic, IEEE Transactions Electron Devices, Vol. ED-26, No. 11, pp. 1677-1683, November (1979)

[2] G. Rupper, S. Rudin, and M. Shur, Response of Plasmonic Terahertz Detectors to Amplitude Modulated Signals. Solid State Electronics, Volume 111, September 2015, Pages 76–79 (2015)

[3] S. Rudin, G. Rupper, M. Shur, Ultimate Response Time of High Electron Mobility Transistors, Journal of Applied Physics 117, 174502 (2015); doi: 10.1063/1.4919706.

[4] S. Rudin, G. Rupper, A. Gutin, and M. Shur, Theoretical and experimental studies of response of plasmonic terahertz detector to large signals. J. Appl. Phys. 115, 064503 (2014)

[5] M. Shur, Introduction to Electronic Devices, Wiley (1998)

[6] M. I. Dyakonov and M. S. Shur, Detection, Mixing, and Frequency Multiplication of Terahertz Radiation by Two Dimensional Electronic Fluid, IEEE Transactions on Electron Devices, Vol. 43, No. 3, pp. 380-387, March (1996)

[7] V. Yu. Kachorovskii and M. S. Shur, Field effect transistor as ultrafast tunable detector of terahertz radiation, Solid State Electronics, Vol 52/2 pp. 182-185, February (2008)

[8] M. Dyakonov and M. S. Shur, Shallow Water Analogy for A. Ballistic Field Effect Transistor. New Mechanism of Plasma Wave Generation by DC Current. Phys. Rev. Lett. Vol. 71, No. 15, pp. 2465-2468, Oct. 11 (1993)

[9] M. Shur, A. Muraviev, G. Rupper and S. Rudin, "THz pulse detection by photoconductive plasmonic high electron mobility transistor with enhanced sensitivity," 2016 74th Annual Device Research Conference (DRC), Newark, DE, 2016, pp. 1-2. doi: 10.1109/DRC.2016.7548489

[10] A. Muraviev, A. Gutin, G. Rupper, S. Rudin, X. Shen, M. Yamaguchi, G. Aizin and M.S. Shur, Sub-picosecond detection of terahertz radiation by plasmonic field effect transistors, Optics Express, 24(12), 12730-12739 (2016)