Plasma Enhanced Terahertz Rectification and Noise in InGaAs HEMTs

Javier Mateos Member IEEE, Tomás González Senior Member IEEE

Abstract—In this work we explore high frequency collective phenomena present in InGaAs HEMTs which lead to a peak in the current noise spectrum and enhance their DC response to THz signals, thus originating a resonance in the rectification of AC signals. These phenomena have been evidenced in recent experiments, in which THz detection as a result of plasma wave resonances has been demonstrated. In this paper, by means of Monte Carlo simulations, the noise spectra and the AC to DC rectification properties of the devices have been calculated and linked to the properties of the plasma oscillations within two distinct parts of the source-gate region: the capped and the recessed sections of the channel.

Index Terms—Monte Carlo simulation, plasma oscillations, THz detection, InGaAs HEMTs

I. INTRODUCTION

The range of the electromagnetic spectrum, sandwiched between microwaves and infrared light, with frequencies between 300 GHz and 3 THz (sometimes called T-rays), has gained scientific and commercial interest in recent years due to its broad range of potential applications, which were previously restricted to radioastronomy and high resolution spectroscopy [1]. The fast development of new technologies, such as pump-probe time domain THz spectroscopy with ultrafast lasers [2] or ultra-high frequency MMICs based on aggressively scaled InGaAs HEMTs and HBTs [3], [4], is making practically available THz systems working at room temperature. The applications which are presently being focused are mainly security-screening tools [5], [6], but also ultra high speed communications and medical imaging (without the damage produced by ionizing radiation such as X-rays). THz sources must be obtained at the limits of electronics from one side and optical systems from the other, resulting in a lack of efficient and practical radiation sources, which most of the times need bulky (or low temperature) equipment. In the last years, the search for more powerful alternatives to the frequency multiplication of electronic sources or the photomixing of optical beams has come to the possibility of exploiting the use of plasma resonances in semiconductor devices in order to fabricate compact room temperature THz sources and detectors. This approach has been confirmed by recent experiments made with classic High Electron Mobility Transistors (HEMTs) based on different material systems (GaAs, InGaAs and GaN), and Si MOSFETS, where THz detection as a result of plasma wave resonances has been demonstrated [7], [8]. Also, some evidence of plasma related THz emission has been found [9], [10].

Up to now, only ideally simple 1D theoretical models, not considering the complex geometry and electron dynamics in HEMTs, have been used for predicting the different types of plasma resonances that can appear in semiconductor devices [7], [11]-[13]. In this work we try to explain the mechanism of plasma-resonant THz detection with HEMTs by means of 2D Monte Carlo (MC) simulations self-consistently coupled with a Poisson solver [14]-[16]. This technique is able to consider the real geometry and layer structure of the HEMTs and provide not only static results but also identify the effect of collective phenomena such as plasma oscillations on both the current noise and the AC to DC rectification properties of the devices. The in-house MC simulator has been already validated through comparison with experimental results obtained in similar devices and concerning static characteristics, small-signal and noise parameters [10], [14].

In previous works [17], [18], MC simulations of an extremely simple device, the so called self-switching diode (SSD), have confirmed the presence of a peak in the current noise spectral density with a dependence on the geometry similar to that exhibited by an analogous peak observed in the AC to DC rectification. These are two different evidences of the same internal phenomena: plasma resonances. In this work, we will show, by means of MC simulations, how this very same relationship between AC to DC rectification and noise linked to plasma oscillations holds also in classical InGaAs based HEMTs, where the phenomena are even more complex due to the presence of a third electrode. In both cases, SSDs and HEMTs, the plasma-related charge fluctuations in regions where the field effects limit the current flowing through the devices establish a link between the noise in the current (given by the plasma oscillations) and the AC to DC rectification (given by the coupling between the internal plasma and the external DC output signal). This work summarizes and completes the theoretical interpretation of the preliminary results presented by the authors in two prior conference papers [15], [16].

II. MONTE CARLO MODEL AND DEVICES UNDER STUDY

The most appropriate numerical technique to deal with room temperature nanometer-size devices where electron transport is potentially ballistic is the MC method, able to provide static results and also the dynamic and noise behavior of the
structures. Thus, to model the devices we make use of a 2D MC, consisting of a semiclassical ensemble MC simulator self-consistently coupled with a 2D Poisson solver. The MC transport description is based on a 3-valley model (Γ-L-X nonparabolic spherical valleys for both InGaAs and AlInAs) and includes ionized impurity, alloy, polar and non-polar optical phonon, acoustic phonon and intervalley scattering, and appropriate carrier injection techniques at the contacts. Details about the model can be found in [14].

The geometry of the simulated 80 nm T-gate InGaAs HEMTs is shown in Fig. 1 and is based on that already used in [10]. This recessed topology is similar to that commonly used in fabricated devices. The length of the source, recess and drain regions (Ls, Lr and Ld, respectively) will be taken as variable parameters in the MC simulations to understand their influence on the plasma effects. Surface charges at the semiconductor-air interfaces, σ, are included in a simple way by fixing the value of σ to the experimentally-extracted equilibrium value, keeping it constant independently of the topology of the structure, bias and time [14]. Simulations are performed at T=300 K.

In order to model the THz detection mechanism (AC to DC conversion), as in [12], a sinusoidal signal of varying frequency is superimposed to the DC gate bias, VGS, and the average drain current is recorded. The detection of the AC signal is therefore accounted for as an increment of the drain current, ΔI. On the other hand, noise calculations are carried out in the standard way. First, from the time series of instantaneous current values obtained from the simulation we evaluate the autocorrelation function of current fluctuations. Then, by Fourier transform, the corresponding spectral density is calculated. To this end, the dynamics of electrons in the device have been simulated during time series of 500,000 steps of 1 fs for each bias point.

III. PHYSICS OF PLASMA-RESONANT THz DETECTION

Initially the results for the AC to DC response will be presented. Fig. 2 shows the variation of the drain current with respect to its static value, ΔI, as a function of the frequency of a sinusoidal signal (of amplitude 0.1 V) superimposed to different DC values of VGS in a HEMT with Ls=200 nm, Lr=100 nm and Ld=0.5 μm (the different curves have been shifted by 20 A/m each to enhance the clarity of the figure). As observed, ΔI shows a resonant peak (mainly for VGS near pinch off) for a frequency around 2.5 THz, in good agreement with similar experimental measurements [19], [20]. In Secs. V and VI, the origin and frequency of this peak will be explained in terms of different plasma modes that are possible in the HEMT.

In order to perform the comparison between the noise of a device and its response to oscillating signals, it must be bear in mind that the spectral density of current fluctuations at equilibrium is related to its small signal behavior [21], whereas the DC response current is caused by the device rectification and corresponds to large signal (non-linear) operating conditions. As such, the amplitude of the signal superimposed to the DC gate voltage has been kept to a low value of 0.1 V (but high enough to produce a significant response for ΔI).

Figure 2: Average drain current variation originated by the superposition of a sinusoidal signal of 0.1 V of amplitude and varying frequency to the DC gate potential (VGS) of a HEMT with Ls=200 nm, Lr=100 nm and Ld=0.5 μm. The different curves have been shifted by 20 A/m each. Vφ=1.0 V.
The resonance between the plasma oscillations in the source-gate region of the HEMT and the excitation applied to the gate voltage is clearly visible in the time evolution of the potential and carrier concentration profiles obtained with the MC simulations, shown in Fig. 3. Figs. 3(a) and (b) show that at low frequency the potential and electron concentration under the gate are in phase (a lower gate potential induces a lower electron concentration). But at the frequency of the resonance, Figs. 3(c) and (d), due to the plasma oscillations within the source-gate region, the injection of the electrons under the gate is in opposite phase to the gate potential variations. As a consequence, when the gate potential decreases the electron concentration is higher, thus increasing the drain current over its stationary value. This effect is mainly visible at gate voltages near pinch off, since it is the non-linearity of the transfer curve which provides a net positive variation of drain current when such an “opposite phase” evolution of potential and electron concentration arises. Finally at very high frequencies the electron concentration does not respond to the changes in the potential profile, Figs. 3(e) and (f), thus leading to a zero value of $\Delta \rho_c$.

IV. LINK BETWEEN NOISE AND RECTIFICATION

For having a deeper insight into the origin of the plasma-resonant detection we have performed simulations of HEMTs with different geometries for $V_{gs}=-0.3$ V, the value of gate bias that provides the optimum AC to DC response. Fig. 4(a) shows the results for the HEMT previously analyzed in Fig. 2 compared with those obtained for devices with a longer gate of 0.2 \mu m and a longer drain region of 1 \mu m. It can be observed that the frequency of the resonant peak does not depend on the gate length, $L_g$, nor on the recess-drain region, $L_d$. On the contrary, the resonant frequency strongly depends on the lengths of the recess-source region, $L_r$, Fig. 4(b), of the source side of the recess, $L_s$, Fig. 4(c). Moreover, the resonant frequency variation with $L_r$ is stronger than with $L_s$; it increases from 1.5 to more than 3 THz when reducing the value of $L_r$ from 300 nm to 10 nm. These results clearly indicate that the source-gate region (including the recess) acts as the plasma-wave cavity that produces the resonant detection of THz radiation in HEMTs. This fact is confirmed by the comparison of the frequency dependence of the source, gate and drain current noise spectra, $S_{is}$, $S_{ig}$ and $S_{id}$, respectively, with that of $\Delta I_d(f)$ shown in Fig. 5 for a HEMT with $L_r=100$ nm and $L_d=0.5$ \mu m and different values of $L_s$. Fig. 5 evidences the presence of a peak in all the three $S_{is}$, $S_{ig}$ and $S_{id}$, analogous to that observed in the case of the rectifying response. However, the dependence of the frequency of the peak in the noise spectra with the length of the source region $L_s$ is analogous to that observed for $\Delta I_d(f)$ only for the case of $S_{is}$, and not for $S_{ig}$ and $S_{id}$. This happens because it is the resonance between the plasma-related charge fluctuations of the source-gate region and the oscillating gate voltage that results in $\Delta I_d$, due to an enhanced injection of electrons over the potential barrier induced by the gate electrode. In fact, the very same characteristic frequency is also visible in the spectra of the electric potential fluctuations in the channel under the source side of the gate (not shown here), which is the point that determines the value of $I_d$, since it is where the carrier injection over the potential barrier takes place.
As can be expected, the peak of $S_{Is}$ observed in Fig. 5(c) is related to the electron dynamics within the drain region and its position depends on its length $L_{d}$ (results not shown here) in the same way as $S_{Is}$ depends on $L_{s}$. However, as explained above, such plasma oscillations in the drain region do not couple to $\Delta I_{d}(f)$ as the source ones do.

It is also remarkable the degradation of the quality factor of the resonant peak observed in Fig. 4(b) as the length of the source region is increased (the FWHM, Full Width at Half Maximum, of the peaks increases), which goes in parallel to the broadening of the peak of the source current noise. This is also observed when the recess region is enlarged, Fig. 4(c), and it is due to the more efficient damping of the plasma resonance by scattering as the size of the resonant cavity increases.

A further confirmation is provided by the results of Fig. 6, which shows the bias dependence of the frequency of the resonant peak of $\Delta I_{d}(f)$ compared with those of the peaks observed in the source, gate and drain current noise spectra for different values of $L_{s}$. As observed, the frequency of the resonant peak in both $\Delta I_{d}(f)$ and $S_{Is}$ shifts towards lower values as $V_{gs}$ is increased (as already seen in Fig. 2). The variation with $L_{s}$ is also coincident, thus confirming the close relationship between source current noise and AC to DC rectification. On the contrary, the peak frequency of $S_{Id}$ does not depend on the source-gate distance, since it is mainly affected by the drain region of the device. The behaviour of the peak in $S_{Id}$ also somewhat follows that of the plasma resonant rectification, but not as closely as that of $S_{Is}$, mainly for the longest values of $L_{s}$ [Fig. 6(c)].
Plasma Enhanced Terahertz Rectification and Noise in InGaAs HEMTs - revised

V. 2D PLASMA FREQUENCIES IN HEMTS

Dyakonov-Shur theory predicts the values of the plasma frequencies that should appear when mobile charges with effective mass \( m^* \) are distributed in an ideally thin channel sandwiched by a dielectric material with permittivity \( \varepsilon_{\text{die}} \) (2D case), and when a gate contact is placed at a distance \( d \) from the channel (gated 2D case); see for example [22]. Their values are, respectively:

\[
f_{p} = \frac{1}{2\pi} \frac{e^2 n_{2D} k}{2m^* \varepsilon_{\text{die}} \varepsilon_0} \quad (1)
\]

\[
f_{p} = \frac{1}{2\pi} \frac{e^2 n_{2D} d}{m^* \varepsilon_{\text{die}} \varepsilon_0} \quad (2)
\]

\( k \) being the plasma wavevector, \( \varepsilon_0 \) the vacuum permittivity and \( n_{2D} \) the sheet electron density. The mechanisms at the origin of plasma oscillations in a sheet of electrons are analogous to those of waves in fluids, since both can be represented by hydrodynamic equations. Indeed, these cases of 2D and gated 2D plasma oscillations can be understood as analogous to deep water and shallow water waves, respectively.

Due to the complex geometry of the simulated HEMT and the different plasma modes that may appear in the source-gate region, sketched in Fig. 7, there is not a unique ideal plasma model for explaining the MC results (mainly because the source-gate region is made up by the series combination of the source and recess regions, with lengths \( L_S \) and \( L_R \), respectively).

For an ideally long source region, the plasma resonance (and the peak in \( S_H \)) should practically not depend on the gate voltage, since \( V_{gs} \) does not in principle affect the electron concentration in that region. This is observed in Fig. 6(c), in which \( L_S=400 \) nm and the resonant frequency hardly changes with \( V_{gs} \). However, when the source length is shortened, the lateral depletion induced by the gate voltage can significantly modify the size of the plasma cavity, thus appreciably modulating the resonant frequency. A higher gate potential reduces the depletion, so that the effective length \( L_S+L_R \) is increased and the plasma frequency decreases. This effect is more evident for the shortest source regions, Fig. 6(b) for \( L_S=200 \) nm and mainly Fig. 6(a) for \( L_S=100 \) nm.

V. 2D PLASMA FREQUENCIES IN HEMTS

Dyakonov-Shur theory predicts the values of the plasma frequencies that should appear when mobile charges with effective mass \( m^* \) are distributed in an ideally thin channel sandwiched by a dielectric material with permittivity \( \varepsilon_{\text{die}} \) (2D case), and when a gate contact is placed at a distance \( d \) from the channel (gated 2D case); see for example [22]. Their values are, respectively:

\[
f_{p} = \frac{1}{2\pi} \frac{e^2 n_{2D} k}{2m^* \varepsilon_{\text{die}} \varepsilon_0} \quad (1)
\]

\[
f_{p} = \frac{1}{2\pi} \frac{e^2 n_{2D} d}{m^* \varepsilon_{\text{die}} \varepsilon_0} \quad (2)
\]

\( k \) being the plasma wavevector, \( \varepsilon_0 \) the vacuum permittivity and \( n_{2D} \) the sheet electron density. The mechanisms at the origin of plasma oscillations in a sheet of electrons are analogous to those of waves in fluids, since both can be represented by hydrodynamic equations. Indeed, these cases of 2D and gated 2D plasma oscillations can be understood as analogous to deep water and shallow water waves, respectively.

Due to the complex geometry of the simulated HEMT and the different plasma modes that may appear in the source-gate region, sketched in Fig. 7, there is not a unique ideal plasma model for explaining the MC results (mainly because the source-gate region is made up by the series combination of the source and recess regions, with lengths \( L_S \) and \( L_R \), respectively).

For an ideally long source region, the plasma resonance (and the peak in \( S_H \)) should practically not depend on the gate voltage, since \( V_{gs} \) does not in principle affect the electron concentration in that region. This is observed in Fig. 6(c), in which \( L_S=400 \) nm and the resonant frequency hardly changes with \( V_{gs} \). However, when the source length is shortened, the lateral depletion induced by the gate voltage can significantly modify the size of the plasma cavity, thus appreciably modulating the resonant frequency. A higher gate potential reduces the depletion, so that the effective length \( L_S+L_R \) is increased and the plasma frequency decreases. This effect is more evident for the shortest source regions, Fig. 6(b) for \( L_S=200 \) nm and mainly Fig. 6(a) for \( L_S=100 \) nm.
Moreover, there are two possible ungated plasma modes (both on the source and the recess regions): on the one hand one can consider that electromagnetic coupling takes place through the semiconductor forming the barrier itself (AlInAs, with $\varepsilon_{AlInAs}=13.88$), on the other hand there is also the possibility that it takes place at a longer scale and therefore one has to consider the external dielectric (in this case SiN, typically used for the passivation of HEMTs, with $\varepsilon_{SiN}=6$). Also, in addition to the gated plasma mode that could come out within the recess region due to the presence of the gate hat (with $d=167$ nm), as suggested in [7], the highly-doped source cap region acts as a quasi-metallic constant voltage region, thus possibly generating a 2D gated plasma in that part of the device (with $d=17$ nm). In Fig. 8 the theoretical dependences of the different plasma modes [Eqs. (1) and (2)] are plotted using the following criteria:

- The value of $k=\pi/2L$, with $L=L_s+L_r$, has been used because: i) as explained before, the plasma-wave cavity corresponds to the source-gate region, and ii) as shown in Fig. 3, the boundary conditions for the electric potential variations is a fixed value at the left and oscillating potential at the right (nearly open circuit, since the channel is almost pinched off). Therefore the source-gate region corresponds to a $k/4$ distance and as such $k$ takes the value $\pi/2(L_s+L_r)$.

- The values of $n_{2D}$ in the channel are different in the source region and under the recess. We have used here the values obtained from the MC simulations (see Fig. 3), of $3.1\times10^{12}$ and $1.8\times10^{12}$ cm$^{-2}$, respectively.

- For the gated 2D plasma at the recess region we have used $d=167$ nm (channel-gate hat distance) while at the source $d=17$ nm (channel-cap distance).

The comparison between the ideal dependences with those obtained from MC simulations shown in Fig. 8 allows extracting interesting information. The dependence of the resonant frequency when varying $L_s$ is similar to the $1/L$ behavior of the 2D plasma case, while when varying $L_r$, the dependence approaches the $1/L$ trend characteristic of the gated 2D plasma. These dependences seem to be in agreement with the geometry of the HEMT, since the recess is covered by the gate contact while the source region is not. However, this contradicts the fact that nearly the same dependencies on $L_s$ and $L_r$ are obtained with MC simulations performed without considering the T-shape of the gate electrode (shown in Fig. 8 by open symbols), where the possibility of the 2D-gated mode in the recess region is cancelled. Moreover, the 2D-gated mode within the source region, caused by the cap acting as a virtual gate, seems to be necessary to explain the MC results (in good agreement with the results of the hydrodynamic model presented in Ref. [7]), since it is the only one providing plasma frequencies below the values obtained from the MC simulations. Therefore, the combination of this mode with that of the recess region is the more plausible possibility to originate the resonances shown in the MC simulations. Regarding the recess region, both coupling possibilities (via the AlInAs barrier and the external dielectric) are considered, even if the latter one provides excessively high values of the ungated 2D plasma mode. The exact way the plasma frequencies of the source and recess regions combine themselves to generate the overall plasma-resonant detection in HEMTs will be explored in the next section.

### VI. Series Combination of Plasma Modes

The previous MC results indicate that the origin of the resonant AC to DC conversion is a series combination of plasma modes in the source and recess regions: i) a gated 2D plasma mode in the source region as a result of the coupling between the channel and the cap layer and ii) a 2D plasma mode within the recess region.

The case of a series connection of two regions with different plasma frequencies has already been studied in the simple case of $n'n''$ diodes, where the 3D plasma frequency of the $n'$ regions is different from that of the $n$ region [23]. In that case the overall peak observed in the current fluctuations appears at an intermediate frequency lying between the 3D plasma values of the two regions considered independently. A phenomenological analytical formula able to predict the global plasma frequency, $f_{res}$, as a function of the length of each of the two regions has been suggested [24]. We will apply that formula to our case in this way:

$$f^2_{res} = f^2_p \frac{L_r}{L_s + L_r} + f^2_p \frac{L_s}{L_s + L_r}$$

where $f_p$ and $f_{ps}$ being the plasma frequencies corresponding to the source and recess regions, respectively.

Eq. (3) predicts that if $L_s=10$ nm then $f_p \rightarrow f_{ps}$ and if $L_s=50$ nm then $f_p \rightarrow f_p$. In order to check this behavior we have performed simulations of those extreme cases: i) with $L_r=10$ nm and variable $L_s$ and ii) $L_s=50$ nm and variable $L_r$. In Fig. 9 we compare these results with the ideal 2D plasma mode in the source region and the two possible 2D plasma modes within the recess region (following the previous discussion the ungated 2D mode in the source region and the gated 2D in the recess have been disregarded). In the case when $L_s<<L_r$, the MC results are close to $f_p$, but do not follow the expected gated 2D dependence. This discrepancy is attributed to the fact that being $L_s=50$ nm, such short cap is not able to completely fix the potential, so that the “virtual gate” is not effective. In the case when $L_s>>L_r$ the results are very close to $f_{ps}$ considering the ungated 2D coupling short recess region, $L_r=10$ nm). From this last result one could expect that the dielectric outside the semiconductor has no effect on the plasma resonant detection. In order to verify its influence, we have performed MC simulations with different dielectric materials placed at the top of the device. The typical SiN
the analytical formulas given in eqs. (2) and (1), corresponding changed by \( \epsilon \) (which has been considered in all the previous cases) has been plays also a significant role (mainly for long recess lengths, Fig. 9) since the dielectric outside the devices result of the coupling between the channel and the source side (20 nm), which can be understood as a border limit between the two regions has been slightly shifted towards MC case. In order to reach the observed good agreement, the been used, so that the comparison should be made with that sufficiently long values of \( L_s \) and constant \( L_r \), the efficiency of the cap layer acting as a “virtual gate” (thus leading to 2D-gated plasma oscillations) and the recess region, where an un gated 2D plasma mode arises. The combination of the two modes has been explained through a phenomenological formula, previously derived for n nm diodes.

VII. CONCLUSIONS

By means of MC simulations we have described the mechanism for the plasma-resonant THz detection observed in HEMTs and linked it to the noise in the source current. Simulations show that the peak of the current noise spectra and the resonant frequency of the AC to DC conversion depend similarly on the topology of the devices and are a consequence of plasma oscillations of free charge in the source-gate region of the HEMT. This peak in the THz detection observed in HEMTs is a result of a combination of the plasma modes of two distinct regions, the source region, with the cap layer acting as a “virtual gate” (thus leading to 2D-gated plasma oscillations) and the recess region, where an un gated 2D plasma mode arises. The combination of the two modes has been explained through a phenomenological formula, previously derived for n nm diodes.

REFERENCES


