SISPAD 2012, September 5-7, 2012, Denver, CO, USA

# Carrier Dynamics Study of Lateral Scaling and the Limiting High-Frequency Performance of GaN-HEMTs

Riccardo Soligo,
Diego Guerra,
David K. Ferry, *Life Fellow, IEEE,*
Stephen M. Goodnick, *Fellow, IEEE,*
and Marco Saraniti, *Member, IEEE.*
School of Electrical, Computer and Energy Engineering,
Arizona State University, Tempe, AZ 85287 $USA$.
(e-mail: rsoligo@asu.edu)

***Abstract*—The effects of access region scaling on the RF performance of millimeter-wave GaN HEMTs is investigated through full band Cellular Monte Carlo simulation. The nanoscale carrier dynamics under the gate controlled region is simulated in devices with different access region lengths in relation with their cut-off frequency. In particular, the cut-off frequency is shown to increase monotonically up to $860\ GHz$ by symmetrically scaling the source to gate and gate to drain distance from $635\ nm$ to $50\ nm$. The electron scattering rates have been studied showing that while polar phonon emission is the overall dominant scattering mechanism, the emission of acoustic phonons is greatly enhanced in shorter devices. By scaling the gate length and the access region at the same time, we find that the cut-off frequency increases further. Moreover, for vanishing access regions, we are able to calculate a maximum velocity, while a limit effective gate length has been obtained for the physical gate length approaching zero. Based on these limits, we calculate the transit time and find a limiting cut-off frequency of $1.49\ THz$ for the GaN HEMT studied in this work.**

***Index Terms*—HEMT, GaN, High Frequency, Scaling, Ultimate Frequency, Monte Carlo, Numerical Simulation, Transit Velocity.**

## I. Introduction

GaN HEMTs are nowadays the most suitable candidates for high-power, high-frequency applications. In fact, the wide bandgap of GaN leads to large breakdown voltages and the possibility for GaN based power devices to operate over a broad range of voltages. A record output power of $40\ W/mm$ was reported in a one micron gate length GaN-HEMT [1]. At the same time, the high electron mobility of the AlGaN/GaN heterojunction enables nanoscale gate GaN HEMTs to reach cutoff frequency ($f_T$) as high as $370\ GHz$ [2] and maximum oscillation frequency, $f_{MAX}$, of $400\ GHz$ [3]. The interest in faster devices has recently led to the development of many techniques to improve the RF performance. Some of them rely on fabrication and material/interface quality improvement, like regrown ohmic contacts [4], or self-aligned gates [5]. However, the main way to increase the HEMTs $f_T$ is through reduction of the gate length. Guerra *et al.* [6] have shown that the thickness of the barrier has to be reduced when downscaling the gate to limit short channel effects. Gate scaling is thus limited by the thickness of the barrier due to increase in the gate leakage current, and depletion of the channel.

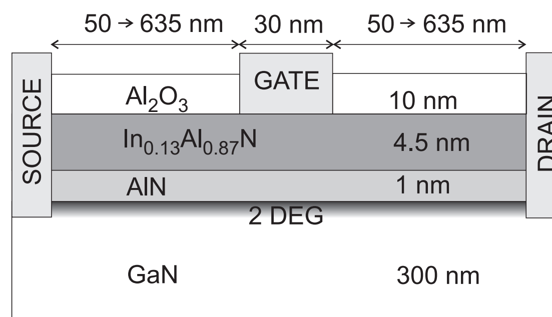


Fig. 1. Generic schematic cross-section of the $GaN$ HEMT simulation domain.

Recently, the scaling of the source to gate and gate to drain length on nitride HEMTs has drawn some attention [7]. In particular, Shinohara *et al.* [8] reported the record $f_T$ for GaN based HEMTs in a deeply scaled self aligned gate device. Several efforts to optimize HEMTs design for high RF performance have been made using accurate simulation tools like full band Monte Carlo. In particular, a physical definition of $f_T$ based on the transit time of electrons under the effective gate length was given by Akis *et al.* [9] that provides a physical interpretation of the dependence between $f_T$ and the gate length. The same definition of $f_T$ was later used by Guerra *et al.* [10] to understand the influence of the passivation dielectric on the RF performance by studying the electron velocity profile under the gate. Moreover, based on the transit time, an estimation of the limit $f_T$ has been performed by Foutz *et al.* [11]. This limit was obtained calculating the transit time across the metallurgic gate length assuming the electrons velocity to be equal to the maximum overshoot velocity recorded in bulk. However, this method assumes an


ISBN 978-0-615-71756-2

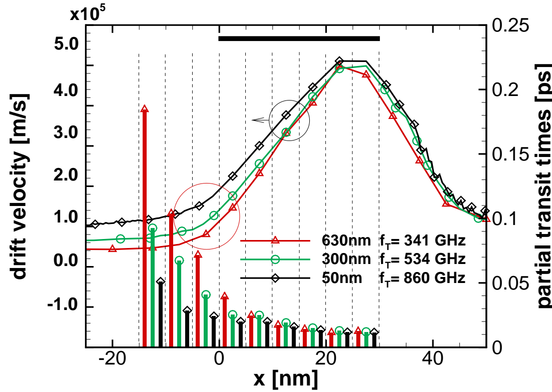


Fig. 2. Electron velocity profile along the effective gate length for 3 Lsg,Lgd with the relative cut off frequency indicated. Histograms represents the partial transit time for the sections separated with the dashed lines. The gate stars at zero and is indicated by the black rectangle. Bias: $V_D = 5\ V$, $V_G = -2\ V$.

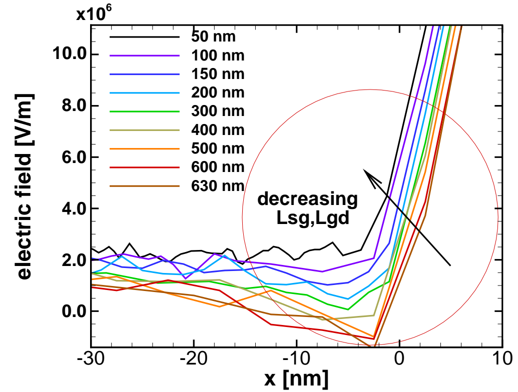


Fig. 3. Electric filed at the gate end of the source access region. The red circle highlights the correspondent area in Fig. 2. Gate start is set to zero. Bias: $V_D = 5\ V$, $V_G = -2\ V$.

abrupt discontinuity of the electric field between the source region and the gate, which is far from what happens in reality. In particular, it has been shown that the velocity transient of the electron entering the gate region is the main factor in determining the device frequency response [10].

In the present work, we show the effect of access region scaling on the carrier dynamic and on the RF performances in Section II in a state-of-art GaN HEMT. In Section II, we extend these results simultaneously scaling the gate length and the access regions. Finally, in Section III, we introduce a novel technique to determine the limiting $f_T$ of nitride HEMTs based on the extrapolation of the transit time for source to gate, gate to drain and the metallurgic gate length approaching zero. The approach adopted provides quantitative guidelines to improve the frequency performance through horizontal scaling.

## II. RF PERFORMANCE ANALYSIS

The layout of the GaN HEMT analyzed in this work is shown in Fig. 1. This device, proposed Lee *et al.* [12], has two symmetrical access regions of 635 $nm$ and a gate length of 30 $nm$. The match of the simulated $Id{-}Vd$, obtained with our full band cellular Monte Carlo [13], with experimental device measurements has been reported in a previous work by our group [14], showing a good agreement except for high values of current. This is due to self heating, which is neglected in the present work, resulting in a small overestimation of the calculated current.

In order to study the effects of the access regions, the source to gate distance, $L_{sg}$, and gate to drain distance, $L_{gd}$, are symmetrically downscaled from 635 $nm$ to 50 $nm$ keeping the gate length constant. In particular, we analyze the effect of the scaling on the cut-off frequency, $f_T$. Multisinusoidal simulation in which the input signal consists of a combination of several sinusoids, allows us to efficiently calculate $f_T$ fitting the short-circuit current-gain with a $-20\ dB/dec$ line and extrapolating the unit gain frequency. The simulation results show that $f_T$ monotonically increases reducing $L_{sg}$ and $L_{gd}$. In order to investigate the reason for this trend, we compare the profile of the electron velocity under the gate for different access regions as shown in Fig. 2. It may be noted that the three curves are relatively close one another, although the $f_T$ values reported next to the legend in the same plot differ by as much as a factor of three. However, we can see that the main difference is on the source side of the velocity profile while the area near the peak is not affected by the scaling. The implication of the last observation on the $f_T$ is clarified by the transit time histograms that are superimposed on the velocity curves in Fig. 2. In particular, the height of each column represents the average transit time that electrons take to go through the 5 $nm$ sections delimited by each pair of dashed lines for the three different access region devices. One notes that the first section does not actually correspond to the beginning of the metallurgical gate. This is because the gate control area, called effective gate length ($L_{eff}$), extends beyond the geometrical projection of the gate due to the presence of fringing capacitances. Moreover, the longest partial times are on the source side of $L_{eff}$, where the velocity of the electrons is lower because they have just begun to accelerate toward the peak velocity at the drain side of the gate. Comparing the three partial transit times in the first section, we can see that this is the quantity that is most affected by the access region scaling. In particular, electrons in the 600 $nm$ $L_{sg}$,$L_{gd}$ device take four time as much time as in the 50 $nm$ one to go across the first 5 $nm$ of the effective gate length. Considering that $f_T$ is inversely proportional to the total transit time, *i.e.* the summation of all partial transit times, we conclude that the reduction of the access regions improves the frequency response of the device by increasing the velocity of the electrons as they enter $L_{eff}$, which are indeed the most dominant terms in the summation.

In Fig. 3 the component of the electric field along the transport direction is shown in the area of interest highlighted by a red circle in Fig. 2, located at the source side of $L_{eff}$. It may be noticed that the electric field is increased by reducing $L_{sg}$ and $L_{gd}$, which explains the higher electron velocity in the source access region.

Figures 4 and 5 show the effect of access region scaling

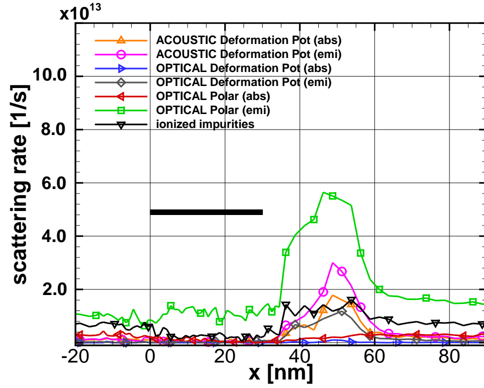


Fig. 4. Rates of the main scattering mechanisms for 600nm Lsg,Lgd device. The gate starts at 0 and is indicated by a black rectangle. Bias: $V_D = 5\ V$, $V_G = -2\ V$.

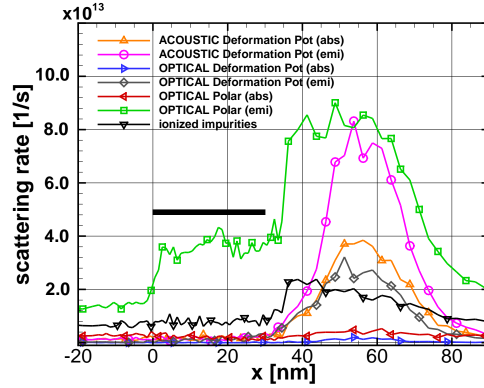


Fig. 5. Rates of the main scattering mechanisms for 150nm Lsg,Lgd device. The gate starts at 0 and is indicated by a black rectangle. Bias: $V_D = 5\ V$, $V_G = -2\ V$.

on the scattering rates, where Fig. 4 correspond to a 600 $nm$ $L_{sg}$,$L_{gd}$, while Fig. 5 correspond to a shorter 150 $nm$ length structure. It is apparent that all scattering mechanisms are enhanced in the shorter device. In both figures, the highest rates are related to the emission of optical polar phonons, which is seen in the GaN bulk simulations to be the dominant scattering mechanism for middle energies, up to $1.5\ eV$. However, it may be observed that the emission of acoustic phonons is greatly increased in the shorter device. This scattering mechanism is the dominant phonon scattering at energies above $1.5\ eV$, and in fact a peak can be seen corresponding with the position of the peak of the carrier energy shown in Fig. 6, while elsewhere it is negligible. The analysis of the scattering mechanisms is an important instrument for the evaluation of hot spots which affect electron transport inside the device. For example, the increase of the emission of acoustic phonons instead of the emission of optical polar phonons in reducing $L_{sg}$ and $L_{gd}$ is an important element for the heat management in the device. Optical phonon are in fact much slower than acoustic phonon, and thus do not contribute to heat transport as stay localized in the spot where the phonon are emitted. Eventually, optical phonons decay in acoustic phonon that are characterized by higher group velocity and can remove the heat from the device.

## III. LIMIT $f_T$ CALCULATION

In order to isolate the effect of the access regions, the gate length was always kept constant in Section II. However, it is well known that shorter gate devices show higher $f_T$. In Fig. 7, the $f_T$ of three devices with 600 $nm$, 400 $nm$ and 100 $nm$ $L_{sg}$,$L_{gd}$ are shown for different gate lengths while keeping to gate to channel distant constant. In particular, each color in the plot identifies one of the three length of the access regions where the gate length has been scaled. The solid curves represent $f_T$ as a function of the inverse of the metallurgical gate length, $L_g$. The dashed curves instead, show the relation between $f_T$ and the inverse of $L_{eff}$. In this analysis, the cut-off frequencies are calculated through AC simulation as explained in Section II, and the effective gate lengths are found in such a way that the $f_T$ derived by the transit time

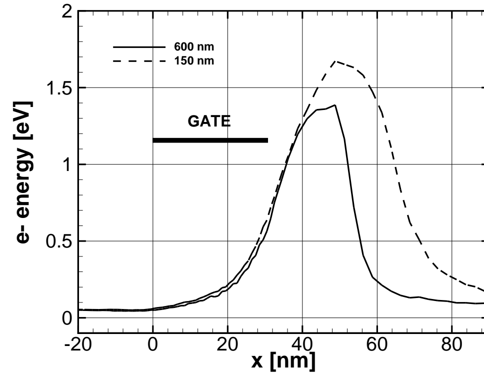


Fig. 6. Average energy of electron in the channel for 600nm(solid) and 150nm(dashed) Lsg,Lgd device. The gate starts at 0 and is indicated by a black rectangle. Bias: $V_D = 5\ V$, $V_G = -2\ V$.

of electrons through $L_{eff}$ matches the $f_T$ given by the AC simulations.

The main difference between the dashed curves and the solid ones is that while $f_T$ changes sub-linearly with $L_g$, it has a more linear dependence with $L_{eff}$ regardless of the value of the device access region. Rather, what the access region changes, is the slope of the linear dependence. In fact, the slope of the $f_T$ versus $L_{eff}$ plot represents a velocity being calculated as a frequency divided by the inverse of a distance. In particular, this slope represents the average transit velocity of electrons under the effective gate length, which as we saw in Section II, increases as $L_{sg}$ and $L_{gd}$ are reduced. Plotting the slope of $f_T$ versus $L_{eff}$ for the different lengths of $L_{sg}$ and $L_{gd}$ considered in our scaling, we can see in Fig. 8 that the data present a linear relation. This allows us to linearly extrapolate the transit velocity as $L_{sg}$ and $L_{gd}$ approach zero. An interesting aspect of this extrapolated limit velocity is that it is a property specific of the material, because it is independent from the device geometry. In fact, it was previously reported by Guerra *et al.* [6] that the slope of the $f_T$ versus $L_{eff}^{-1}$ line is not changed by scaling the gate-to-channel distance. In the inset of Fig. 8, we show the effective

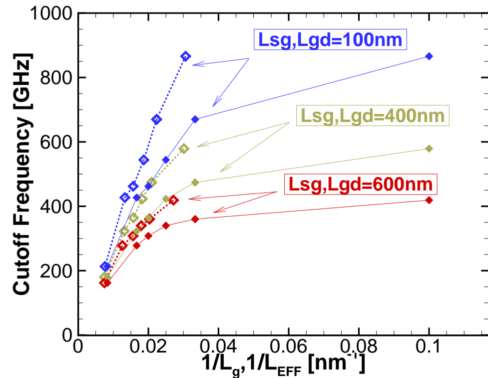


Fig. 7. $f_T$ versus the inverse of metallurgical (solid line solid symbols) and effective (dashed line empty symbols) gate length obtained at $f_T$ peak bias $V_D = 3.5\ V$, $V_G = -2\ V$.

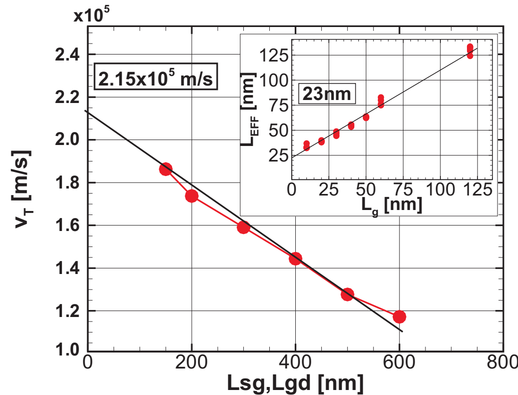


Fig. 8. Limit velocity extrapolation for 0 access region length. (Inset) Extrapolation of the limit effective gate length from the dependence with the metallurgic gate length. Bias: $V_D = 5\ V$, $V_G = -2\ V$.

gate lengths calculated at different metallurgic gate lengths. Similarly with the $f_T$ versus $L_{sg}$,$L_{gd}$ plot, the data present a linear relation and a limiting $L_{eff}$ can be extrapolated, this time for $L_g$ approaching zero. Dividing the limit effective gate length of 23 $nm$ by the limit transit velocity of $2.15 \times 10^5\ m/s$ we calculate a limiting transit time of $0.11\ ps$ corresponding to a maximum cut-off frequency of $1.49\ THz$. It has to be pointed out that the limiting cut-off frequency that we calculated is related specifically to this device, because the limit $L_{eff}$ is not a property of the material alone but it is influenced by the geometry as well. However, the device studied has a layout optimized for high RF performance with one of the thinnest barrier reported in literature that minimizes the fringing capacity and so $L_{eff}$.

## IV. Conclusions

In this work, we showed the effect of reduction of the access region length in GaN HEMTs on the carrier dynamics under the gate. In particular, we found that by reducing the source to gate and gate to drain spacing, the velocity of the electrons on source side of the gate controlled region is increased, which results in an overall increase of the device frequency response. Moreover, simultaneously downscaling the gate length and the access regions, we calculated the limiting transit velocity of the electrons under the gate as both the gate length and the access region length approaches zero, and we noticed that this velocity is a function only of the material in which the transport occurs. We also extrapolated a minimum effective gate length as the metallurgical gate length approaches zero, which allows, together with the limiting velocity, to calculate a limit transit time and the related maximum cut-off frequency of $1.49\ THz$.

## References

[1] Y.-F. Wu, M. Moore, A. Saxler, T. Wisleder, and P. Parikh, “40-W/mm double field-plated GaN HEMTs,” in *Device Research Conference, 2006 64th*, June 2006, pp. 151–152.

[2] Y. Yue, Z. Hu, J. Guo, B. Sensale-Rodriguez, G. Li, R. Wang, F. Faria, T. Fang, B. Song, X. Gao, S. Guo, T. Kosel, G. Snider, P. Fay, D. Jena, and H. Xing, “InAlN/AlN/GaN HEMTs with regrown ohmic contacts and $f_T$ of 370 GHz,” *Electron Device Letters, IEEE*, vol. 33, no. 7, pp. 988 –990, july 2012.

[3] K. Shinohara, A. Corrion, D. Regan, I. Milosavljevic, D. Brown, S. Burnham, P. Willadsen, C. Butler, A. Schmitz, D. Wheeler, A. Fung, and M. Micovic, “220 GHz $f_T$ and 400 GHz $f_{max}$ in 40-nm GaN DH-HEMTs with re-grown ohmic,” in *Electron Devices Meeting (IEDM), 2010 IEEE International*, Dec. 2010, pp. 30.1.1 –30.1.4.

[4] D. Brown, A. Williams, K. Shinohara, A. Kurdoghlian, I. Milosavljevic, P. Hashimoto, R. Grabar, S. Burnham, C. Butler, P. Willadsen, and M. Micovic, “W-band power performance of AlGaN/GaN DHFETs with regrown n+ gan ohmic contacts by MBE,” in *Electron Devices Meeting (IEDM), 2011 IEEE International*, dec. 2011, pp. 19.3.1 –19.3.4.

[5] U. Singisetti, M. H. Wong, S. Dasgupta, Nidhi, B. Swenson, B. Thibeault, J. Speck, and U. Mishra, “Enhancement-mode N-Polar GaN MISFETs with self-aligned Source/Drain regrowth,” *Electron Device Letters, IEEE*, vol. 32, no. 2, feb. 2011.

[6] D. Guerra, R. Akis, F. Marino, D. Ferry, S. Goodnick, and M. Saraniti, “Aspect ratio impact on rf and dc performance of state-of-the-art short-channel GaN and InGaAs hemts,” *IEEE Electron Device Letters*, vol. 31, no. 11, pp. 1217–1219, Nov. 2010.

[7] Y. Yamashita, I. Watanabe, A. Endoh, N. Hirose, T. Mimura, and T. Matsui, “Effect of source-drain spacing on dc and rf characteristics of 45 nm-gate AlGaN/GaN MIS-HEMTs,” *Electronics Letters*, vol. 47, no. 3, pp. 211 –212, 3 2011.

[8] K. Shinohara, D. Regan, I. Milosavljevic, A. Corrion, D. Brown, P. Willadsen, C. Butler, A. Schmitz, S. Kim, V. Lee, A. Ohoka, P. Asbeck, and M. Micovic, “Electron velocity enhancement in laterally scaled GaN DH-HEMTs with $f_T$ of 260 GHz,” *Electron Device Letters, IEEE*, vol. 32, no. 8, pp. 1074 –1076, Aug. 2011.

[9] R. Akis, J. Ayubi-Moak, N. Faralli, D. K. Ferry, S. M. Goodnick, and M. Saraniti, “The upper limit of the cutoff frequency in ultrashort gate-length InGaAs/InAlAs HEMTs: A new definition of effective gate length,” *IEEE Electron Device Letters*, vol. 29, no. 4, pp. 306–308, Apr. 2008.

[10] D. Guerra, M. Saraniti, D. Ferry, S. Goodnick, and F. Marino, “Carrier dynamics investigation on passivation dielectric constant and RF performance of Millimeter-Wave power GaN HEMTs,” *Electron Devices, IEEE Transactions on*, vol. 58, no. 11, pp. 3876 –3884, nov. 2011.

[11] B. E. Foutz, S. K. O’Leary, M. S. Shur, and L. F. Eastman, “Transient electron transport in wurtzite GaN, InN, and AlN,” *Journal of Applied Physics*, vol. 85, no. 11, pp. 7727–7734, 1999.

[12] D. S. Lee, J. Chung, H. Wang, X. Gao, S. Guo, P. Fay, and T. Palacios, “245-GHz InAlN/GaN HEMTs with oxygen plasma treatment,” *IEEE Electron Device Letters*, vol. 32, no. 6, pp. 755–757, Jun 2011.

[13] M. Saraniti and S. Goodnick, “Hybrid full-band Cellular Automaton/Monte Carlo approach for fast simulation of charge transport in semiconductors,” *IEEE Transactions on Electron Devices*, vol. 47, no. 10, pp. 1909–1915, October 2000.

[14] D. Guerra, F. Marino, D. Ferry, S. Goodnick, M. Saraniti, and R. Soligo, “Large-signal full-band monte carlo device simulation of millimeter-wave power GaN HEMTs with the inclusion of parasitic and reliability issues,” Dec. 2011, pp. 34.2.1 –34.2.4.