Received October 29, 2019; accepted December 7, 2019. Date of publication 10 December 2019; date of current version 3 January 2020.
The review of this article was arranged by Editor C. C. McAndrew.

*Digital Object Identifier 10.1109/JEDS.2019.2958915*

# Computational Modelling-Based Device Design for Improved mmWave Performance and Linearity of GaN HEMTs

**ANKIT SONI AND MAYANK SHRIVASTAVA (Senior Member, IEEE)**

Advanced Nanoelectronics Device and Circuit Research Laboratory, Department of Electronic Systems Engineering, Indian Institute of Science, Bengaluru 560012, India

CORRESPONDING AUTHOR: A. SONI (e-mail: soni@iisc.ac.in)

This work was supported by the Department of Science and Technology, Government of India, under Grant DST/TSG/AMT/2015/294.

**ABSTRACT** In this work, a comprehensive, TCAD based design approach for mmWave (mmW) GaN HEMTs is presented. Unique trade-offs between epi-layer design and HEMT's mmW performance are discussed. Effect of surface states on cut off frequency is modeled and presented. We have found that carrier trapping by the donor type interface states causes RF performance drift at high drain fields, which particularly leads to the non-linear behavior of mmW HEMTs at high drain bias. Moreover, we have observed that channel electrostatics, barrier layer, and UID GaN channel design govern the linearity and scaling behavior of such GaN HEMTs. To improve channel electrostatics, which improves the linearity and cut-off frequency, a partially recessed barrier under the gate is studied. A relative study of AlN/GaN HEMT and AlGaN/GaN HEMTs is performed to investigate the nonlinearity behavior. In addition, the dependence of cut-off frequency on contact resistance and lateral scaling is studied for partially-recessed barrier and conventional design for both AlN and AlGaN barrier types. The mmW performance is found to be a strong function of barrier design in the gate and recess regions. Unique design trends and physical behavior was observed for AlN and AlGaN barriers, which signifies that design guidelines derived for one epi-stack can't be deployed to the other.

**INDEX TERMS** AlGaN/GaN, AlN/GaN, TCAD, HEMT, device simulation, RF HEMT, device design, non-linearity.

## I. INTRODUCTION

While GaN HEMT has emerged as an alternative to InP and GaAs technologies for mmW applications, there are limited reports on the design of mmW HEMTs. Design guidelines for mmW HEMTs to maximize performance and improve linearity behavior is hardly existent in the literature. Attributed to a number of design/technology/growth parameters to engineer, the design of such RF HEMT has become a multi-dimensional engineering problem, which is non-trivial to address from the experimental design of experiments. The evolution of Silicon RF technology has demonstrated the advantage of TCAD based co-design approach for engineering complex device architectures [1]. A TCAD based approach to design and engineer RF power devices, depending on the application-specific requirements, significantly saves device design time, as well as offers improved system performance. This eventually brings down the cost and time to market while potentially maximizing the performance beyond set limits.

While there have been limited efforts to develop TCAD based design approach [2]–[9], the same however is largely non-existent for mmW HEMTs. Optimization studies related to field plated devices by Karmalkar and Mishra [2], were only limited to the design of field plate for maximizing breakdown voltage. In another work, Medjdoub *et al.* [3] correlated 2DEG density ($n_s$) with AlInN/GaN HEMT barrier layer scaling. Saito *et al.* [4] carried out TCAD based studies to analyze the effect of contact on the breakdown voltage of AlGaN/GaN HEMT. Moreover, TCAD computations have been used to carry out $R_{ON} - V_{BD}$ optimization [5], [6] &

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see http://creativecommons.org/licenses/by/4.0/

mitigate drain−to−source punch through [10]. Joshi *et al.* presented a computational modeling approach for HEMTs in general [7] and physics of Carbon doping [8], [9]. While these works independently address one or the other parameters to enhance DC or breakdown performance, there is no report on the approach for maximizing RF figures of merit parameters of HEMTs for mmW applications, while accounting for design − performance − nonlinearity trade-offs. At the RF front there have been only a few reports on TCAD modeling of GaN HEMTs, limited to large signal analyses [11], [12]. There is however no report on detailed small-signal modeling of RF HEMT devices for mmW applications. In this paper, we have also investigated the RF performance of a partially recessed architecture by carrying out thorough comparative analyses of design parameters such as barrier type, lateral scaling, and contact resistance for both AlN and AlGaN barriers. The previous reports on partially recessed HEMT for RF applications experimentally demonstrate the improvement in transconductance [13], [14]. However, it's correlation with other design aspects is missing. The goal of this work is to bridge the existing gap and develop deeper understanding of RF HEMT device design to push the performance for mmW applications.

This work attempts to address this gap. The manuscript has been arranged as follows: Section II presents a design strategy for RF HEMT design. Section III summarizes the computational framework used in this work, whereas Section IV uses the developed approach and design insights to understand the impact of buffer and channel layers on the mmW performance. Section V discusses the design parameters- barrier layer, surface states and drain induced nonlinearity. A partially recessed HEMT has been modelled and studied in Section VI. Finally, Section VII concludes this work.

## II. TCAD-DESIGN STRATEGY FOR RF HEMTS

The device cut off frequency ($f_T$) is the measure of maximum frequency at which device current can be effectively modulated by the gate and is given as [16]: $\frac{1}{2\pi f_T} = \frac{C_{gs}+C_{gd}}{g_m} + C_{gd}.(R_s + R_d).[1 + (1 + \frac{C_{gs}}{C_{gd}})\frac{g_d}{g_m}]$, where $C_{gs}$ and $C_{gd}$ are gate to source and gate to drain capacitance, respectively, $g_m$ is transconductance, $g_d$ is output conductance, and $R_s$ and $R_d$ are source and drain resistances, respectively. The expression represents total delay corresponding to carrier transport in the device. It can be classified into 3 components- intrinsic delay, parasitic delay, and extrinsic delay. It is critical to understand the contribution of each delay component for improving the frequency performance of the device. While the intrinsic and extrinsic delays are governed by device transconductance ($g_m$) and capacitances ($C_{gs}$ and $C_{gd}$), the extrinsic delay is a function of device parasitic resistance ($R_s, R_d$) and output conductance ($g_d$). The RF design mandates to address challenges such as- how does the contribution of delay components change with physical parameters and bias conditions? What are the trade-offs associated with it? The device nonlinearity performance is another key design aspect, for the device to function at large signal operating conditions. In addition to electrostatics in the channel, nonlinearity in short channel devices is also affected by intrinsic resistivity of the buffer region, which is defined by unintentional doping during epi-growth. Surface states in HEMT play a significant role in determining DC as well as the RF performance of the device. The device co-design design approach can greatly help in predicting the physical behavior by modeling the impact of surface traps on the cut-off frequency of the device. We have also observed that the contact resistance dependency on cut-off frequency is co-related to HEMT design. Earlier works reporting record high cut-off frequencies [15], [17], [18] have often compromised in terms of breakdown voltage, leading to breakdown voltage and cut-off frequency product always below theoretical limits. A well-framed co-design approach is required to thoroughly engineer design while exploring various trade-off and inter-dependencies of parameters. In this work, the product of breakdown voltage and cut-off frequency ($V_{BD}.f_T$) is taken as a figure of merit while presenting the said co-design approach.

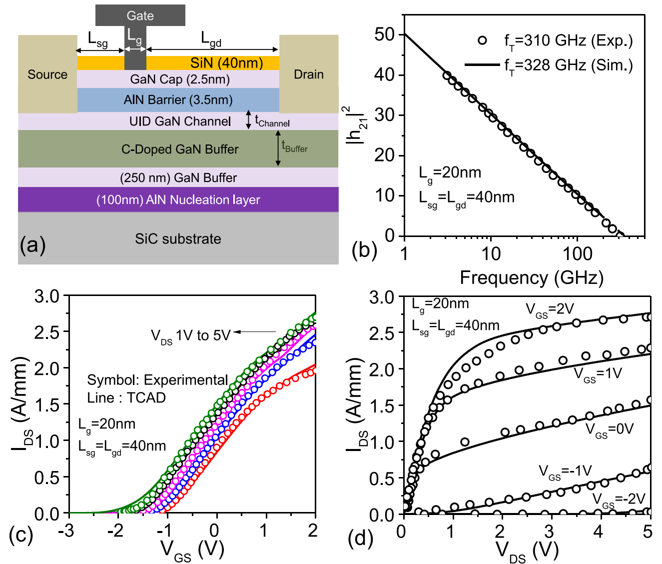


**FIGURE 1. (a) Cross-sectional view of the studied normally-ON HEMT device. (b) Experimentally extracted current-gain versus frequency characteristics of the fabricated HEMT device (Symbols) plotted together with simulated I-V characteristics (Line). Calibrated (c) transfer and (d) output characteristics of the device. The device cross-section and the experimental data has been taken from literature [15].**

## III. COMPUTATIONAL FRAMEWORK

For TCAD simulations, a well-calibrated TCAD setup has been used. Details of the computational framework used are presented in earlier works [7]–[9]. The device structure used for the investigation is shown in Fig. 1(a), adopted from the report by Shinohara *et al.* [15]. The simulation setups for the small signal as well as DC operation are calibrated with the experimental results [15] as shown in

Fig. 1(b)-(d). To summarize, polarization at all the hetero-interfaces is considered for accurately estimating the energy band profile, 2-dimensional electron gas (2DEG) density and for accounting surface charge. The impact of surface and buffer traps is also captured in this study to get insights and precise estimation of breakdown voltage and frequency performance. Contacts are physically modeled as Schottky interfaces, as explained in detail in [7]. Carrier transport [7] accounts for both carrier as well as lattice heating by enabling hydrodynamic and thermodynamic transport models. Finally, gate leakage due to Fowler-Nordheim tunneling and Poole-Frenkel tunneling is also considered. C-doping induced buffer traps and resulting avalanche behavior has been accounted to predict breakdown voltage [8], [9]. Breakdown simulations are performed using an impact ionization model according to Chynoweth law [19] with the critical electric field for GaN as 3 MV/cm. The breakdown voltage is calculated by considering the drain current limit of 1mA/mm while applying off-state stress at the gate. To study the impact of C-doped buffer, a compensating doping profile is considered in the buffer layer as modeled in our earlier works [8], [9]. The acceptor and donor type trap concentrations are taken as $10^{18} cm^{-3}$ and $5 \times 10^{17} cm^{-3}$ respectively. The carrier scattering due to the presence of C-dopant atoms is accounted for in the simulations according to Masetti Model [20]. The source to gate distance $(L_{sg}) = 40$nm, gate to drain distance $(L_{gd}) = 40$nm and gate length $(Lg) = 20$nm are used unless specified otherwise.

## IV. DESIGN OF BUFFER AND CHANNEL LAYERS

### A. GAN BUFFER DESIGN

GaN Buffer is a key parameter for high voltage device applications. However, the buffer has not been considered a critical design aspect for RF HEMT devices. Carbon-doped buffer is employed for the majority of the high voltage HEMT stacks. It consists of a Carbon doped GaN layer sandwiched between channel and UID buffer regions. It is well known, C doping mitigates the buffer leakage current at higher drain bias voltage and improves the breakdown voltage of the device. For RF applications, it is imperative to optimize the buffer and channel thickness as C doping can influence the 2DEG concentration and channel mobility, affecting the cut-off frequency of the device. Here two critical design parameters channel thickness $(t_{Channel})$ and C-doped buffer thickness $(t_{Buffer})$ are studied as follows.

#### A.1. ROLE OF C-DOPED BUFFER

In order to study the impact of C-doped buffer thickness, the channel thickness is kept constant at 150nm and the device RF characteristics are studied as a function of C-doped buffer thickness $(t_{Buffer})$. As observed in Fig. 2(a), buffer thickness does not affect the RF performance significantly as the lateral transport properties of the device are not perturbed if C-dopants are deep into the buffer region, away from the active channel. The parasitic capacitances are independent of the buffer thickness as deeply situated C-doped buffer does not participate in channel charging or discharging.

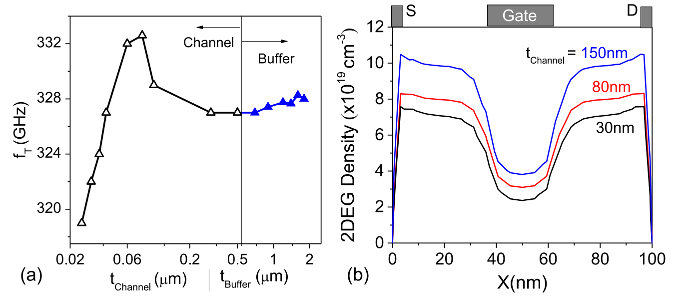


**FIGURE 2. (a) Effect of channel and buffer thickness on cut-off frequency. (b) Distribution of 2DEG in channel as function of channel thickness. Carriers in the shallow channel are trapped by Carbon dopants and do not contribute to the channel conduction.**

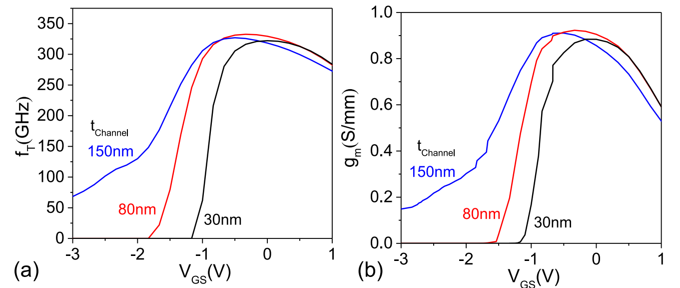


**FIGURE 3. (a) Non-linearity in cut-off frequency and (b) transconductance ($g_m$) as function of channel thickness.**

#### A.2. ROLE OF CHANNEL THICKNESS

The channel thickness $(t_{Channel})$ determines the distance of the C-doped buffer layer from the 2DEG channel. The C dopants introduce acceptor type traps that limit the spreading of hot carriers into the buffer region at high drain bias. It leads to a reduction in output conductance, but at the same time due to carrier trapping, the effective 2DEG concentration also falls as depicted in Fig. 2(b). Besides, increased carrier scattering due to Carbon dopants further degrades the on-state performance of the HEMT. All these factors are imperative to RF device performance. Improvement in $f_T$ is observed initially as the channel thickness is increased (Fig. 2(a)). It can be attributed to reduced carrier scattering and channel depletion, which in turn improves the on-state characteristics of the device. Hence higher 2DEG concentration is the key factor for improved cut-off frequency in HEMT. It should be noted that further increasing the channel thickness beyond 60nm, the output conductance starts to dominate and ceases any improvement in $f_T$.

Another desirable critical aspect of an RF device is having a fixed cut-off frequency over a large gate swing. It is observed that the nonlinearity with gate voltage increases significantly with decreasing channel thickness as shown in Fig. 3(a). The lower 2DEG density as a result of carrier trapping degrades the device transconductance as depicted in Fig. 3(b). The fall in $g_m$ is responsible for a steeper $f_T$

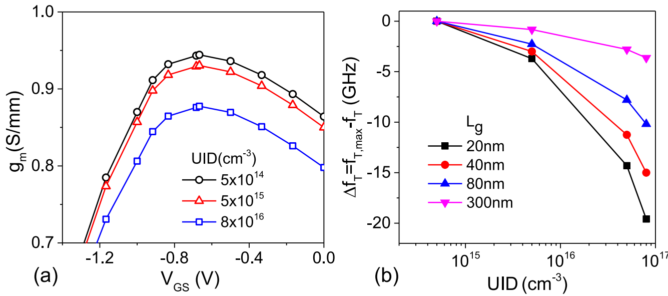


**FIGURE 4.** (a) Effect of unintentional doping (UID) on transconductance ($g_m$) of the device. (b) The dependency of $f_T$ roll-off with gate length as a function of UID. The degradation in $f_T$ is increases with gate length scaling at higher UID due to poor gate control.

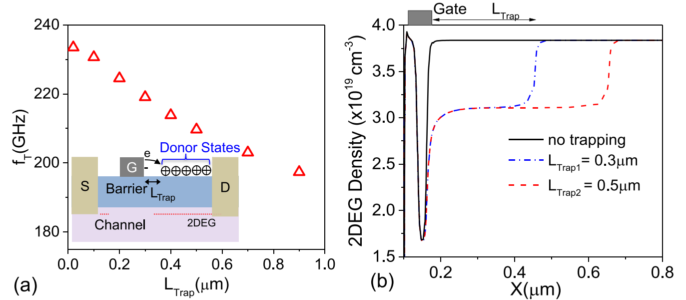


**FIGURE 5.** (a) Cut-off frequency degradation with $L_{Trap}$, inset represents the virtual gate formation used for device simulation. $L_{Trap}$ is the length up to which electrons from the gate are injected in the donor states and deplete the region underneath. (b) Simulated electron density in the channel as a function of trapping length $L_{Trap}$. As the carrier trapping near the gate increases, the channel depletion further extends towards the drain side.

roll-off as a function of gate bias below a critical channel width. Hence it is imperative to optimize the channel layer to achieve high breakdown voltage in conjunction with good linearity characteristics.

### B. HOW CRITICAL IS BACKGROUND DOPING OF CHANNEL?

High unintentional doping (UID) in GaN buffer results in parasitic conducting paths and are responsible for drain to source leakage current. It lowers the breakdown voltage by providing excess carriers for early impact ionization. Leaky channel causes the hot electrons to spread out of the channel, deteriorating the carrier confinement. The increased output conductance decreases the total drain current available for charging the channel capacitance, which introduces an additional delay component. It is also evident by the fall in transconductance at higher doping concentrations as shown in Fig. 4(a). This eventually leads to reduction in $f_T$ as depicted in Fig. 4(b). The effect of UID is more pronounced for short channel devices. The increased parasitic coupling between the gate and drain results in further degradation of $f_T$.

## V. BARRIER LAYER DESIGN, SURFACE STATES AND DRAIN FIELD INDUCED NON-LINEARITY

### A. SURFACE STATES AND NON-LINEARITY

2DEG in HEMT is very sensitive to surface states. To study this effect, the virtual gate concept [21] was applied. The virtual gate formation is attributed to the trapping of electrons in the surface states leading to a decrease in net positive charge on the surface. This charge trapping phenomenon is particularly dominant at large signal device operation. The RF devices operate in the saturation region, resulting in depletion near the drain side of gate edge to support the high drain bias. The resultant field at the gate generates hot carriers, that may ionize the surface traps and deplete the 2DEG in the localized region. To emulate this behavior, the surface states up to the length $L_{Trap}$ from the gate edge, are deionized as illustrated in Fig. 5(a) inset. Surface traps were allowed to ionize starting from the drain side of the gate edge, as a function of the drain field, which gradually extended towards the drain as the lateral field increased. At higher drain bias, the high energy channel electrons or hot electrons get trapped by the donor states present at the surface of the barrier. This is enhanced by the high electric field at the gate edge and aids in virtual gate formation. Consequently, the 2DEG charge in the channel region depletes in proportion to trapping at the surface as shown in Fig. 5(b). The rate of carrier trapping or virtual gate extension is determined by the surface trap concentration, characteristics, and electric field in the channel. Higher trapping rate at the surface leads to extended depletion of the 2DEG in channel, as a consequence the gate to drain region resistance increases. It translates to added drain delay, resulting in a slower device response at high frequencies. Cut off frequency behavior studied as a function of distance $L_{Trap}$, shows substantial degradation with virtual gate length, as depicted in Fig. 5(a), which is expected to significantly add to nonlinearity as a function of the drain field.

The characterization technique for the measurement of depletion width extension because of surface trap ionization is reported in the literature [22]. The series of floating gates have been used to capture the trap ionization length as function of gate bias. However, the reported technique is valid for off-state device characterization, where the surface trap ionization only due to gate injected carrier is accounted. In RF operation, the device operates in saturation region, resulting in depletion near drain side of gate edge. One method to capture this by calculating the depletion region length as function of drain bias by employing drain delay extraction model as proposed in [23]. This technique, however, does not isolate the impact of surface traps and drain bias in extension of depletion region. We propose to employ this method in a passivated device or device under UV exposure. It can help to suppress the impact of surface traps, and the true drain delay contributed by applied drain bias can be measured. By subtracting it from the total delay time of device measured without passivation or without UV exposure, can predict the contribution of surface trap in total delay.

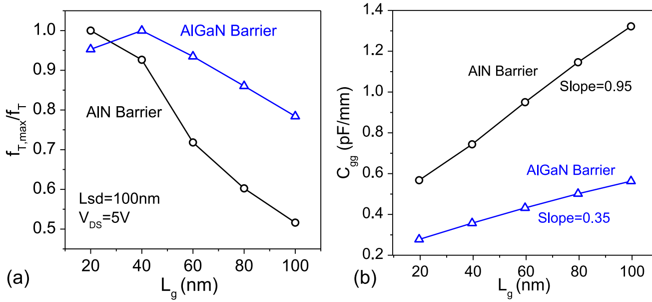


**FIGURE 6.** **(a) Drop in relative cut-off frequency and (b) total gate capacitance as function of gate length for AlN/GaN HEMT and AlGaN/GaN HEMT. The parasitic capacitance increases at much higher rate for AlN barrier design.**

Further experiments need to be carried out to confirm the hypothesis.

### B. BARRIER LAYER DESIGN AND DRAIN FIELD INDUCED $F_T$ DRIFT

The barrier layer is a vital design aspect, which controls the key device performance parameters such as transconductance, gate leakage, 2DEG sheet density and device linearity at high-frequency operation. AlGaN and AlN are the barrier layer materials that have been adopted for RF HEMT devices. In this section, we have analyzed the RF performance of the devices with aforementioned barrier types. The comparative study of the impact of gate length scaling in both types of devices is depicted in Fig. 6(a). The AlN barrier offers better cut-off frequency compared to the AlGaN barrier device owing to higher polarization induced sheet density. In addition, the barrier thickness of 3.5nm used in this study for AlN barrier is sufficient to induce sheet charge density in excess of $1 \times 10^{13} cm^{-2}$ unlike the AlGaN barrier, in which case the barrier thickness of 25nm is employed for similar 2DEG concentration. The reduced gate to channel distance, significantly amplifies the transconductance of the device, leading to high cut-off frequencies. However, the cut-off frequency rolls-off as a function of gate length, and the effect is much severe for the AlN barrier as shown in Fig. 6(a). It can be explained by examining the total gate capacitance ($C_{gg} = C_{gs} + C_{gd} + C_{channel}$) for both the devices. The $C_{channel}$ is higher for the AlN barrier due to a relatively thin barrier. However, the remaining component of $C_{gg}$ consists of the parasitic capacitances ($C_{gs}$+$C_{gd}$), which results in lowering of the cut-off frequency. Fig. 6(b) clearly predicts the increase in parasitic capacitance with gate length for both AlGaN and AlN barrier devices. The parasitic capacitances, however, increase at a much higher rate in the AlN barrier device with a slope of 0.95 compared to 0.35 for that of the AlGaN barrier. Similar to the channel capacitance, the parasitic capacitance is also a function of barrier thickness, dielectric constant and 2DEG density. Hence, the parasitic capacitance rises at a much faster rate in the AlN barrier, resulting in poor frequency response for long channel devices.

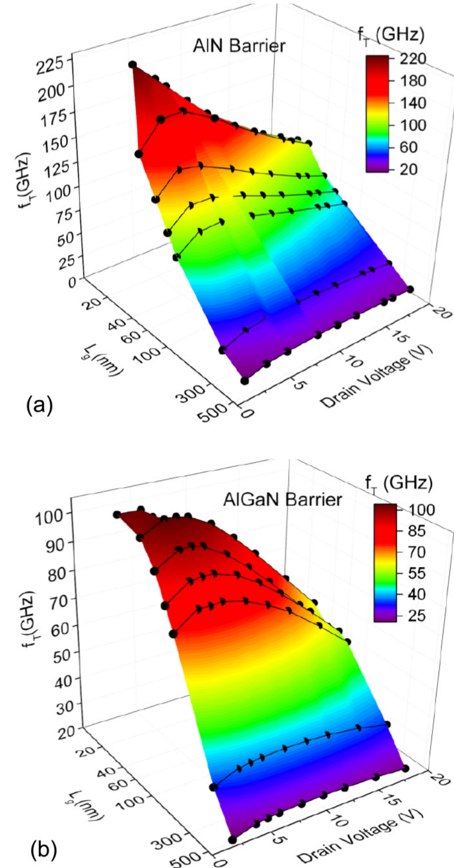


**FIGURE 7.** **HEMT cut-off frequency as a function of channel length and drain voltage depicting nonlinearity in short channel devices in case of (a) AlN/GaN HEMT and (b) AlGaN/GaN HEMT.**

RF performance drift with drain voltage for short channel length device depicts the increased nonlinearity. It is observed for AlN as well as AlGaN barrier HEMT devices as shown in Fig. 7(a) and (b) respectively. This can be explained as follows. As the drain field is increased, carriers drift with higher velocity, which improves the transconductance of the device. It translates to higher cut-off frequencies until velocity saturation is reached. With further increase in drain voltage, the depletion region below the gate extends towards the drain, to support additional electric field. This in-turn gives rise to additional carrier transit time that is characterized by drain delay. For short-channel devices drain delay was observed to be significantly increased at higher drain voltages, which lowers the cut-off frequency and adds to nonlinearity.

As discussed above, nonlinearity is a direct consequence of short-channel effects in HEMT, which arises due to poor gate control. The barrier thickness plays a decisive role in gate charge modulation. It is imperative to maintain good electrostatics in the gate region to avoid short channel effects for smaller gate lengths. Fig. 8(a) and (b) depict the cut-off frequency dependence for AlN and AlGaN barriers at varying gate lengths. It is noted that devices show significantly less

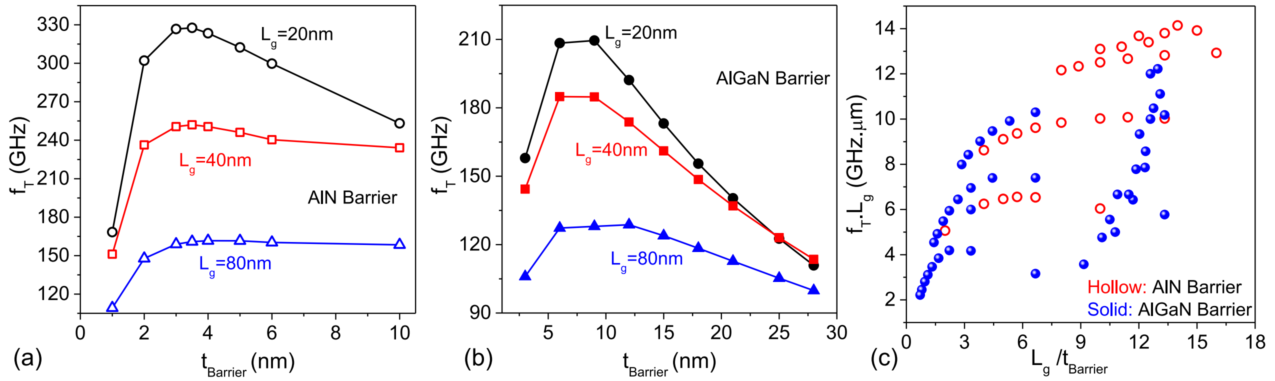


**FIGURE 8. Extracted cut-off frequency by simultaneous scaling of both gate length and barrier thickness in case of (a) AlN barrier and (b) AlGaN barrier. (c) Optimization of vertical and lateral electric field in the channel to maximize $f_T$. For $\frac{Lg}{t_{Barrier}}$ ratio > 5, $f_T.L_g$ product was found to be higher in case of AlN barrier, depicting larger optimum scaling window compared to AlGaN barrier.**

degradation in $f_T$ as a function of barrier thickness with the AlN barrier compared to the AlGaN barrier device. Below a critical barrier thickness, $f_T$ roll-off is attributed to incomplete surface trap ionization that results in low 2DEG in the channel. However, the critical thickness for the AlGaN barrier is much higher compared to the AlN barrier as shown in Fig. 8(a) and (b). It enables much larger aspect ratio ($\frac{L_g}{t_{Barrier}}$) in AlN/GaN HEMT devices. The impact of barrier layer on the device linearity can be described as follows.

For the short gate length below 100 nm, the lateral electric field becomes dominant and the vertical electric field component starts diminishing. This results in poor gate controllability and lower gate modulation efficiency, which collectively degrade the cut-off frequency performance as depicted in Fig. 8(a) and (b) for AlN barrier and AlGaN barrier respectively. Hence to mitigate the short channel effects and nonlinearity behavior, it is essential to enhance the vertical electric field, which can be achieved by top barrier layer scaling. As discussed above due to the properties of the AlN barrier, the barrier width can be scaled down to 2nm. It allows wider design windows for AlN/GaN RF HEMT design for given specifications as evident from the $f_T.L_g$ (GHz.$\mu$m) product in Fig. 8(c). For a smaller aspect ratio, the short channel effects dominate, deteriorating the RF performance. Hence, barrier layer scaling is a critical design aspect when the gate length is shrunk to maximize the cut off frequency without compromising with device linearity.

## VI. PARTIAL RECESSED GATE HEMT FOR IMPROVED MMW PERFORMANCE AND NON-LINEARITY

As discussed earlier, to maximize cut-off frequency without compromising the linearity of the device, the barrier layer, and channel length must be scaled simultaneously. Barrier layer scaling, however, has an adverse impact on the device's ON state performance due to reduced 2DEG concentration in channel. Besides, it also amplifies the virtual gate effect, which can increase the nonlinearity contributed by drain field-dependent ionization of surface states. We propose a partially recessed gate stack design, as depicted in Fig. 9(a),

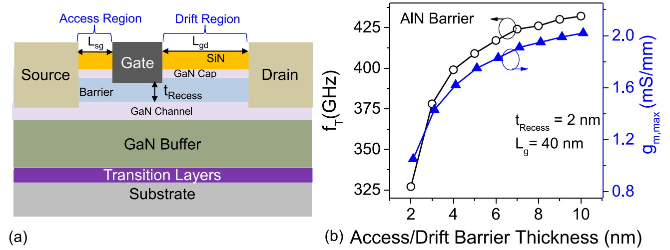


**FIGURE 9. (a) Cross-sectional view of HEMT with a partially recessed barrier under the channel. (b) Unity-gain frequency as a function of drift/access region barrier thickness while using a partially recessed channel with a recess depth of 2nm. $f_T$ improves substantially as transconductance ($g_m$) increases with access region thickness.**

to recover the device's ON-state performance, improve RF performance and the linearity behavior. Fig. 9(b) shows the cut-off frequency ($f_T$) of the partially recessed design. Here the AlN barrier layer below the gate is etched to have a reduced thickness of 2nm, whereas the thickness of the barrier layer in access/drift regions is increased. Fig. 9(b) shows that $g_m$ and as a result of which $f_T$ improves when the barrier thickness in the access/drift regions is increased, owing to improved $R_{ON}$ and ON current. The partial gate recess ensures better gate control and optimum electrostatics in the channel, which mitigates short channel effect induced nonlinearity. The access and drift region resistance is determined by the 2DEG density in these regions. As the barrier thickness is well above the critical thickness, the un-etched barrier regions do not suffer from high resistivity. This, in combination with the lateral scaling result in high cut off frequency. It is worth highlighting that an improved DC performance results in higher power density at a given frequency. The optimized design offers both - improved frequency performance and power density.

### A. LATERAL DEVICE SCALING

Fig. 10 shows the added advantages of partial recess gate stack design in terms of maximizing RF performance by

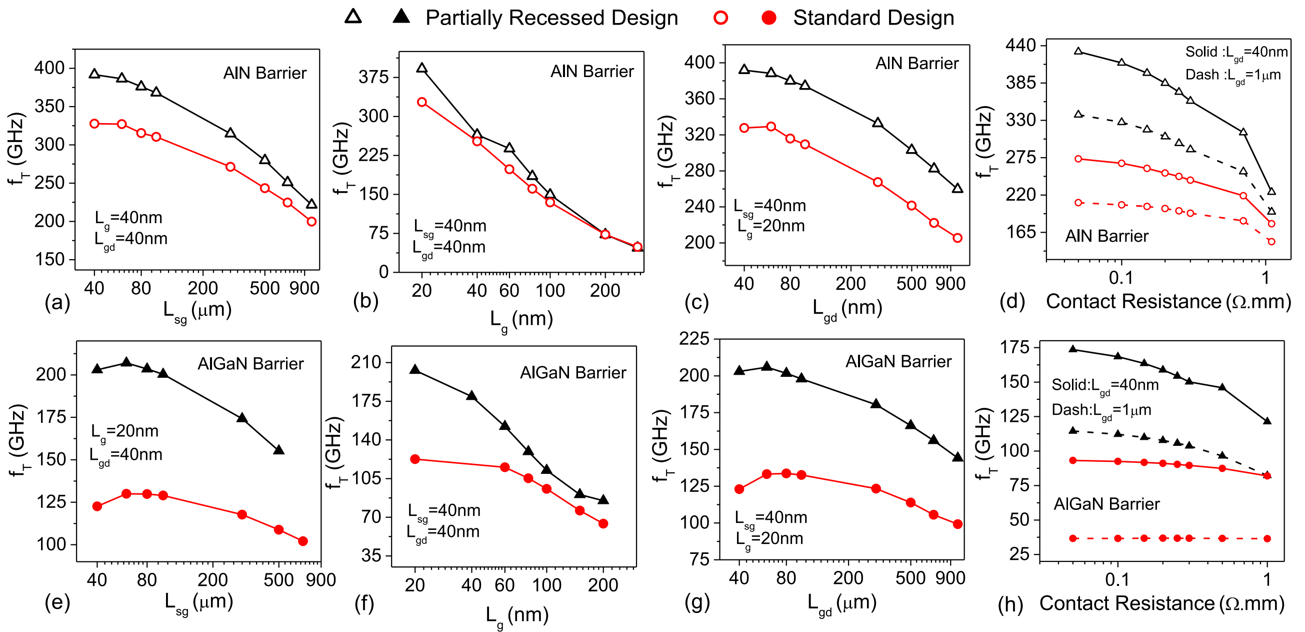


**FIGURE 10. Impact of scaling (a),(d) Gate to Source spacing ($L_{sg}$); (b),(f) Channel length ($L_g$); (c)(f) Gate to Drain spacing ($L_{gd}$) and (d),(f) contact resistance on the unity gain frequency ($f_T$) for AlN/GaN and AlGaN/GaN HEMT having partially-recessed and conventional design architectures.**

device scaling. Here the partially recessed design is implemented for both AlN and AlGaN barrier with recess thickness of 2nm ($t_{Barrier,AlN} = 3.5nm$) and 10nm ($t_{Barrier,AlGaN} = 25nm$) respectively. The performance of standard design (unrecessed) is also evaluated while scaling the lateral design parameters.

Figs. 10(a)-(c) and (e)-(g) show significant improvement in cut-off frequency for partially recessed gate design while scaling the lateral device dimensions - source to gate spacing ($L_{sg}$), gate length ($L_g$) and gate to drain distance ($L_{gd}$). Although the cut-off frequency for the AlN barrier is higher compared to the AlGaN barrier, the relative improvement in $f_T$ from standard to the partially-recessed device is significant in AlGaN HEMT. It is ascribed to deeper localized barrier etching below gate in case of AlGaN barrier compared to the AlN barrier, therefore higher relative increase in transconductance in AlGaN/GaN devices. It is worth observing that the short channel effects as results of lateral scaling are considerably suppressed in the AlN barrier and recessed device designs. It corroborates with the earlier discussion on the importance of better electrostatic control in the gate region. The short channel effects are dominant in standard AlGaN barrier design as evident by Fig. 10(e)-(g). The recessed architecture effectively mitigates the non-uniform field distribution in the channel by improving the aspect ratio ($\frac{L_g}{t_{Barrier}}$).

### B. ROLE OF S/D CONTACTS

It is desirable to minimize the ohmic contact resistance to reduce the external parasitic delay in the device. However, the impact of contact resistance needs to be studied from the device design perspective. There are several questions that have to be addressed for instance - In what conditions the contact resistance becomes an indispensable parameter so that the techniques such as regrowth must be used to maximize the performance? Does the contact resistance have any dependence on the barrier type or device architecture? We have performed a systematic study using simulations to understand the role of contacts on cut-off frequency. Both AlN and AlGaN barrier architectures have been analyzed with varying contact resistance. In addition, the device designs- standard and partially recessed barrier are simulated and the cut-off frequency behavior is shown in Fig. 10(d) and (h) for AlN and AlGaN barrier respectively. Following observations can be made from the results: (1) the increase in $f_T$ for AlN barrier is higher compared to AlGaN barrier design, (2) devices with low drift region length ($L_{gd}$ = 40nm), show strong dependence on contact resistance, (3) for high drift region length ($L_{gd} = 1\mu m$) devices, the cut-off frequency dependence decreases significantly, (4) the recessed devices show rapid improvement in cut-off frequency as contact resistance is reduced compared to standard barrier architecture, These observations are interpreted as follows.

AlN barrier, accounting for higher polarization charges, induces larger 2DEG compared to the AlGaN barrier. It signifies, that the ratio of $\frac{R_{on}}{R_{contact}}$ is considerably higher in the AlGaN barrier compared to the AlN barrier design. Any improvement in contact resistance, thus has a larger contribution in the reduction of parasitic resistance in the case of the AlN barrier, as observed in Fig. 10(d) and (h). Similarly, increasing the drift length ($L_{gd}$) results in high $\frac{R_{on}}{R_{contact}}$ ratio.

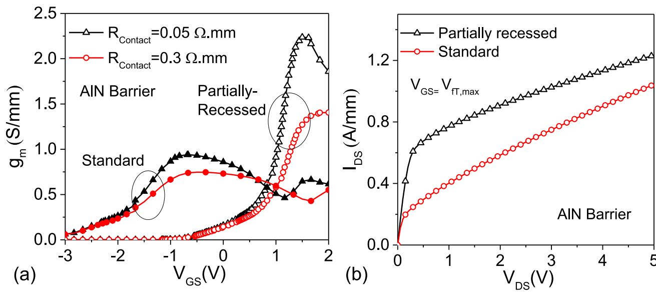


**FIGURE 11. (a) Change in transconductance with improving contact resistance in AlN barrier design with standard and partially recessed architectures. (b) Device output characteristics extracted at respective $f_{T,max}$ gate bias voltages in case of partially and recessed barriers. The simulations are performed for $L_g = 20$nm.**

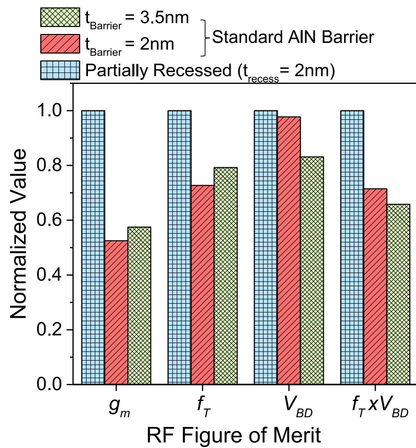


**FIGURE 12. RF Figure of merit parameters for standard design with 3.5nm and 2nm barrier, compared with the partially recessed design ($t_{recess} = 2$nm).**

It reduces the sensitivity of the cut-off frequency on contact resistance. In the case of the recessed device, the impact of contact resistance is considerable compared to the standard barrier device. Comparing the transconductance of the two architectures with varying contact resistance reveals that for partially recessed design, improvement in $g_m$ with decreasing contact resistance is substantial compared to standard architecture as depicted in Fig. 11(a). Due to a positive shift in threshold voltage, the gate bias voltage at which $f_T$ peaks shifts to positive in case of partially recessed HEMT. At their respective $f_{T,max}$ bias voltages, the partially recessed device has higher drain current as shown in Fig. 11(b). It translates to low $\frac{R_{on}}{R_{contact}}$ ratio for partially recessed architecture as discussed above, hence higher cut-off frequency at low contact resistance is observed.

To summarize, Fig. 12 compares the AlN/GaN HEMT device's RF figure of merit parameters of the three device structures – partially recessed gate structure ($t_{recess} = 2nm$) and standard structures with a barrier thickness of 3.5 nm and 2 nm. It shows that the optimized partially recessed gate design outperforms the other two standard designs in terms of all the RF performance FOM parameters.

## VII. CONCLUSION

Using a physics-based computational modeling framework, design guidelines for AlGaN/GaN and AlN/GaN HEMTs have been presented to improve it's mmW performance and linearity behavior. The impact of surface traps on the HEMT's mmW performance is studied. We have found that the spread of ionized surface traps across the GaN surface, above the drift region, increases with the drain field. This was found to add serious non-linearity in HEMT's RF behavior. Cut-off frequency roll-off in devices with aggressive gate scaling was attributed to short-channel effects and poor gate electrostatics. This in conjunction with drain field-dependent surface trap ionization magnified non-linearity. The AlN/GaN HEMT devices were found to have better linearity characteristics compared to AlGaN/GaN HEMT at large drain bias and sub 100nm channel lengths. It was found that the key to improve mmW performance and reduce non-linearity was improved electrostatics in the channel. This was achieved by demonstrating a partially-recessed gate design. Moreover, compared with standard (non-recessed gate) design, improvement in RF performance was found to be higher in AlN/GaN HEMT and partially-recessed barrier design in conjunction with lateral device scaling and contact resistance reduction. Physical insights to explain these trends have been developed. Finally, the computational modeling based device design approach has demonstrated the potential to achieve higher $f_T.V_{BD}$ product for mmW HEMTs, when compared to conventional designs.

## REFERENCES

[1] M. Shrivastava, M. S. Baghini, H. Gossner, and V. R. Rao, "Part II: A novel scheme to optimize the mixed-signal performance and hot-carrier reliability of drain-extended MOS devices," *IEEE Trans. Electron Devices*, vol. 57, no. 2, pp. 458–465, Feb. 2010, doi: 10.1109/TED.2009.2036799.

[2] S. Karmalkar and U. K. Mishra, "Enhancement of breakdown voltage in AlGaN/GaN high electron mobility transistors using a field plate," *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1515–1521, Aug. 2001, doi: 10.1109/16.936500.

[3] F. Medjdoub *et al.*, "Barrier-layer scaling of InAlN/GaN HEMTs," *IEEE Electron Device Lett.*, vol. 29, no. 5, pp. 422–425, May 2008, doi: 10.1109/LED.2008.919377.

[4] W. Saito, M. Kuraguchi, Y. Takada, K. Tsuda, I. Omura, and T. Ogura, "Influence of surface defect charge at AlGaN-GaN-HEMT upon Schottky gate leakage current and breakdown voltage," *IEEE Trans. Electron Devices*, vol. 52, no. 2, pp. 159–164, Feb. 2005, doi: 10.1109/TED.2004.842710.

[5] W. Saito *et al.*, "High breakdown voltage AlGaN-GaN power-HEMT design and high current density switching behavior," *IEEE Trans. Electron Devices*, vol. 50, no. 12, pp. 2528–2531, Dec. 2003, doi: 10.1109/TED.2003.819248.

[6] W. Saito *et al.*, "Suppression of dynamic on-resistance increase and gate charge measurements in high-voltage GaN-HEMTs with optimized field-plate structure," *IEEE Trans. Electron Devices*, vol. 54, no. 8, pp. 1825–1830, Aug. 2007, doi: 10.1109/TED.2007.901150.

[7] V. Joshi, A. Soni, S. P. Tiwari, and M. Shrivastava, "A comprehensive computational modeling approach for AlGaN/GaN HEMTs," *IEEE Trans. Nanotechnol.*, vol. 15, no. 6, pp. 947–955, Nov. 2016, doi: 10.1109/TNANO.2016.2615645.

[8] V. Joshi, S. P. Tiwari, and M. Shrivastava, "Part I: Physical insight into Carbon-doping-induced delayed avalanche action in GaN buffer in AlGaN/GaN HEMTs," *IEEE Trans. Electron Devices*, vol. 66, no. 1, pp. 561–569, Jan. 2019, doi: 10.1109/TED.2018.2878770.

[9] V. Joshi, S. P. Tiwari, and M. Shrivastava, "Part II: Proposals to independently engineer donor and acceptor trap concentrations in GaN buffer for ultrahigh breakdown AlGaN/GaN HEMTs," *IEEE Trans. Electron Devices*, vol. 66, no. 1, pp. 570–577, Jan. 2019, doi: 10.1109/TED.2018.2878787.

[10] E. Bahat-Treidel, O. Hilt, F. Brunner, J. Würfl, and G. Tränkle, "Punchthrough-voltage enhancement of AlGaN/GaN HEMTs using AlGaN double-heterojunction confinement," *IEEE Trans. Electron Devices*, vol. 55, no. 12, pp. 3354–3359, Dec. 2008, doi: 10.1109/TED.2008.2006891.

[11] C. Nasser, P. Luo, F. Schnieder, M. Rudolph, and D. Ritter, "Simulation of the RF power performance of a GaN HFET and comparison to experiment," *IEEE Trans. Electron Devices*, vol. 66, no. 5, pp. 2146–2150, May 2019, doi: 10.1109/TED.2019.2907484.

[12] D. Guerra, M. Saraniti, D. K. Ferry, and S. M. Goodnick, "Linearity analysis of millimeter wave GaN power transistors through X-parameters and TCAD device simulations," in *IEEE MTT-S Int. Microw. Symp. Dig.*, 2013, pp. 1–4, doi: 10.1109/MWSYM.2013.6697725.

[13] M. L. Schuette *et al.*, "Gate-recessed integrated E/D GaN HEMT technology with $f_t/f_{max} > 300$ GHz," *IEEE Electron Device Lett.*, vol. 34, no. 6, pp. 741–743, Jun. 2013, doi: 10.1109/LED.2013.2257657.

[14] J. W. Chung, W. E. Hoke, E. M. Chumbes, and T. Palacios, "AlGaN/GaN HEMT with 300-GHz $f_{\max}$," *IEEE Electron Device Lett.*, vol. 31, no. 3, pp. 195–197, Mar. 2010, doi: 10.1109/TED.2019.2907484.

[15] K. Shinohara *et al.*, "Deeply-scaled self-aligned-gate GaN DH-HEMTs with ultrahigh cutoff frequency," in *Proc. Int. Electron Devices Meeting (IEDM)*, 2011, pp. 1–4, doi: 10.1109/IEDM.2011.6131582.

[16] P. J. Tasker and B. Hughes, "Importance of source and drain resistance to the maximum $f_T$ of millimeter-wave MODFETs," *IEEE Electron Device Lett.*, vol. 10, no. 7, pp. 291–293, Jul. 1989, doi: 10.1109/55.29656.

[17] K. Shinohara *et al.*, "Electron velocity enhancement in laterally scaled GaN DH-HEMTs with $f_T$ of 260 GHz," *IEEE Electron Device Lett.*, vol. 32, no. 8, pp. 1074–1076, Aug. 2011, doi: 10.1109/LED.2011.2158386.

[18] Y. Yue *et al.*, "InAlN/AlN/GaN HEMTs with regrown ohmic contacts and $f_T$ of 370 GHz," *IEEE Electron Device Lett.*, vol. 33, no. 7, pp. 988–990, Jul. 2012, doi: 10.1109/LED.2012.2196751.

[19] A. G. Chynoweth, "Ionization rates for electrons and holes in silicon," *Phys. Rev.*, vol. 109, no. 5, pp. 1537–1540, 1958, doi: 10.1103/PhysRev.109.1537.

[20] G. Masetti, M. Severi, and S. Solmi, "Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon," *IEEE Trans. Electron Devices*, vol. ED-30, no. 7, pp. 764–769, Jul. 1983, doi: 10.1109/T-ED.1983.21207.

[21] R. Vetury, N. Q. Zhang, S. Keller, and U. K. Mishra, "The impact of surface states on the DC and RF characteristics of AlGaN/GaN HFETs," *IEEE Trans. Electron Devices*, vol. 48, no. 3, pp. 560–566, Mar. 2001, doi: 10.1109/16.906451.

[22] R. Vetury *et al.*, "Direct measurement of gate depletion in high breakdown (405 V) AlGaN/GaN heterostructure field effect transistors," in *Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, 1998, pp. 55–58, doi: 10.1109/IEDM.1998.746245.

[23] D. S. Lee *et al.*, "Device delay in GaN transistors under high drain bias conditions," *IEEE Electron Device Lett.*, vol. 34, no. 7, pp. 849–851, Jul. 2013, doi: 10.1109/LED.2013.2262034.