Part I: On the Unification of Physics of Quasi-Saturation in LDMOS Devices

B. Sampath Kumar and Mayank Shrivastava, *Senior Member, IEEE*

***Abstract*— There have been a lot of ambiguities related to physics of quasi-saturation (QS) in laterally diffused MOS (LDMOS) devices in the published literature. For example, models that explain QS in input characteristics do not explain the same in output characteristics and vice versa. In addition to this, none of the earlier models explain early onset of QS at higher temperatures nor the models were validated using counter arguments. Attributed to this, a need for unified theory explaining physics of QS is justified in this paper. Furthermore, this paper for the first time, while addressing missing links between the observations reported in the past, develops a unified theory to explain physics of QS behavior. The theory presented here is independent of device architecture and covers all voltage-current–temperature trends. While considering velocity saturation and space charge modulation, we have discovered key role of high field mobility degradation of majority carriers and electric field screening, which is found to be the root cause of QS in LDMOS devices. The theory presented is further validated with numerous counter arguments. Finally, based on the new physical insight developed, we have proposed different approaches to mitigate QS effect. A detailed device design guideline to mitigate QS and its correlation with analog/RF performance, electo static discharge, hot-carrier reliability, self-heating, and safe operating area concern is presented in Part II of this paper.**

***Index Terms*— Drain extended MOS, electric field screening, Kirk effect, laterally diffused MOS (LDMOS), mobility degradation and space charge modulation (SCM), quasi-saturation (QS).**

I. Introduction

HIGH-voltage (HV) integrated/discrete semiconductor devices such as drain extended MOS or laterally diffused MOS (LDMOS) cater to a wide range of HV/high-power switching and RF applications like RF power amplifier, level shifters, line driver, power management, and motor drives. These devices are commonly designed with extended drain to block higher drain voltages by reducing the surface electric field, which mandates careful design of the drain extension or drift region. However, in order to obtain higher breakdown voltage by drift profile engineering, one has to sacrifice ON resistance (R_{ON}), output resistance ($1/g_ds$), and transconductance (g_m) of the device, which eventually lowers the switching as well as RF performance of these devices. Fundamentally, such a design challenge is imposed by early quasi-saturation (QS) in the HV devices. QS in LDMOS devices can be visualized as gate independent drain current as depicted using the I_D–V_G and I_D–V_D characteristics in Fig. 1. When the device enters into the QS regime, gate relinquishes its control on channel and hence the drain current, which severely degrades the g_m [1] and does not allow R_{ON} to reduce further as a function of gate voltage. This undesirable effect not only severely affects g_m but also creates unintended nonlinearities in the device's miller capacitance [2]. Moreover, though the safe operating area (SOA) defines the electrical and thermal extremes [3] of the device, QS further limits the operation boundaries. Therefore, mitigating QS in LDMOS devices is the key to its usability for desired applications. However, this requires a thorough understanding of physics of QS. There have been several investigations on physics of QS in the past [4]–[8], as summarized below.

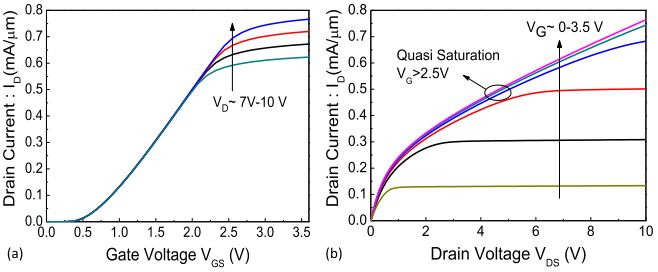

Fig. 1. *I–V* characteristics of an LDMOS device depicting QS phenomena. (a) Insensitive drain current with respect to gate voltage for different drain voltages. (b) Drain voltage-dependent drain current for different gate voltages. Here onset of QS is defined at the (a) gate voltage after which transconductance drops significantly or (b) gate voltage at which drain current does not saturate with drain voltage.

Darwish [9] attributed QS to early velocity saturation of mobile carriers in the drift region and highlighted that it depends on body spacing and drift region doping concentration. Kreuzer *et al.* [10] proposed that QS is due to drift resistance modulation under high current injection condition, however, failed to offer physical insights. Evans and Amaratunga [11] proposed formation of dipole under the gate/in the drift region, which causes charge imbalance in the drift region and leads to drain voltage-dependent drain current.

Manuscript received September 3, 2017; revised October 21, 2017; accepted November 20, 2017. Date of current version December 27, 2017. This work was supported in part by the IMPRINT Program of Government of India and in part by the Department of Science and Technology. The review of this paper was arranged by Editor G. Ghione. *(Corresponding author: B. Sampath Kumar.)*

The authors are with the Advance Nanoelectronic Device and Circuit Research Group, Department of Electronic Systems Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: bsampath@dese.iisc.ernet.in; mayank@dese.iisc.ernet.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2017.2777004

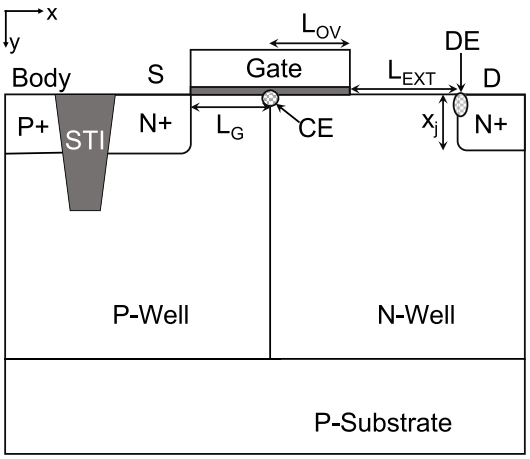

Fig. 2. Schematic of an LDMOS device depicting critical positions like CE and DE to study the physics of QS. Here $L_G = 200$ nm, $L_{OV} = 200$ nm, $L_{EXT} = 250$ nm, $x_j = 50$ nm, N-well doping $= 7 \times 10^{16}$ cm^{-3}, N-well depth $= 500$ nm, and oxide thickness $= 4$ nm. DE is the boundary between N-N+ junction; similarly, CE is the boundary between channel and N-well. At moderate currents, it is relatively broader, which, however, converges to a narrower high-field spot as the device is forced into deep QS.

Few other works attributed QS to Kirk effect [space charge modulation (SCM)] [12], [13], which reduces the effective V_{DS} seen by the MOS channel and moves the transistor into linear region. Besides these, finger spacing dependence is reported in [14] and impact of device heating on QS is reported in [15], which narrates worse extremes of QS.

Most of the theories highlighted above contradict in one or the other aspects. For example, the theories presented in the context of input characteristics do not explain output characteristics in the QS region, and vice versa. The physical events presented elsewhere are indeed correct and can be reproduced using the technology CAD; however, they do not offer a unified understanding of the QS behavior. This mandates developing a unified theory to explain the physics of QS, which—as presented in Section IV—also helps in drawing design guidelines to mitigate QS. The ambiguities and contradictions in various theories presented in the past are discussed in Section II. Based on detailed physical insights, a well-connected theory that unifies the physics of QS is presented in Section III. Section IV discloses few device designs to mitigate QS behavior. Finally, this paper is concluded in Section V.

II. Physics of Quasi-Saturation: Ambiguities and Contradictions

This section revisits the physics of QS presented in earlier works and highlights the contradictions or missing links in different theories presented till date. Fig. 2 depicts the schematic of a lateral LDMOS device, used in this paper to study the physics of QS. Critical locations within the LDMOS device, to study the physics of QS, are marked as channel edge (CE) and drain edge (DE), respectively. The device design is chosen in such a way that all attributes that are discussed in previous works can be captured and explored further.

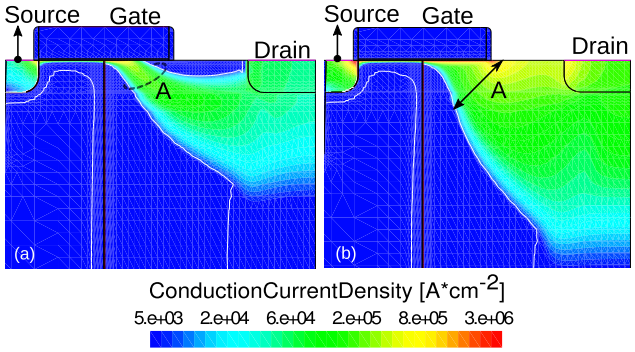

Fig. 3. Conduction current density depicting change in the conduction cross section with the applied gate bias. (a) Before QS, simulation depicts restricted current conduction due to depletion under the gate oxide with cross section A. (b) After QS, simulation shows carrier spreading deep into the drift region with an increased conduction area.

A. Velocity Saturation

Darwish [9] proposed velocity saturation of mobile carriers in the drift region to be the root cause for early QS in vertical LDMOS devices. The model was later extended to lateral LDMOS devices as well. It was proposed that the spacing between the cells/fingers of vertical transistors and doping concentration in the drift region sets the onset of velocity saturation. At higher gate voltages, attributed to proportionately increased inversion charge in the channel, channel's conductivity increases, which increases excess carrier injection in the drift region. This is depicted in Fig. 3 as change in conduction cross section "A" as a function of gate bias. It was shown that beyond a given gate voltage, with an increase in the drain voltage, the electric field in the drift region increases and accelerates the carriers to achieve maximum velocity (υ_{sat}), which results in early velocity saturation. Once υ_{sat} is reached, drain current becomes insensitive to gate field and hence the device experiences QS [9].

1) Observations: The theory presented above for IDS–VGS characteristics was found not complete and also contradicts with the same presented elsewhere for IDS–VDS characteristics. It is trivial to visualize from (2) and (2) [16] that velocity saturation of mobile carriers in the drift region should not make the drain current insensitive to V_{GS} as the channel carrier concentration (n) is independent of υ_{sat} and increases only with respect to gate voltage. Hence, this model does not explain insensitivity of drain current at higher gate voltages [Fig. 1(a)]. Moreover, explanation for shift in onset of QS with increasing drain voltage [Fig. 1(a)] is missing too. As per the proposed theory above, at higher drain voltage (i.e., drain field), drift region should experience an early velocity saturation; in other words: 1) with increase drain voltage, onset of QS should come down and 2) under QS condition, drain current should be insensitive to drain field. However, Fig. 1 depicts an opposite trend. Under QS, drain current increases as a function of drain voltage; moreover, gate voltage required for onset of QS increases with increased drain field. Finally, it was earlier presented that the QS becomes even severe at elevated temperatures [17]; since the elevated temperature does not drastically change υ_{sat}, the theory presented above does not explain an early QS with elevated lattice

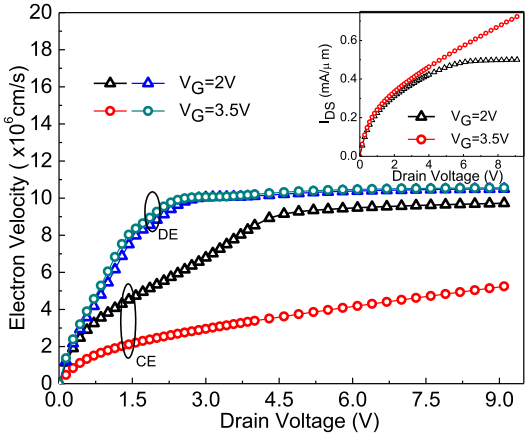

Fig. 4. Carrier velocity at CE and DE, extracted before and after QS. Inset: the I_D versus V_D before and after QS.

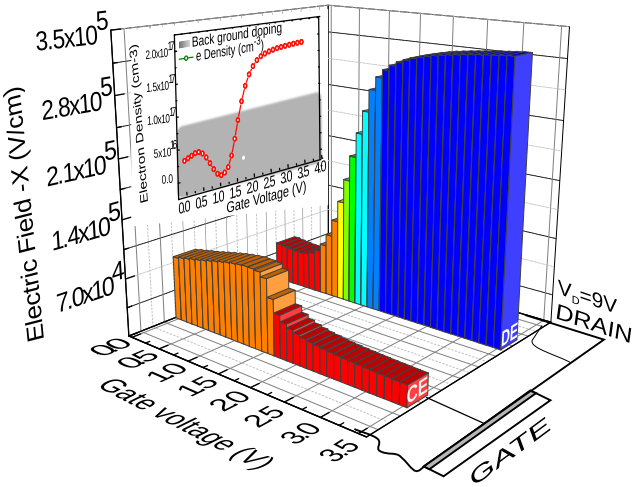

Fig. 5. Electric field at DE and CE as a function of gate voltage. Inset: electron density as a function of V_G, where the shaded region represents the background doping.

temperature

$$\frac{I}{A} = nq\upsilon_d \tag{1}$$

$$\upsilon_d(E) = \frac{\mu_n E}{[1 + (\mu_n E/\upsilon_{\text{sat}})]}. \tag{2}$$

To validate our arguments further, carrier velocity at the CE and DE was extracted as a function of drain voltage before and after QS. Fig. 4 depicts the presence of velocity saturation in drift region and υ_{sat} at DE being independent of the presence or absence of QS. Based on this observation, one can conclude that drift region of LDMOS device in QS can experience velocity saturation; however, the converse may not be true always. The model presented in [9] highlights that the onset of QS is linked with carrier velocity saturation at the DE. Though the observation is found to be correct under certain bias conditions, but is found not to be universal. Moreover, it does not explain why ID–VD characteristics under QS regime show a liner I–V relationship at higher drain field when the carrier velocity is expected to saturate. Similarly, it does not explain shift in the onset (gate voltage and drain current) for QS with increasing drain potential, as for higher drain field, carrier velocity is expected to saturate faster (i.e., at lower gate voltage and drain current).

B. Drift Region Resistance

Conductivity modulation of drift region is one of the reasons presented in the past to explain QS behavior [10], [11], [14], [17]. The model explains that at higher gate overdrive channel's conductivity is significantly higher than the same of drift region and therefore it can be virtually treated as a short [10]. In this case, the overall resistance is then due to drift region resistance. Considering drift carriers to be under velocity saturation, linear dependency of I_{DS} on V_{DS} is explained using change in carrier concentration n and conduction cross section "A" with drain voltage [10], [11]. It was proposed that A dramatically increases with drain voltage and n falls below drift region doping concentration (N_D). This uneven charge distribution along the drift region leads to formation of a dipole in the drift region under the gate, which in turn increases electric field and causes velocity saturation. Linear rise in drain current with V_{DS} after QS, even when υ_d was saturated, was attributed to increase in conduction cross section "A" [see (2)].

1) Observations: The model presented above explains I_{DS} versus V_{DS} characteristics; however, it fails to explain I_{DS} versus V_{GS} characteristics. Even if the carriers are driven into velocity saturation and depend on 'A', n strongly depends on V_{GS}. Therefore, as per the model presented above, the current must increase with increase in n and hence with increase in V_{GS}. It is worth highlighting again that cross section "A" increases with gate voltage, as depicted in Fig. 3. To validate this argument, a simulation by restricting N-well depth and hence cross-sectional area "A" was performed. According to the model presented above, if "A" is restricted to a constant value and hence does not allow it to change with V_{DS}, drain current should not change with V_{DS}. However, simulation predicted (data not shown here) just opposite of this hypothesis, which challenges the validity of this model. Moreover, since the elevated temperature does not change current cross-sectional area "A," the theory presented above does not explain the onset of QS versus temperature trends. Impact of conductivity modulation is discussed in detail in Section III. At this stage, it would not be an exaggeration to conclude that the observations presented in [10], [11], [14], and [17] might be correct; however, they were not well connected with each other and do not derive a convincing or unified theory to explain physics of QS.

C. Kirk Effect or Space Charge Modulation

Wang *et al.* [13] suggested Kirk effect or SCM in the drift region to be responsible for QS. In principle, by increasing V_{DS}, the electrostatic potential at the CE (V_{CE}) increases and drives the transistor from linear to saturation region. However, it was observed that after SCM, which occurs when majority carrier concentration is higher than background doping, peak electric field shifts to DE and V_{CE} falls below the pinch-off voltage. This phenomenon can be seen in Fig. 5, which depicts change in: 1) electric field at CE and DE and 2) electron density as a function of gate

voltage. When the injected charge density is higher than the effective background doping, electric field at DE increases significantly whereas the same is lowered at CE. This phenomenon is called Kirk effect [12], [13] or SCM [3]. As V_{CE} falls below a threshold (pinch-off voltage), it takes the device into linear region and hence drain current increases as drain voltage is increased after a certain gate voltage, which was presented as QS.

1) Observation: The model presented above does not explain insensitivity of drain current as a function of gate voltage after the onset of QS, as depicted in the I_{DS} versus V_{GS} characteristics (Fig. 1). As per this model, for fixed V_{DS}, once the device enters into linear region, drain current saturates with V_{GS}. However, SCM is solely a current injection phenomenon over the background doping. The model does not correlate, why there is a different on set current required for different V_{DS}. Finally, given the fact that SCM, which solely depends on background doping and injected carrier concentration, is independent of change in lattice temperature, the theory presented above does not explain early onset of QS with increasing temperature.

D. Summary

The observations presented in earlier works are indeed correct; however, the theories derived out of that have ambiguities and are not well connected. For example, models that explain I_{DS} versus V_{GS} characteristics do not explain I_{DS} versus V_{DS} and vice versa. In addition to this, none of these models explain early onset of QS at higher temperatures. Moreover, these models have never been validated using counter arguments. Therefore, a need for unified theory explaining physics of QS, while connecting all attributed and observations, is highly justified. Section III attempts to probe the root cause of all observations presented so far, connects earlier observations with new findings, and develops the unified understanding of QS, while also using counter arguments and experiments to validate the proposed theory.

III. Unification of Physics of Quasi-Saturation

A. Space Charge Modulation and Hot Spot Formation

Among various observations presented before, one distinct feature that stands throughout is the presence of SCM at the onset of QS, as depicted in Fig. 5. The electrostatic as well as carrier transport behavior across the device undergoes a number of changes after SCM. Such as, an early *velocity saturation* in the drift region can be observed due to the increased electric field after the SCM. Moreover, depletion region at the well junction shrinks to a narrower region as a result of SCM, which increases the conduction cross-sectional area "A," as depicted in Fig. 6. This further enhances SCM [11] and increases electric field as depicted in Figs. 3 and 6(c). These observations indicate that the increased electric field, attributed to SCM, is a necessary condition for the onset of QS. However, an increased electric field is not sufficient enough to unify the physics of QS and clarify the interdependency between the input and output characteristics. In order to present a unified theory, the next sections present the missing links/observations.

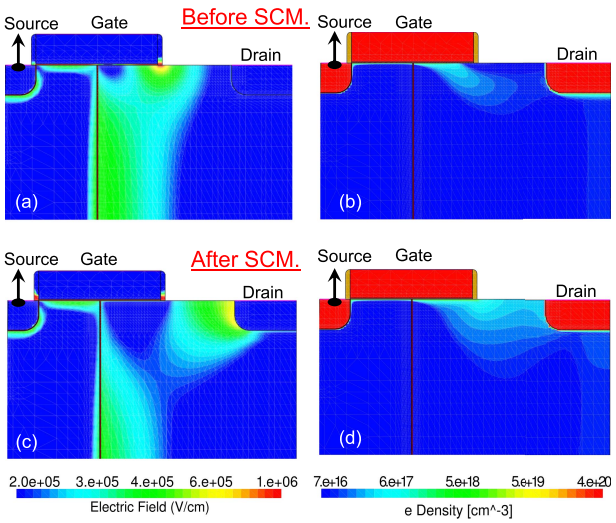

Fig. 6. (a) and (c) Electric field distribution across the device before and after SCM. (b) and (d) Electron density before and after SCM.

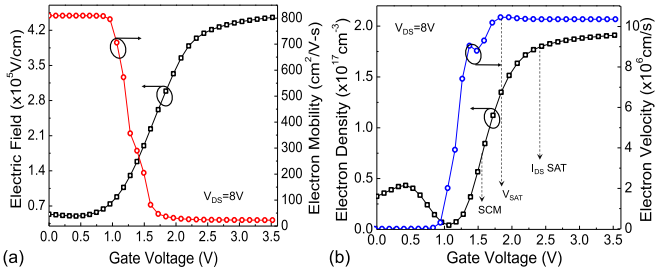

Fig. 7. (a) Change in electric field and electron mobility as a function of gate voltage. (b) Change in electron density and carrier velocity as a function of gate voltage (data were extracted at DE).

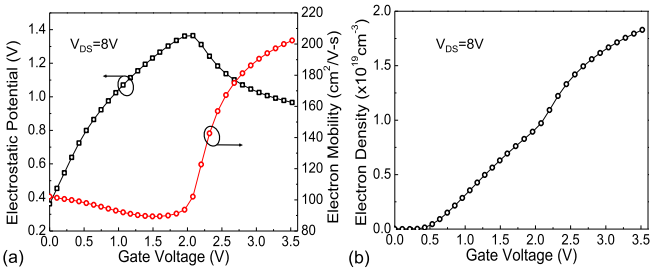

Fig. 8. Dynamics of (a) electrostatic potential and electron mobility and (b) electron density as a result of SCM. The data are extracted at CE, and the gate voltage at which channel potential drops is termed onset of QS.

B. Mobility Degradation and Field Screening

Aftermath SCM, one of the crucial parameters to look into is mobility degradation.

Figs. 6(c) and 7(a) show localized SCM and hot spot formation near DE, as the gate voltage is increased. Interestingly, a different trend can be seen at CE. Fig. 8(a) shows lowering of potential at CE as a function of gate bias, which recovers the mobility at CE. As a result, current density at CE continues

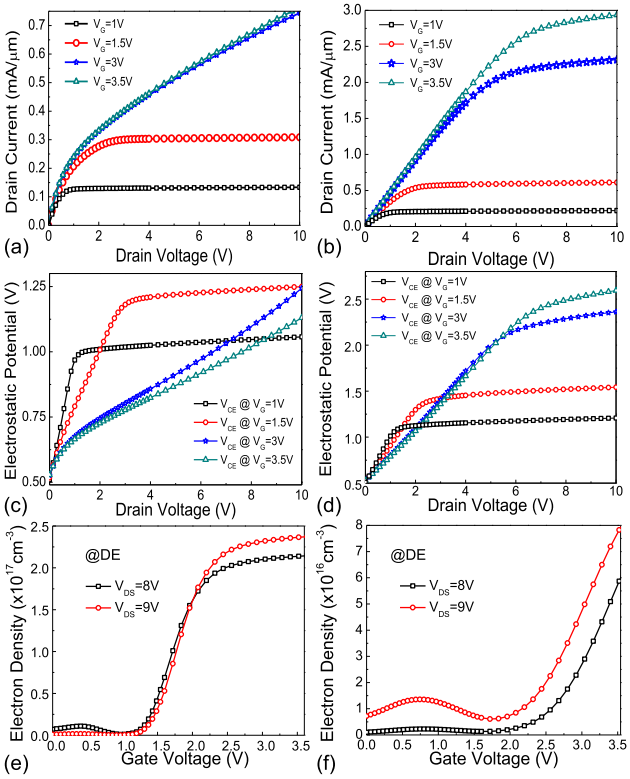

Fig. 9. (a) and (b) I_{DS} versus V_{DS} characteristic at higher and lower gate voltages. (c) and (d) V_{CE} versus V_{DS} relation at higher and lower gate voltages. (e) and (f) Electron density at DE versus V_{GS} at different drain voltages with [(a), (c), and (e)] and without [(b), (d), and (f)] high field mobility degradation model.

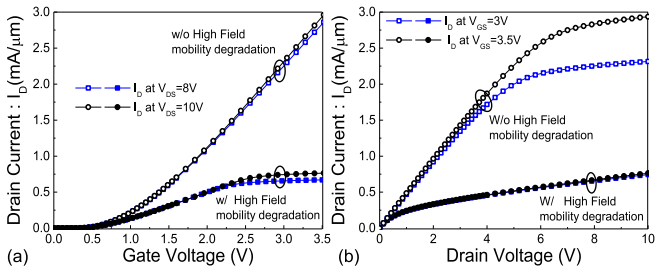

Fig. 10. (a) I_{DS} versus V_{GS} and (b) I_{DS} versus V_{DS} characteristics of the LDMOS device with and without high field mobility degradation effect.

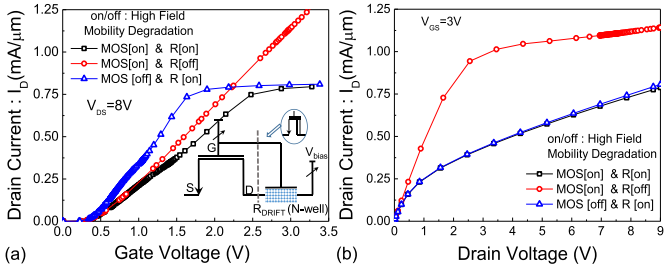

Fig. 11. (a) I_{DS} versus V_{GS} and (b) I_{DS} versus V_{DS} characteristics of LDMOS equivalent device, as depicted in the inset, for the following three conditions: 1) mobility degradation turned ON in both MOS and N-well resistor; 2) mobility degradation accounted only in MOS; and 3) mobility degradation accounted only in N-well resistor. Inset of (a) a MOS device in series with N-well resistor used to validate our model. N-well resistor mimics the drift region.

to grow with gate voltages, as depicted in Fig. 8(b), unlike the case at DE. These observations indicate QS to be related to DE, not CE, unlike discussed in previous works. A recovered mobility at CE, after SCM, attempts to inject an increased current in the drift region as gate voltage increases further. This, however, strengthens the SCM and continues to increase electric field inside the hot spot near DE. This leads to an early saturation of carrier velocity, which in turn degrades mobility near DE severely, as depicted in Fig. 7(b), and hinders flow of electrons through the DE region. It is worth highlighting that though the carrier velocity saturates early, drain current and current density continue to increase, as depicted in Fig. 7(b). Hence, velocity saturation alone cannot be the root cause of QS. As the injected carrier density increases further (defined by potential and mobility at CE), it gradually screens the electric field at DE ($\nabla^2\phi = -\nabla.E = -(\rho/\epsilon)$) and does not allow it to increase further. This slows down the mobility degradation at DE and causes mobility to saturate at a finite value. These competing events eventually limit the flow of current as soon as field screening and mobility saturation are achieved at DE. This is seen as saturated, i.e., insensitive, drain current as a function of gate voltage.

One should note that the driving force for carrier injection from channel to drift region (V_{CE}), as well as peak electric field at DE, in the QS region, increases with applied drain voltage [Fig. 9(c)], and hence the saturated drain current after QS increases with drain voltage [Fig. 9(a)]. This is because: 1) increased V_{CE} strengthens carrier injection, which increases electric field at DE and 2) higher electric field at DE requires higher carrier density to screen the electric field. This explains the linear dependence between I_{DS} and V_{DS} in the QS region [Fig. 9(a)]. To validate our model further, it is worth highlighting that when high field mobility degradation was not considered in simulations, QS effect was missing. This is evident from missing linear dependence between I_{DS} and V_{DS}, even at currents higher than required for QS [Fig. 9(b)]. Moreover, when high field mobility degradation was not considered, V_{CE} was always higher than pinch-off voltage and current density never saturates at higher gate voltages, independent of drain voltage applied [Fig. 9(f)].

C. Model Validation

It is worthwhile to validate the model presented above, before extending it further. To validate the role of mobility degradation in the onset of QS, LDMOS device is simulated with and without mobility degradation effect. This approach was specially instigated by omitting high field saturation model, so as to avoid mobility degradation in the drain extension region. Fig. 10 shows that without mobility degradation effect, QS completely disappears. This proves the role of mobility degradation in QS, as presented above. Furthermore, it is worth localizing and validating the weakest spot causing QS. Fig. 11 shows (a) I_{DS}–V_{GS} and I_{DS}–V_{DS} characteristics of a MOS device in series with an N-well resistor [inset of Fig. 11(a)]. The connections and design were made in such a way that the MOS and resistor stack mimics

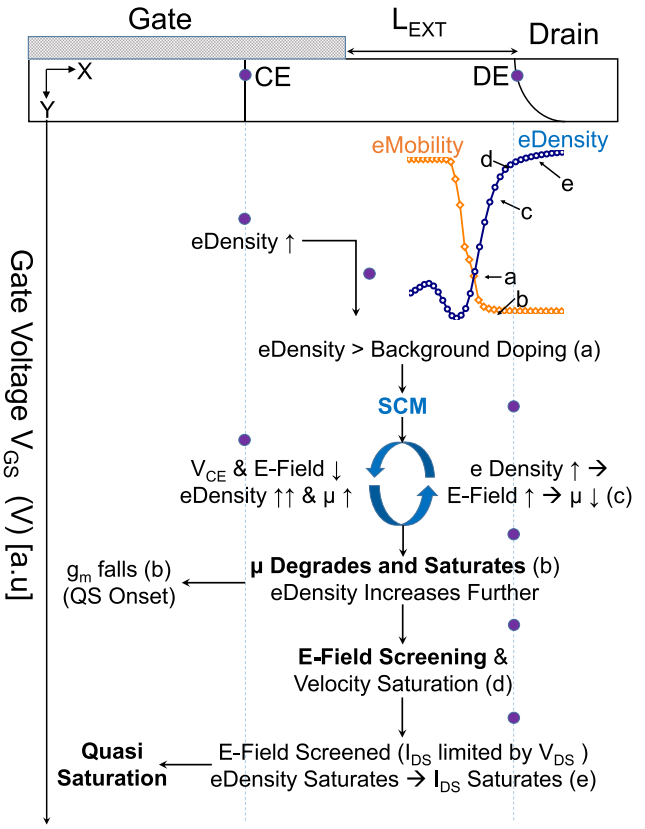

Fig. 12. Illustration to unify physics of QS by depicting all physical events with respect to the gate voltage at different locations across the device. Solid circles depict the physical location of respective physical event and arrows indicate mutual dependence or sequence of those events. QS onset is considered as the point (or gate bias) where g_m starts to roll off.

its parent LDMOS device's I–V characteristics. The following three cases were simulated: 1) mobility degradation in MOS and resistor was turned ON; 2) mobility degradation only in MOS was turned ON; and 3) mobility degradation only in N-well resistor was turned ON. Fig. 11 depicts that QS behavior vanishes only when mobility degradation was switched OFF in the N-well resistor, which mimics the drift region of the device. This proves that the mobility degradation after SCM in the drift region, in particular near DE, is the root cause for onset of QS.

The unified model for physics of QS, as explained above, is systematically summarized in the illustration presented in Fig. 12. The illustration depicts the series of physical events with respect to gate bias that take place at different locations in the device and lead to QS. It is worth highlighting that the proposed unified model is suitable for all voltage classes as well as types of LDMOS architectures. To validate this, similar investigations on a 120–600-V class SOI LDMOS device [18] and multiple reduced surface field device [19] were performed (data not shown here), which revealed the same trends in terms of SCM, mobility degradation, and electric field screening [20]. In fact, the QS behavior or SCM and mobility degradation were found to be increasingly severe for HV devices. Hence, it can be concluded that irrespective of the voltage class and device architectures, the underlying physics remains the same.

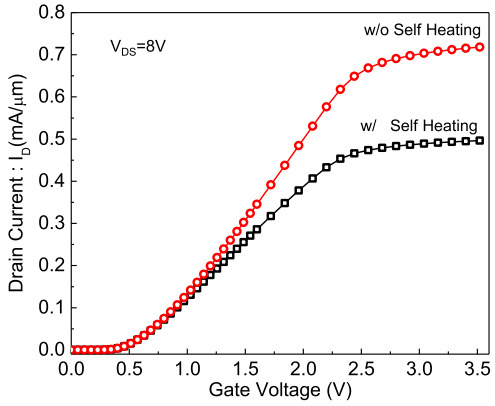

Fig. 13. I_{DS} versus V_{GS} characteristics of LDMOS device with and without self-heating to depict temperature dependence of QS.

D. Lattice Heating Versus Quasi-Saturation

Phonon population, i.e., lattice vibrations, increases when electrons (or holes) accelerated in the high electric field region undergo inelastic scattering with acoustic phonons. This event leads to hot spot formation, which elevates lattice temperature and (semiclassically) degrades electron mobility ($\mu \propto (1/{m_c}^{*5/2}T^{3/2})$, where m_c^* is the effective mass and T stands for temperature [16]). As mentioned earlier, mobility degradation at DE is found to be the root cause for QS. Inelastic scattering with acoustic phonons aids to mobility degradation along with field driven mobility degradation, and hence intensifies the SCM. Therefore, lattice heating drives the device into an early QS with an onset at lower drain current when compared with case without lattice heating, as depicted in Fig. 13. Once the device enters into QS regime, attributed to electric field peak at DE, lattice heating localizes at the same location. This also leads to peculiar electo static discharge (ESD) reliability and SOA concerns, which we have found to be closely linked with onset of QS and hence will be investigated in detail in Part II of this paper.

IV. How to Mitigate Quasi-Saturation?

This section proposes several ways of mitigating QS, i.e., engineering onset of QS, using the understanding build so far. It is evident that SCM is pushed to higher gate voltages or channel is forced to remain pinched off even if SCM exists, QS will get mitigated or delayed. Besides design guidelines, the idea is to correlate the unified theory presented in this paper with the improvements (in terms of onset of QS) gained from various design modifications (Fig. 14). The guidelines presented here have their own demerits; however, the idea to validate the theory is presented. Part II of this paper presents novel guidelines to mitigate QS and correlates QS with analog/RF performance, ESD, hot carrier reliability, and SoA concerns. A significant improvement in performance and reliability by mitigating QS will be presented in Part II by without significantly lowering the breakdown characteristics.

A. Drift Region Doping

Onset of QS was found to have a strong dependence on drift region doping and doping profile. Fig. 14(a) shows that

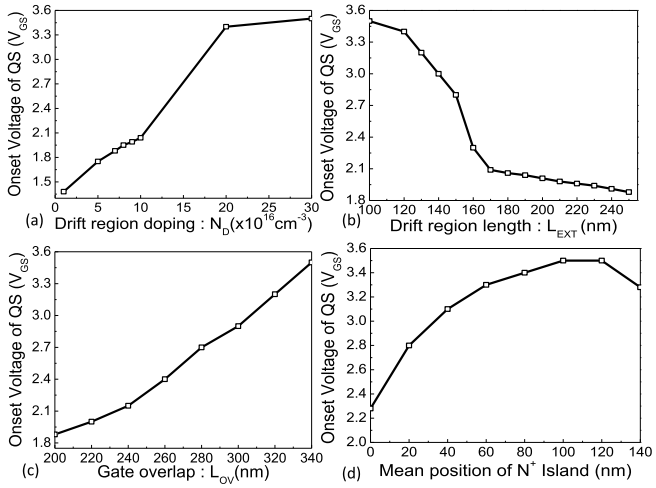

Fig. 14. Onset of QS as a function of (a) drift region doping, (b) drift region length, (c) gate overlap length, and (d) N-type pocket placement [as per the schematic depicted in Fig. 15(a)].

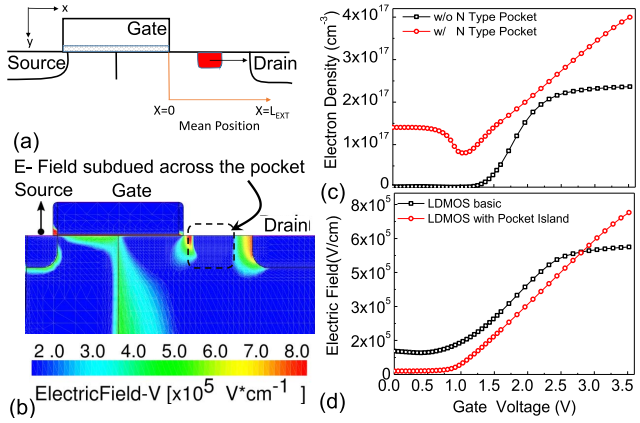

Fig. 15. (a) LDMOS device with the proposed N-type pocket. (b) Electric field profile across the device with N-type pocket after SCM. (c) Electron density at DE with and without N-type pocket. (d) Electric field at DE with and without N-type pocket.

the onset of QS shifts to higher gate voltages, i.e., higher drain currents, when drift region doping was increased. This is attributed to shift in onset of SCM to higher carrier densities, i.e., higher drain current, when the background concentration was increased. This improves the ON resistance of the devices; however, it sacrifices breakdown voltage. This validates the role of SCM in QS.

B. Drift Length

Drift region length (L_{EXT}) influences the onset of QS directly, as depicted in Fig. 14(b), by controlling the peak electric field at DE, after SCM. It should be clear, based on earlier discussion, that onset of SCM mostly depends on the background doping concentration and the injected charge density. Therefore, reducing L_{EXT} should not affect the onset of SCM. However, as L_{EXT} is reduced (increased), the peak electric field at DE, after SCM, falls (increases), which is attributed to an increased (reduced) sharing of space charge by well junction. Moreover, the potential at CE also increases (reduces). Reduced peak electric field at DE for shorter L_{EXT} delays the mobility degradation. This pushes the onset of QS to higher gate voltages as the device requires higher current for SCM to get stronger or the electric field at DE to grow further. Once very high electric field at DE is achieved, mobility drops significantly, which leads to QS, as explained in the previous section.

C. Gate Overlap

Fig. 14(c) depicts linear dependence between gate overlap and onset of QS. An increased gate overlap promotes mobile carriers to spread deep into the drift region. The carrier spreading relaxes the injected carrier concentration, which delays the SCM. This eventually delays the onset of QS and improves the device characteristics. It is worth highlighting that an increased gate overlap is not preferred for analog/RF applications because of higher miller capacitance and from hot carrier reliability point of view, which will be discussed in detail in Part II of this paper.

D. N-Type Pocket: Drift Surface Engineering

As depicted in Fig. 6(d) the onset of SCM is dominated by the carrier accumulation at the surface of drift region. An increased drift region doping near the device surface can mitigate QS, but will seriously hamper the breakdown characteristics. Fig. 15(a) depicts that the drift region surface can be intelligently engineered by placing an N-type pocket between the gate and DE, with doping higher than drift region doping. One should note that the SCM triggers in the region between gate and DE [Fig. 6(a)], before the peak electric field is shifted to DE. Such an island, placed in the region where SCM triggers, attributed to its higher background doping as shown in Fig. 15(a), mitigates the QS. Moreover, as this is placed far away from the impact ionization hot spot, it does not seriously hamper the breakdown voltage. Fig. 15(b) and (d) shows that the electric field in and around the pocket island and at DE is lowered due to mitigated charge modulation, which in fact is redistributed to CE when compared with the case without N-type pocket [Fig. 6(a)]. This allows high drain current or drift current density before the device experiences QS [Fig. 15(b)]. Fig. 14(d) depicts the guidelines to design this N-type pocket. By keeping this pocket between gate and DEs, the onset of QS can be improved by 1.2 V. It is worth highlighting that moving it close to DE or gate edge will bring the onset back to lower gate voltages. Moreover, moving it very close to gate edge can affect the breakdown voltage.

E. Deep Drain Diffusion

As discussed earlier, relaxing the injected carrier density by spreading mobile carriers deep in the drift region helps mitigating QS. Similarly, if the majority carrier density can be relaxed at DE, QS can be avoided. Fig. 16(a) shows that increasing drain diffusion depth (x_j) relaxes the carrier density at DE. This relaxes the peak electric field at DE and delays

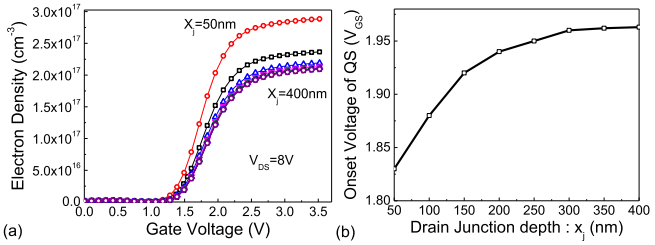

Fig. 16. Impact of drain junction depth on the QS behavior. (a) Relaxed carrier density at DE when drain junction depth was increased from 50 to 400 nm. (b) Onset voltage for QS as a function of drain junction depth.

mobility degradation, which improves the onset of QS as depicted in Fig. 16(b).

V. Conclusion

While acknowledging the criticality of QS in power devices and ambiguities related to physics of QS presented in the published literature, we have developed a unified theory with deeper physical insights to uncover the root cause of QS. It was found that carrier velocity saturation and SCM are necessary conditions for QS; however, they are not the root cause, unlike presented earlier. We found that early SCM leads to electric field peaking at the DE of the device, which causes high field mobility degradation and hinders the current conduction. On the other hand, as gate voltage is increased, channel attempts to inject higher and higher current into the drift region, which screens the electric field at the DE and mitigates mobility degradation. These competing factors were found to be the root cause of QS. In case of joule heating across the device, the onset drain current for QS falls attributed to simultaneous temperature as well as field-dependent mobility degradation. Finally, using the new insights, key device design parameters to mitigate QS have been disclosed. It was found that avoiding early SCM and electric field peaking by spreading the injected carriers deep in the drift region and near the drain diffusion region is the key to mitigate QS. This was shown to be achieved by engineering the drift region design. In general, QS can be mitigated by suppressing electric field at DE, which was achieved by engineering drain extension region or by incorporating an N-type pocket between gate and DEs.

References

[1] Y. Shi, N. Feilchenfeld, R. Phelps, M. Levy, M. Knaipp, and R. Minixhofer, "Drift design impact on quasi-saturation amp; HCI for scalable
N-LDMOS," in *Proc. IEEE 23rd Int. Symp. Power Semiconductor Devices ICs (ISPSD)*, May 2011, pp. 215–218.

[2] C.-M. Liu and J. B. Kuo, "Quasi-saturation capacitance behavior of a DMOS device," *IEEE Trans. Electron Devices*, vol. 44, no. 7, pp. 1117–1123, Jul. 1997.

[3] M. Shrivastava and H. Gossner, "A review on the ESD robustness of drain-extended MOS devices," *IEEE Trans. Electron Devices*, vol. 12, no. 4, pp. 615–625, Dec. 2012.

[4] W. Wang, B. Tudor, X. Xi, W. Liu, and F. J. Lee, "An accurate and robust compact model for high-voltage MOS IC simulation," *IEEE Trans. Electron Devices*, vol. 60, no. 2, pp. 662–669, Feb. 2013.

[5] W. El Manhawy and W. Fikry, "Power MOSFET macromodel accounting for saturation and quasi saturation effect," in *Proc. Can. Conf. Elect. Comput. Eng.*, vol. 4. May 2004, pp. 1839–1843.

[6] A. C. T. Aarts and W. J. Kloosterman, "Compact modeling of high-voltage LDMOS devices including quasi-saturation," *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 897–902, Apr. 2006.

[7] T. Lekshmi, A. Mittal, A. DasGupta, A. Chakravorty, and N. DasGupta, "Compact modeling of SOI-LDMOS including quasi-saturation effect," in *Proc. 2nd Int. Workshop Electron Devices Semiconductor Technol. (IEDST)*, Jun. 2009, pp. 1–4.

[8] C. Anghel *et al.*, "Physical modelling strategy for (quasi-) saturation effects in lateral DMOS transistor based on the concept of intrinsic drain voltage," in *Proc. Int. Semiconductor Conf.*, vol. 2. Oct. 2001, pp. 417–420.

[9] M. N. Darwish, "Study of the quasi-saturation effect in VDMOS transistors," *IEEE Trans. Electron Devices*, vol. 33, no. 11, pp. 1710–1716, Nov. 1986.

[10] C. Kreuzer, N. Krischke, and P. Nance, "Physically based description of quasi-saturation region of vertical DMOS power transistors," in *IEDM Tech. Dig.*, Dec. 1996, pp. 489–492.

[11] J. Evans and G. Amaratunga, "The behavior of very high current density power MOSFETs," *IEEE Trans. Electron Devices*, vol. 44, no. 7, pp. 1148–1153, Jul. 1997.

[12] M. Knaipp, G. Rohrer, R. Minixhofer, and E. Seebacher, "Investigations on the high current behavior of lateral diffused high-voltage transistors," *IEEE Trans. Electron Devices*, vol. 51, no. 10, pp. 1711–1720, Oct. 2004.

[13] L. Wang *et al.*, "Physical description of quasi-saturation and impact-ionization effects in high-voltage drain-extended MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, no. 3, pp. 492–498, Mar. 2009.

[14] K. H. Lou, C. M. Liu, and J. B. Kuo, "Analysis of the quasi-saturation behavior considering the drain-to-source voltage and cell-spacing effects for a vertical DMOS power transistor," *Solid-State Electron.*, vol. 36, no. 1, pp. 85–91, 1993.

[15] C. M. Liu, J. B. Kuo, and Y.-P. Wu, "An analytical quasi-saturation model considering heat flow for a DMOS device," *IEEE Trans. Electron Devices*, vol. 41, no. 6, pp. 952–958, Jun. 1994.

[16] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*. Hoboken, NJ, USA: Wiley, 2006.

[17] C.-M. Liu, K.-H. Lou, and J. B. Kuo, "77 K versus 300 K operation: The quasi-saturation behavior of a DMOS device and its fully analytical model," *IEEE Trans. Electron Devices*, vol. 40, no. 9, pp. 1636–1644, Sep. 1993.

[18] S. Xu, K. P. Gan, G. S. Samudra, Y. C. Liang, and J. K. O. Sin, "120 V interdigitated-drain LDMOS (IDLDMOS) on SOI substrate breaking power LDMOS limit," *IEEE Trans. Electron Devices*, vol. 47, no. 10, pp. 1980–1985, Oct. 2000.

[19] W. Chen, B. Zhang, and Z. Li, "Novel SJ-LDMOS on SOI with step doping surface-implanted layer," in *Proc. Int. Conf. Commun., Circuits Syst.*, Jul. 2007, pp. 1256–1259.

[20] E. J. Yoffa, "Screening of hot-carrier relaxation in highly photoexcited semiconductors," *Phys. Rev. B*, vol. 23, no. 4, p. 1909, 1981.

B. Sampath Kumar was born in Hyderabad, India. He received the M.Tech. degree from the Visvesvaraya National Institute of Technology, Nagpur, India, in 2013. He is currently pursuing the Ph.D. degree in electronic engineering with the Advanced Nanoelectronic Devices and Circuits Research Group, Indian Institute of Science, Bangalore, India.

Mayank Shrivastava (S'09–M'10–SM'16) received the Ph.D. degree from IIT Bombay, Mumbai, India.

He joined as a Faculty Member with the Indian Institute of Science, Bangalore, India, in 2013. He has over 70 international publications and 40 patents. His current research interests include experimentation, design, and modeling of beyond CMOS devices using Graphene, TMDCs, WBG materials, and ESD reliability.