Cellular Monte Carlo study Lateral Scaling Impact of on the DC-RF Performance of High-Power 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 performance of millimeter-wave GaN HEMTs is investigated through nanoscale carrier dynamics description obtained by full band Cellular Monte Carlo simulation. The drain current and transconductance have shown to increase monotonically up to respectively $5500\ mA/mm$ and $1500\ mS/mm$ by symmetrically scaling the source to gate and gate to drain distance from $635\ nm$ to $50\ nm$. The electric field distribution has been studied for the shorter access regions and it was seen to be still far from the GaN breakdown limit. The access region scaling is found to greatly improve the frequency response of the device as well: from $340\ GHz$ up to $860\ GHz$. Detailed simulation of the carrier dynamics in the area under the gate showed that these improvements are due to higher transit velocity of electrons at the source end of the gate.**

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

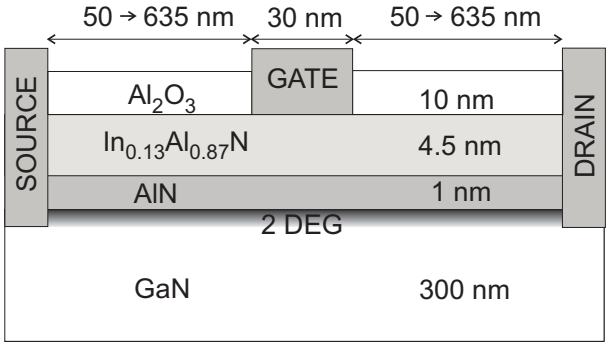

Fig. 1. Generic schematic cross-section of the GaN HEMT simulation domain.

I. Introduction

GAN HEMTs have recently emerged as leading candidates for high power and high frequency applications due to their combination of large breakdown voltages and high electron mobilities. Current cut-off frequencies of $300\ GHz$ [1] and f_{MAX} of $400\ GHz$ [2] have been achieved in nanoscale gate devices, while a $40\ W/mm$ output power was recorded in power GaN-HEMTs [3]. Several techniques have been adopted to improve the performance of HEMTs based on AlGaN/GaN heterostructure: self aligned gates, back barriers, low resistance ohmic contacts, vertical and gate length scaling. In particular, the reduction of the gate length has shown to provide the largest improvement in terms of cut off frequency, but, on the other hand, the reduction of the aspect ratio L_g/d (where L_g is the gate length and d is the channel to gate distance) increases short-channel effects. Vertical scaling preserves the aspect ratio, but the reduction of the barrier thickness increases the gate leakage current. Moreover, a reduced gate to channel distances induce depletion in the 2D electron gas. However, heterojunctions with new barrier materials like InAlN are characterized by a large polarization discontinuity and thus channels with large charge density. By using these materials it is possible to have thinner barrier preserving the aspect ratio, but it has been shown that for shorter gate length short channel effects cannot be neglected [4]. Degradation of the performance has been mitigated with the introduction of InGaN and AlGaN back barriers, which provide higher confinement of the carriers. However, back barriers in general reduce the thermal conductivity in comparison with bulk GaN [5]. The scaling of the source to gate and gate to drain length has recently drawn some attention as a technique to improve GaN HEMTs performance. For example, Shinohara and coworkers measured a 25% increase of the cut-off frequency by reducing the access region of a GaN HEMT [6]. In the present work, the effects of access region downscaling in a state-of-art InAlN/GaN HEMT [4] are investigated with RF and DC analysis through our full band Cellular Monte Carlo simulator [7]. Both the DC and the RF performance has shown to improve with such a scaling. Furthermore, the RF results have been explained in terms of nano-scale carrier dynamics providing guidelines for any future optimization efforts.

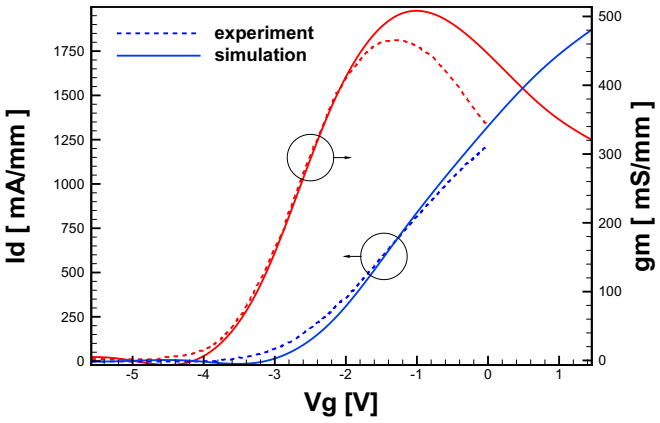

Fig. 2. Comparison between the experimental and simulated $Id-Vg$ for $Vd = 5\ V$.

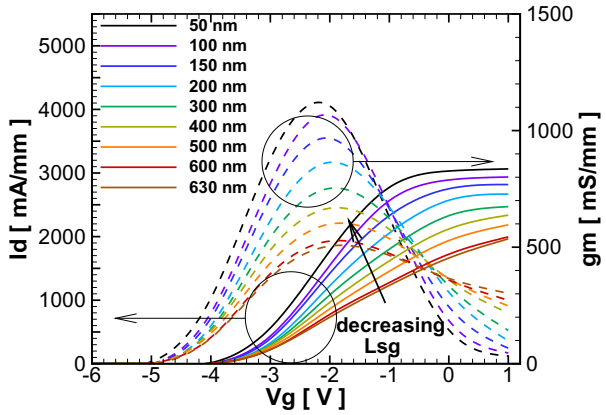

Fig. 3. Comparison of $Id-Vg$ and g_m for different source-gate access regions lengths corresponding to different colors in the legend and decreasing according to the arrow. The devices are biased in saturation at $V_D = 5\ V$.

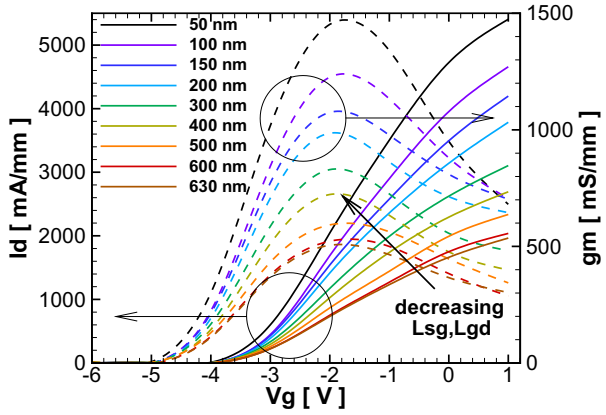

Fig. 4. Comparison of $Id-Vg$ and g_m for different source-gate, gate-drain access region lengths corresponding to different colors in the legend and decreasing according to the arrow. The devices are biased in saturation at $V_D = 5\ V$.

II. DC ANALYSIS

The GaN HEMT analyzed in this work is shown in Fig. 1 and was proposed by Lee *et al.* [4]. The source to drain

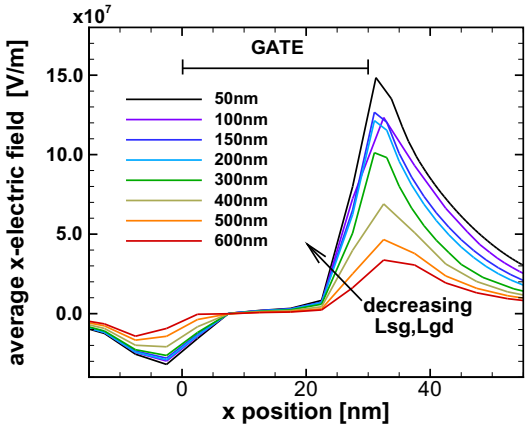

Fig. 5. Profile of the electric field component along the transport direction for a slice extracted under the gate for different access region lengths. The gate source-end is set as origin of the axis. The devices biased at $V_G = -2\ V$ and $V_D = 5\ V$.

distance is $1.3\ \mu m$ and the gate length is $30\ nm$. In Fig. 2, the match of the simulated $Id-Vg$ with the experimental measurement is shown for a drain bias of $5\ V$. It may be noticed that the simulated curve starts diverging from the experimental curve for high values of the current. This difference is due to self-heating which is neglected in the simulation. In fact, for values of current above $800\ mA/mm$, the self-heating reduces the current and the calculated current is consequently overestimated as it was previously reported in [8]. Fig. 3 shows the $Id-Vg$ when the distance between the source and the gate (L_{sg}) is reduced to $50\ nm$ from $635\ nm$ of the original device, while the drain access region length (L_{gd}) and the gate length maintains the original dimension of $30\ nm$ and $635\ nm$, respectively. We can see from Fig. 3 that the current and g_m increase by as much as 2 due to the scaling, while the threshold voltage remains constant. A different asymmetric scaling is then performed on the gate to drain access region this time preserving the gate length and Lsg. In this case, the $Id-Vg$ characteristic does not show any relevant variation and the plot is omitted. Fig. 4 shows the $Id-Vg$ for a symmetrical scaling of the access region, where L_{sg} and L_{gd} were scaled at the same time, keeping the gate length constant. It may be noticed that the current and g_m values achieved with this scaling are almost twice as much as scaling L_{sg} alone. The effect of the different scaling on the $Id-Vd$ characteristic can be understood in terms reduced source access region resistance R_{sg} and drain access region resistance R_{gd}. In fact, the access region can be modeled as a series resistor between the contact and the gate area, which is the active part of the device. When a current is flowing through the device, part of the bias applied to the contact drops along the access region. The voltage seen by the active part of the device is called intrinsic to be distinguished from the extrinsic bias applied at the contacts, which appears in the $Id-Vg$ and $Id-Vd$ plots. In particular, when L_{sg} is reduced, the intrinsic source-gate voltage V_{gs} increases. Therefore the output current of the device with reduced L_{sg} will be larger than in the unscaled device, as seen in Fig. 3. However, curves with higher

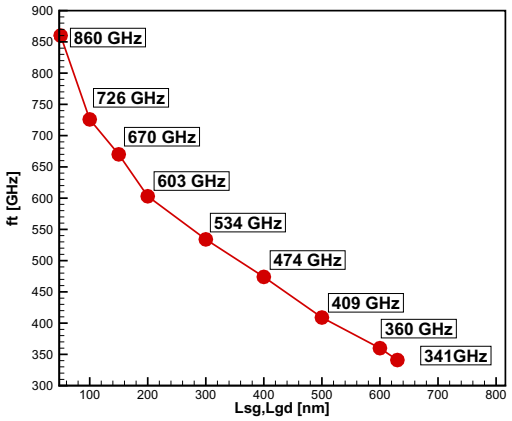

Fig. 6. f_T calculated for different symmetrically scaled access regions. f_T have been extracted from the short circuit current gain. The devices are biased for maximum f_T at $V_G = -2\ V$ and $V_D = 3.5\ V$.

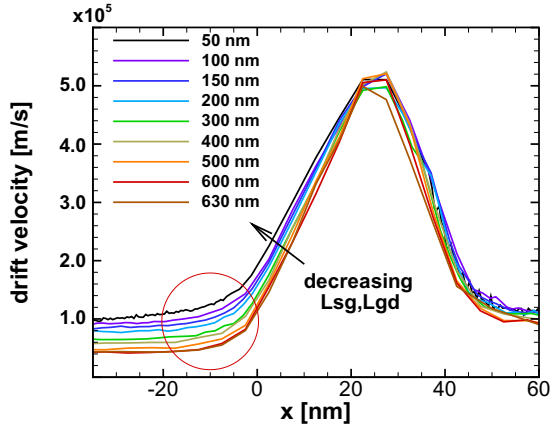

Fig. 7. Average electron velocity profile for different access region lengths decreasing according to the arrow in the plot. The gate source-end is set as origin of the axis. Metallurgic gate from 0 nm to 30 nm.

V_{gs} have higher saturation voltage, so another effect of scaling L_{sg} is to draw the device outside the saturation regime. On the other hand, the reduction of the drain resistance due to downscaling L_{gd} decreases the saturation voltage so that the saturation condition is achieved for a lower drain voltage, and the scaled device returns in saturation regime. Adversely, a decrease of the saturation voltage in devices that are operating in saturation will not have any impact in the $Id - Vg$. Therefore, it can be concluded that most of the improvements in DC performance are due to the Lsg scaling, but the simultaneous scaling of L_{gd} is required to achieve the maximum current and g_m. An important aspect that has to be considered when scaling down the horizontal dimensions of a device is breakdown. To this purpose Fig. 5 shows the electric field along a slice just underneath the gate, where the field reaches the highest values. In particular, it can be seen that even for the 50 nm L_{sg},L_{gd} the peak field is less than 3 times the GaN breakdown field.

III. AC ANALYSIS

The AC analysis has been performed through a multisinusoidal simulation. In this technique, the device is perturbed

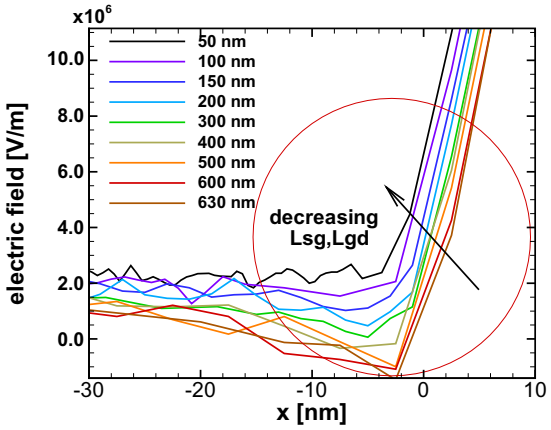

Fig. 8. Profile of the electric field component along the transport direction for different access region lengths decreasing according to the arrow in the plot. The gate source-end is set as origin of the axis. The red circle highlights the source end of L_{eff} which corresponds to the same red circle in Fig. 7.

around the bias point corresponding to the peak g_m by a combination of sinusoids with different frequencies appropriately selected. The contribution of the single sinusoids to the output current is then separated by Fourier decomposition, and the different gains are calculated under the assumption that the device response is linear for small perturbation. The multi-sinusoids approach is a good compromise between the single sinusoid and the step perturbation, being computationally more efficient then the first and offering better SNR than the latter. In order to establish the influence of the access region scaling on the AC performance, the cut-off frequency (f_T) has been extracted from the short-circuit current-gain. In Fig. 6, we can see that f_T monotonically increases almost 3 times by symmetrically downscaling L_{sg} and L_{gd}. In this scaling, the gate length and the gate to channel distance are kept constant. A physical interpretation of f_T is related to the electron transit time under the gate controlled area. In Fig. 7, the profile of the average electrons velocity under the gate is shown. In particular, comparing the velocity profiles for devices with different L_{sg} and L_{gd}, we can see that the curves near the peak are very similar for the different cases. On the other hand, the average velocity on the source side of the gate, is monotonically increasing for shorter access regions. Moreover, we can see that the electrons in the source access region are more than 2 times faster in the device with 50 nm L_{sg},L_{gd} than in the original one. This is due to the higher electric field originating from the reduced distance over which the source-gate potential drop takes place. Fig. 8 shows the field value calculated in proximity of the gate end of the source access region for devices with different L_{sg} and L_{gd}. However, we saw in Fig. 6 that the frequency increases substantially with the scaling, but in the velocity profile one notes that only the short initial region under the gate is affected by the scaling. A similar relation between f_T and the velocity profile has been observed also by Guerra *et al.* [9]. In order to understand this relation we recall the definition of f_T with respect to the transit time:

$$\tau_T = \frac{1}{\omega_T} = \frac{1}{2\pi f_T} = \int_{L_{eff}} \frac{1}{v_{ave}(x)} dx, \qquad (1)$$

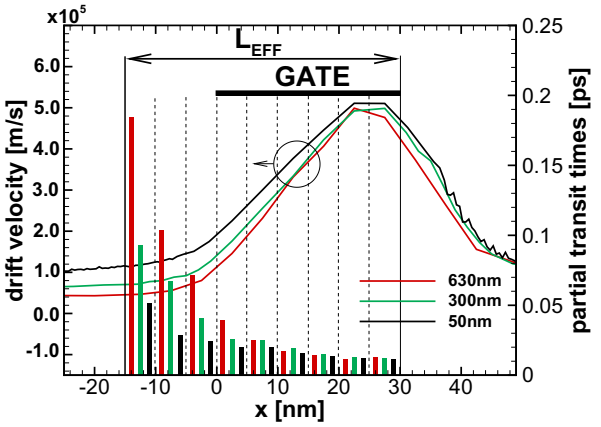

Fig. 9. Average electron velocity profile for three different L_{sg} L_{gd}. The histograms represents the partial contribution to the total transit time of 5 nm sections of L_{eff}, for different symmetrical access region lengths. The gate source-end is set as origin of the axis. Metallurgic gate from 0 nm to 30 nm.

where L_{eff} is the effective gate length, and $v_{ave}(x)$ is the velocity calculated along the channel (x-direction) and is obtained as a weighted average along the vertical y-direction with respect to the carrier density [10]. Particular attention has to be paid to the range over which the integration is performed. Indeed, the electron velocity is integrated along the effective gate length and not the metallurgic gate length, since the gate modulation of the electrons in the channel is extended beyond its geometrical projection, due to the presence of fringing fields. According to Eq. (1) the total transit time is a summation of the electron transit times through infinitesimal lengths in which the velocity can be considered constant. If we decompose the total transit time into a summation of partial transit times along the effective gate length, we notice that the largest terms are due to the slow electrons located near the source side as reported by Guerra *et al.* [11]. In Fig. 9, a comparison of the partial transit time associated with 5 nm section along L_{eff} is shown for devices with three different L_{sg},L_{gd}. In particular, it may be seen that for shorter access regions the partial transit times related to the initial portion of L_{eff} are greatly reduced. For example, the transit time along the first 5 nm section in the 50 nm device is only 30% of the one in the 635 nm device. In other words, electrons spend less time in the lower velocity area when the length of the access region is reduced, improving the speed of the device.

IV. Conclusions

In this work, the impact of access region scaling in a InAlN/GaN HEMT has been investigated in terms of drain current, g_m and f_T. The DC performance is influenced by the reduction of access region resistance and this yield as much as a three-fold increase in current and g_m values for a symmetrical scaling down to 9% the original source to gate and gate to drain length. From the velocity profile, it was shown that the nano-scale carrier dynamics is affected by the scaling. In particular, electrons in the scaled devices have a higher velocity in the area near the source side of the gate, which significatively improves their RF performances. Therefore, the scaling of the access regions is a promising technique to improve the DC and RF performance of GaN HEMTs.

References

[1] D. S. Lee, X. Gao, S. Guo, D. Kopp, P. Fay, and T. Palacios, "300-GHz InAlN/GaN HEMTs with InGaN back barrier," *Electron Device Letters, IEEE*, vol. 32, no. 11, pp. 1525 –1527, Nov. 2011.

[2] 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.

[3] 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.

[4] 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.

[5] S. G. D. Lee, X. Gao and T. Palacios, "InAlN/GaN HEMTs with AlGaN back barriers," *IEEE EDL*, vol. 32, no. 5, pp. 617–619, May 2011.

[6] 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.

[7] 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.

[8] 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.

[9] D. Guerra, M. Saraniti, N. Faralli, D. Ferry, S. Goodnick, and F. Marino, "Comparison of N-face and Ga-face GaN HEMTs through cellular monte carlo simulations," *IEEE Transactions on Electron Devices*, vol. 57, no. 12, pp. 3348–3354, December 2010.

[10] 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.

[11] 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.