Performance Analysis of 60-nm Gate-Length III-V InGaAs HEMTs: Simulations Versus Experiments

Neophytou Neophytos  
*Purdue University - Main Campus*

Rakshit Titash  
*Intel Corp.*

Mark S. Lundstrom  
*School of Electrical and Computer Engineering, Birck Nanotechnology Center, Purdue University, lundstro@purdue.edu*

Follow this and additional works at: [https://docs.lib.purdue.edu/nanopub](https://docs.lib.purdue.edu/nanopub)  
Part of the *Nanoscience and Nanotechnology Commons*

[https://docs.lib.purdue.edu/nanopub/538](https://docs.lib.purdue.edu/nanopub/538)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Performance Analysis of 60-nm Gate-Length III–V InGaAs HEMTs: Simulations Versus Experiments

Neophytos Neophytou, Titash Rakshit, and Mark S. Lundstrom, Fellow, IEEE

Abstract—An analysis of recent experimental data for high-performance In$_{0.7}$Ga$_{0.3}$As high electron mobility transistors (HEMTs) for logic applications is presented. By using a fully quantum mechanical ballistic model, we simulate In$_{0.7}$Ga$_{0.3}$As HEMTs with gate lengths of $L_G = 60, 85, 135$ nm and compare the result to the measured $I–V$ characteristics, including drain-induced barrier lowering, subthreshold swing, and threshold voltage variation with gate insulator (wide-bandgap barrier layer) thickness, as well as on-current performance. To first order, devices with three different oxide thicknesses and channel lengths can all be described by a ballistic model for the channel with appropriate values of parasitic series resistance. For high gate and drain voltages ($V_{GS} − V_T = 0.5$ V and $V_{DS} = 0.5$ V), however, the ballistic simulations consistently overestimate the measured on-current (a sign of higher transconductance), and they do not show the experimentally observed decrease in on-current with increasing gate length. With no parasitic series resistance at all, the simulated on-current of the $L_G = 60$ nm device is about twice the measured current. According to the simulation, the estimated ballistic carrier injection velocity for this device is about $2.7 \times 10^7$ cm/s. Because of the importance of the semiconductor capacitance, the simulated gate capacitance is about 2.5 times the insulator/buffer capacitance. Possible causes of the transconductance degradation observed experimentally under high gate voltages in these devices are also explored. In addition to a possible gate-voltage-dependent scattering mechanism, the limited ability of the source to supply carriers to the channel and the effect of nonparabolicity are likely to play a role. The drop in the on-current at higher gate biases with increasing gate length, is an indication that the devices operate below the ballistic limit.

Index Terms—Apparent, ballistic, high electron mobility transistor (HEMT), InAs, InGaAs, mobility, nonequilibrium Green’s function (NEGF), nonparabolicity, quantum, series resistance, source exhaustion, source starvation, III–V.

I. INTRODUCTION

FIELDEFFECT transistors (FETs) with III–V channel materials have recently received much attention because of their potential as switching devices for future digital technology nodes. Traditionally, III–V high-electron-mobility-transistor (HEMT)-based devices (mostly based on high In concentration InGaAs channels) have been used in high-frequency/low-noise applications because of their high $f_T$ [1]–[7], which is also shown to increase as the devices are scaled [8]–[10]. Recently, however, both III–V heterostructure-based HEMTs [11]–[14] and III–V-based MOSFETs [15], [16] have been reported targeting future logic-based applications. Due to their higher mobility, the III–V channel materials should reach the ballistic limit at longer channel lengths than Si devices. The low effective mass of the III–V materials should also boost the ballistic carrier velocity and improve the $I_D$–$V_D$ characteristics. Tradeoffs are involved (e.g., the light effective mass leads to a density-of-states bottleneck [17], [18] and to source–drain tunneling [19]), but III–V FETs have the potential to outperform Si MOSFETs under low-voltage operation. In that regard, high-performance HEMTs based on III–V compounds with channel lengths below 90 nm have recently been demonstrated [12], [14], [20], [21]. Good control of the wide-bandgap barrier thickness down to 3 nm was achieved while still maintaining relatively low gate leakage currents—even under high biases. This paper is a simulation study of the results reported by Kim and del Alamo [12]. Our objective is to examine the experimental data with a fully quantum mechanical ballistic model in order to understand what can and cannot be explained.

In this paper, a 2-D ballistic quantum transport HEMT simulator based on the real-space nonequilibrium Green’s function (NEGF) approach [22], [23], coupled self-consistently with a 2-D Poisson solver, is employed. Simulation results show that for these In$_{0.7}$Ga$_{0.3}$As HEMTs with a gate length of 60 nm and zero series resistance, a ballistic device of this kind would deliver about twice the on-current of the measured device. With external series resistors added, the simulated $I–V$ characteristics are close to the measured results, except at the highest gate voltage. While the discrepancy at high gate voltages might be due to scattering, source design and conduction band nonparabolicity are equally likely explanations. The ballistic simulations show good agreement with the subthreshold swing (SS) and drain-induced barrier lowering (DIBL) versus channel length, but they do not show the drop in on-current with increasing gate length that is observed experimentally. This suggests that scattering is important in the longer channel length devices. Although the drop in on-current with increasing gate length is an indication that the devices operate below the ballistic limit at higher gate lengths, the saturation current for the 60-nm gate-length device is close to what is predicted by the ballistic simulations. These results provide guidance toward a light effective mass short channel III–V HEMT device design for logic. The conclusions drawn on the near-ballistic device...
from the channel layer, effectively dopes the source/drain regions (noted observations and measurements. and these results help provide understanding of experimental evident in both the carrier velocity and the inversion charge, itance seem to be important in these devices, and they are mechanical effects (tunneling, reflection), and quantum capac-

operation at short channel lengths and the importance of source design indicate that the focus of future experiments should be on the optimization of extrinsic part of the device. Quantum mechanical effects (tunneling, reflection), and quantum capacitance seem to be important in these devices, and they are evident in both the carrier velocity and the inversion charge, and these results help provide understanding of experimental observations and measurements.

II. APPROACH

For simulation purposes, the device geometry was simplified as shown in Fig. 1. In the experimental device [12], the source and drain contacts are located on the top of the device, and the current flow is 2-D through a doped heterostructure stack. Rather than attempting to simulate the contacts (and the associated metal–semiconductor contact resistance), we placed ideal contacts at the two ends of the channel, as shown in Fig. 1, and added extrinsic series resistors to the source and drain. The simulated HEMT consists of a 15-nm In$_{0.52}$Ga$_{0.48}$As layer between two In$_{0.52}$Al$_{0.48}$As buffer layers. The gate electrode in the simulated device is placed on top of the In$_{0.52}$Al$_{0.48}$As layer, which (in the simulated structure) has the same thickness throughout the entire length of the device. A silicon δ-doped layer in the In$_{0.52}$Al$_{0.48}$As buffer layer effectively dopes the source/drain regions of the device that extend from the gate edge to the edge of the ohmic contact (denoted as $L_{\text{side}}$ in Fig. 1, located below the n$^+$ region) to 2.1 $\times$ 10$^{12}$/cm$^2$ [24]. The δ-doped layer is located 3 nm away from the channel layer. Devices with barrier thicknesses of $t_{\text{ins}}$ = 3, 7, and 11 nm were reported in [12]. Later, better estimates of the 7- and 11-nm devices were given as 6.5 and 10 nm [24], and these were the values used in our simulations. When simulating the $t_{\text{ins}}$ = 3 nm device, the δ-doped layer was placed on top of the barrier, and the gate electrode on top of a thin layer residing on top of the δ-doped layer. In the simulation, the δ-doped layer is given a finite thickness of 0.40 nm, and the thin layer on top of the δ-doped layer was 0.40 nm. The result was a barrier thickness of $t_{\text{ins}}$ = 3.8 nm in the simulation. This is within the experimental uncertainty in the barrier thicknesses of ±1 nm [24]. The uncertainty in barrier thickness is not substantial for the thinnest barrier device, and this is discussed in Section IV.

In the simulations, the far left/right regions of the δ-doped layer (denoted as n$^{++}$ in Fig. 1) are doped to 1 $\times$ 10$^{13}$/cm$^2$ to mimic additional doping from the n$^{++}$ stack of cap layers used in the experimental device to facilitate ohmic contacts to the source and drain. There are, therefore, two different doping regions in the simulated device. The region that is directly adjacent to the channel to its left/right ($L_{\text{side}}$) extends from the gate edge to the ohmic edge and has a carrier density of 2.1 $\times$ 10$^{12}$/cm$^2$, which is the value specified by the experimental group [24]. The far left/right region has a larger doping of 10$^{13}$/cm$^2$. This level of doping is unrealistically high for this type of material, but it favors numerical stability of the simulation and does not affect the device. This is a way to mimic the extended source/drain regions of the actual device. In addition, although the lightly doped region in the experiment has $L_{\text{side}}$ = 1 $\mu$m [24], for computational efficiency, $L_{\text{side}}$ is set to $L_{\text{side}} = 60$ nm in the simulations (its length is irrelevant in the case of ballistic simulations). In more recent experimental devices, the length of this region is greatly reduced [14], [20]. As will be discussed later, under high gate bias, the source design becomes important. Because of the simplified source design used in the simulation, we will be able to draw only qualitative conclusions about the high gate bias performance. The channel region is the region directly under the gate electrode and has $L_G = 60$ nm as in the experimental device. Longer channel lengths were also examined experimentally, and these devices are briefly considered in Section IV.

It should be pointed out that the source and drain contacts in the simulated device should not be regarded as real contacts with an associated contact resistance. Rather, they are idealized contacts to the extended source/drain regions, which are assumed to be maintained in thermodynamic equilibrium by strong scattering. Venugopal et al. [25] examined this assumption for silicon transistors and found that scattering in typical contacts of heavily doped silicon is sufficient to maintain thermodynamic equilibrium. Nevertheless, this assumption may
need to be reconsidered as devices continue to shrink and for new channel materials such as the III–V materials considered here. Indeed, Fischetti et al. have discussed the phenomenon of “source starvation” which is a manifestation of nonequilibrium contacts in III–V FETs [26]. For this study, we assume extended contacts that are maintained in thermodynamic equilibrium.

Fig. 1(b) shows the simulated conduction band profile normal to the channel taken at a location near the source end of the channel (near the top of the potential energy barrier between the source and the channel) when the device is under large gate bias. The workfunction difference between the gate and the In$_{0.52}$Al$_{0.48}$As buffer layer is adjusted to $\Delta \Phi_B = 0.5$ eV in order to match the threshold voltage of the simulated devices to the experimental measurements. The thickness of the In$_{0.52}$Al$_{0.48}$As layer in this case is 3 nm, and the $\delta$-doped layer is adjacent to the gate/In$_{0.52}$Al$_{0.48}$As interface. The conduction band discontinuity between the In$_{0.7}$Ga$_{0.3}$As/In$_{0.52}$Al$_{0.48}$As layers is assumed to be $\Delta E_c = 0.6$ eV [24], [27]. The dielectric constant of In$_{0.52}$Al$_{0.48}$As is assumed to be $\varepsilon = 14$, and that of the In$_{0.7}$Ga$_{0.3}$As is $\varepsilon = 14.5$ [28]. The dotted line shows the charge distribution profile under high-gate-bias operating conditions ($V_G = 0.5$ V). Noticeable here is that even under high biases, the amount of charge penetration into the buffer layer is negligible; therefore, at first order, the barrier plays the role of the insulator gate oxide of a traditional MOSFET. Connection to “MOSFETs” and “insulators” further on in this paper is therefore safely made so that the reader can relate this device to the MOSFET operation for logic devices. Another point to make is that the barrier (“insulator”) thickness is 3.8 nm. The $\delta$-doped layer is very near the gate electrode, and its effect is minimal. When the gate is further away from the $\delta$-doped layer, a local energy minimum forms, right at the position of the $\delta$-doped layer—the donor doping pushes the bands down. This is not evident when the doping is adjacent to the gate electrode [Fig. 1(b)].

The effective mass of the In$_{0.7}$Ga$_{0.3}$As channel is an input to the simulation. Because of conduction band nonparabolicity, quantum confinement increases the effective mass as compared to its value in the bulk [29]. In principle, the appropriate effective mass could be extracted from atomistic calculations (e.g., tight-binding atomistic methods); however, this is difficult because the masses depend on the exact placement of the atoms in the structure and the distortions within the structure. In this paper, a simplified approach is followed; we extract the effective mass from atomistic tight-binding [30] calculations for a 15-nm-wide In$_{0.7}$Ga$_{0.3}$As quantum well structure without assuming any lattice distortions. The dispersion of the quantum well is shown in Fig. 2. The wafer orientation is (100), and the transport orientation is [011]. The parabolic band (red-dotted) of $m^* = 0.048 m_0$ is adjusted to match the density of states up to 0.2 eV above the conduction band edge.

Fig. 2. Dispersion of the composite 15-nm-thick In$_{0.7}$Ga$_{0.3}$As structure calculated using atomistic tight-binding calculations with no distortions taken into account. The wafer orientation is (100), and the transport orientation is [011]. The parabolic band (red-dotted) of $m^* = 0.048 m_0$ is adjusted to match the density of states up to 0.2 eV above the conduction band edge.

There are various values reported in the literature for the $\Gamma$ to $L$ valley separation of InAs varying from 0.73 [31], [32] to 1.16 eV [31], [33]–[35]. This uncertainty will affect the In-rich In$_{0.7}$Ga$_{0.3}$As compounds too and will translate to variations from ~0.6 to ~0.93 eV (Fig. 2). Our calculations show the value of 0.93 eV since our TB parameters are calibrated to the data reported in [33] (see [36]). As we show later on, however, it turns out that only the first ~0.2 eV above the conduction band edge ($\Gamma$ valley) is important in transport for this type of logic devices operating at a $V_{cc}$ of only 0.5 V. Since, in all estimations in this In-rich compound, the $L$ valley is much higher than 0.2 eV from the $\Gamma$ valley, it does not affect the device operation and it can safely be ignored.

The NEGF approach [22], [23] for ballistic quantum transport is self-consistently coupled to a 2-D Poisson solver for treatment of the electrostatics. Since the channel is relatively thick (15 nm), potential variations are expected in the cross section along the transport orientations. The NEGF Hamiltonian uses a real-space technique in the parabolic effective mass approximation (EMA) and accurately accounts for the mode coupling when large potential variations exist. The NEGF transport equation is solved in the channel area as well as in the upper In$_{0.52}$Al$_{0.48}$As buffer layer in order to capture the wavefunction penetration in that layer. The 2-D Poisson’s
equation is solved in the entire cross section of the device in order to accurately capture the 2-D electrostatics of the device.

III. RESULTS

In order to compare the measured to the simulated data, two fitting parameters were used: 1) a value of external series resistance \( R_{SD} \) that is added to the device and 2) the workfunction difference \( \Delta \Phi_B \) between the gate and the In0.52Al0.48As buffer layer. The effect of the series resistance will be discussed in the following section. The workfunction difference \( \Delta \Phi_B \) is used to adjust the \( V_T \) of the simulated to that of the experimental data. The adjustment is done once for the \( L_G = 60 \) nm device with a 3-nm-thick barrier. The result, \( \Delta \Phi_B = 0.5 \) eV, is a reasonable number for the workfunction difference between the two materials. Fig. 3(a) shows the \( I_D-V_G \) characteristics for the three devices—each with a 60-nm channel length but with three different barrier thicknesses. The measured and simulated curves agree fairly well. In the case of the 7-nm barrier device, the simulated and measured \( V_T \) differs by \( \sim 0.04 \) V. This small deviation might be due to various reasons such as interface traps, charged impurities, or uncertainties in the thickness of the layers in the experimental device. As the barrier thickness increases, there is a large negative shift in the \( V_T \) by almost 0.25 V, which is attributed to the \( \delta \)-doped layer and its increasing effect on the electrostatics of the channel as the gate electrode moves farther away. The threshold voltage shift is well described by

\[
\Delta V_T = \frac{qN_W}{C_{\text{ins}}} \frac{\hat{x}}{T_{\text{ins}}} \tag{1}
\]

where \( N_W \) is the \( \delta \)-doping concentration per square centimeter and \( \hat{x} \) is the centroid of the charge distribution in the barrier [37]. In this case, \( \hat{x} \) is defined as the distance of the donor \( \delta \)-doped layer from the gate electrode. The closer it is to the gate electrode, the lesser its effect on the \( V_T \) shift.

The DIBL and the SS extracted from the simulated data are shown in Fig. 3(b), and both are seen to increase as the barrier thickness increases, which is expected from 2-D electrostatics. The simulated results agree with the experimental data both qualitatively and quantitatively.

The second adjustable parameter in the simulation is the series resistance. Since the ohmic contacts are not included in our simulation domain, we added a fixed external series resistance \( R_{SD} \) to the ballistic simulated results. The value of that resistance is such that the total resistance of the device is matched (by fitting the slope of the linear region of the \( I_D-V_D \) characteristics at high gate bias for both simulated and experimental data). This value is somewhat lower than the experimentally reported value for \( R_{SD} \). To understand this discrepancy, it is helpful to consider the various components of the experimentally measured \( R_{SD} \). In experiments, the series resistance consists of the following [20]: 1) contact resistance from S/D metal to the cap layer; 2) negligible resistance in the heavily \( n^+ \) doped cap layer; 3) tunneling resistance from \( n^+ \) cap layer into the channel layer through the tunnel barrier layer; (4) a quantum mechanical resistance coming from the price that electrons pay when injected from a bulk into quantized modes of the channel [12]; and (5) resistance in the un gated \( (R_{\text{side}}) \) part of the channel/contact. It is nontrivial to separate out these components or have a one-to-one mapping for both measurements and simulations. Our simulations capture resistance (4) and to some extent parts of (5) (by capturing the number of occupied modes). We note that the doping in the barrier layer close to the gate modulates both (4) and (5) and is a variable in the experiments studied [38]. In the simulations, this is modulated depending on the number of occupied modes, which varies with doping concentration. The rest of the resistances as described earlier are lumped into a bulk external resistance and added to the \( I_D-V_D \) at the end of the ballistic simulations. Therefore, similar but somewhat higher values for the series resistance were measured for the experimental devices [12], [38].

Fig. 4(a) shows the experimental \( I_D-V_D \) data and the simulated ballistic \( I_D-V_D \) characteristic at the same gate overdrive \( (V_G - V_T = 0.5 \) V). The simulated ballistic on-current (measured at \( V_G = V_D = 0.5 \) V) is almost double than the experimental value, and the channel resistance of the simulated ballistic device is \( R_B = 170 \) \( \Omega \cdot \mu \text{m} \) (inverse slope of the linear region). We note here that the “ballistic resistance \( R_B \)” is inherently built into the quantum ballistic model and represents the quantum mechanical resistance of the conducting modes.

It is a quantity that depends on how many modes are occupied, and not on the length of the channel or the source/drain regions. In order to fit the simulated results to the experimental data, a series resistance (source plus drain) of \( R_{SD} = 400 \) \( \Omega \cdot \mu \text{m} \) was added to the ballistic data in order to match the total resistance measured in the experimental data (inverse slope of the high \( V_G \) experimental \( I_D-V_D \)).
R I NEOPHYTOU overdrive. (a) The barrier thicknesses [12]. As more of the barrier sidewall is respectively). The value of the fitted series resistance increases with the highest gate voltage data, the simulated data at low drain volt-

(b) The device. Data for *V* = 0.1, 0.3, and 0.5 V are shown. The black-dashed curve indicates the ballistic *I*–*V* at *V* = 0.5 V with *R* = 0 Ω · μm. A *R* = 400 Ω · μm is added to the simulated data. (b) The *t* = 7 nm device. Data for *V* = 0 V, 0.2 V and 0.4 V are shown. An *R* = 350 Ω · μm is added to the simulated data. (c) The *t* = 11 nm device. Data for *V* = 0.1 V, 0.3, and 0.3 V are shown. An *R* = 310 Ω · μm is added to the simulated data.

Once the series resistance is fit to the linear region of the highest gate voltage data, the simulated data at low drain volt-

ages show very good agreement with the experimental observations for all three gate bias cases reported experimentally (*V* = 0.1, 0.3, and 0.5 V). The agreement at high drain voltage is also good, except for a ~20% discrepancy between on-currents of the measured and simulated data. For this *L* = 60 nm device, the experimental results can, to a reasonable approximation, be explained by an intrinsic ballistic FET with two series resistors attached to it, except for the overestimate of the on-current, which will be discussed in Section IV. Longer channel lengths appear to operate at a lower fraction of the ballistic limit, as will also be discussed in Section IV.

Similarly, the experimental *I*–*V* data for the *t* = 7 nm and *t* = 11 nm devices can be explained by using slightly different values of *R* (in the case of *R* = 350 Ω · μm and *R* = 310 Ω · μm, respectively). The value of the fitted series resistance increases as the barrier thickness decreases. This was also observed in the experiments and was attributed to the isotropic etching that was used before gate deposition to produce the three different barrier thicknesses [12]. As more of the barrier sidewall is etched, the series resistance tends to increase. Fig. 4(b) and (c) shows the experimental *I*–*V* for various *V* values compared to the simulated results after the series resistance has been fit. Good agreement between the experimental and simulated data is observed, but for each of the three cases, the on-current of the simulated device is ~10%–20% more than that of the measured device. The fact that these devices operate close to the ballistic limit has implications in device optimization, shifting the focus from the intrinsic channel to the extrinsic part of the HEMT structure.

The mobility of a FET is often extracted from the linear region current. Although mobility has no physical meaning in our ballistic simulations, the simulated ballistic drain current is linearly proportional to the drain voltage at low *V*, so we can extract a “mobility” by equating the channel resistance to a conventional MOSFET expression

\[
R_{ch} = \frac{V_{DS}}{I_{DS}/W} \approx \frac{L}{\mu B C_{ins}(V_G - V_T)} \tag{2}
\]

where *B* is the so-called ballistic mobility by the studies in [39]–[41]. From our simulations, *R* (same as *R*) at high gate bias (before adding the effect of *R*) varies between *R* = 170 Ω · μm and *R* = 240 Ω · μm as the barrier thickness varies from 3 to 11 nm. The increase in this resistance as the insulator thickness increases is due to the fact that there are fewer occupied modes in the channel at ON-state for a thicker insulator device. From these channel resistances, a value of the ballistic mobility is extracted to be *B* ~170–450 cm²/V · s. Although the mobility of bulk InGaAs is measured to be ~10 000 cm²/V · s, the “apparent” mobility [in the sense of (2)] that a short channel HEMT can display is limited to a few hundreds. Alternatively, one could deduce a mobility for the device by plotting the total resistance between the source and drain as a function of channel length. The y-intercept of this curve would be the fixed external series resistance, and the inverse of the slope would be proportional to the channel mobility. In that case, a ballistic FET would show zero slope, corresponding to an infinite mobility, meaning that the resistance of the device is as expected independent of the length of the ballistic channel (see the slopes of the simulation results in Fig. 11). Therefore, the mobility, a quantity directly related to scattering, loses relevance when the device operates close to the ballistic limit. The carrier velocity, however, as we show later, is the relevant feature of channel optimization.

IV. DISCUSSION

Within the uncertainties of the simplified structure used in the simulations and in our knowledge of various device parameters, the results presented in the previous section show that the *L* = 60 nm HEMTs reported by Kim et al. [12] can be approximately described as FETs with ballistic channel regions connected to two external series resistors. The only significant discrepancy between the simulated and experimental results is the consistent 10%–20% overestimate of the on-currents. The experimental transconductance *g* vs gate voltage characteristic is shown in Fig. 5 for the *t* = 3 nm device. The observed degradation in *g* at high gate voltages might be
attributed to various causes. Scattering at high gate biases could degrade $g_m$. Another possibility is population of heavy effective mass upper valleys. Fig. 2 shows, however, that the $L$ valleys are too high in energy to be populated. Parallel conduction in the upper layer, which has much heavier masses (even up to $\sim$5 times heavier) than the channel layer, could also be a possibility. As shown in Fig. 1(b), however, our simulations show no significant wavefunction penetration in the upper layer—even under high inversion conditions ($V_G - V_T = 0.5$ V). Series resistance could be yet another possibility. Fig. 5 shows the simulated $g_m$ versus $V_G$ characteristics for three different values of series resistance ($R_{SD} = 0$, 400, and 800 $\Omega \cdot \mu m$). For the $R_{SD} = 0$ and 400 $\Omega \cdot \mu m$ cases, $g_m$ follows the experimental curve but saturates at higher $V_G$ than the experimental curve. For the 800-$\Omega \cdot \mu m$ characteristic, we obtain roughly the correct magnitude of $g_m$, but this value of $R_{SD}$ is too large to be consistent with the experimental measurement. The fact that $g_m$ degradation occurs even in the ballistic simulation tells us, however, that there might be other possibilities. Two other plausible causes, the design of the source, and the effects of nonparabolicity are discussed hereinafter.

For III–V transistors, the design of the source can be an important factor [26], [42]. Transistors operate by modulating potential energy barriers [43], [44]. As the gate voltage increases, the potential energy barrier decreases, and the charge in the channel increases. When the gate voltage increases to the point where the barrier is removed and the channel charge is equal to the charge in the source, transistor action degrades significantly. For the HEMT under consideration here, the charge in the source ($2.1 \times 10^{12}$/cm$^2$) is much lower than typical for Si MOSFETs, so these source exhaustion effects become apparent at relatively low gate voltages.

Source design considerations are illustrated by the ballistic simulation shown in Fig. 6. Fig. 6(a)–(c) shows the energy-resolved current versus position for the HEMT under different gate voltages. The conduction and valence bands are indicated (white-dot lines), and the current flows above the top of the conduction band. (We note that the white-dotted lines are estimates of the conduction and valence bands and are plotted for illustration purposes using the electrostatic potential as a reference. In principle, they can be slightly shifted in energy due to potential variations in the cross section of the channel.) The source/drain regions consist of two portions, an $n^+$ region near the ideal contacts and an $n^+$ region adjacent to the channel. Fig. 6(a) shows the OFF-state of the device, where the source Fermi level ($E_{FS}$) is well below the top of the source to channel energy barrier. As $V_G$ increases, the barrier in the channel decreases—eventually reaching the same level as the $n^+$ source region [Fig. 6(b)]. The top of the barrier has, in this case, shifted to the beginning of the $n^{++}$ source region. The $L_{side}$ region below the $n^+$ layer cannot provide any additional carriers to the channel since, now, their concentrations are of similar value (the bands in the $L_{side}$ region and the top of the barrier are at the same level below $E_{fs}$). When $V_G$ increases even more [Fig. 6(c)], the gate can only modulate the energy barrier at the $n^{++}$ to $n^+$ junction through weak fringing fields. Transistor action is lost, and $g_m$ drops as shown in Fig. 5 for both the simulated and measured characteristics. In our simulations, these effects are exaggerated by the assumption of ballistic transport in the $n^+$ source, but the effect is primarily an electrostatic one and is also observed in drift–diffusion simulations [45].

The gate voltage at which the transconductance begins to degrade is strongly dependent on the barrier between the channel and the source, which depends on the doping of the source. Fig. 7 shows the simulated $g_m$ for structures with different $\delta$-doping densities above the source/drain. As the doping in the source decreases, this effect shows up at smaller gate voltages. The low gate bias part of the $g_m$ versus $V_G$ characteristic is not
doping dependent because under low gate voltage, the source is able to supply the charge demanded by the gate voltage. In the experimental results, the n$^+$ source region was $L_{\text{side}} = 1 \mu\text{m}$ in length, whereas in our simulation, $L_{\text{side}} = 60$ nm was used. The differences/uncertainties in the source doping profiles and design may explain why the transconductance is experimentally observed to degrade $\sim 0.1$ V before the simulated transconductance. Although we cannot unambiguously conclude that the observed transconductance degradation is due to source exhaustion, our simulations clearly demonstrate that source design is an important issue for III–V HEMTs. Higher doping in the barrier layers would increase the maximum $g_m$. Finally, note that the effects discussed here are purely electrostatic in nature and occur in both ballistic and drift–diffusion simulations. Fischetti et al. [26] have discussed “source starvation,” which results from a difficulty in injecting carriers into longitudinal momentum states in the channel. Those effects were not included in our study and would only make source design an even more important issue.

Two important parameters for a FET are the charge and velocity at the beginning of the channel. Two questions arise. The first is as follows: How close is the charge at the top of the potential barrier to the equilibrium MOS capacitor value of $Q = C_G(V_D - V_T)$? The second question is as follows: How the velocity extracted from the numerical simulator compares to the ballistic injection velocity expected from the bandstructure of the channel? To answer both of these questions, the top of the potential barrier in the numerical results needs to be identified. Doing so is not as trivial because of the large variation of the $E_C$ across the depth of the 15-nm channel width. We employ two different methods to locate the top of the barrier. The first is to take the weighted average of the charge distribution with the 2-D $E_C(x,y)$ profile with the 2-D charge density $n(x,y)$ according to

$$\langle E_C(x) \rangle = \frac{\int n(x,y)E_C(x,y)dy}{\int n(x,y)dy}. \quad (3)$$

Fig. 8(a) shows the resulting $\langle E_C(x) \rangle$ (white-dotted line) superimposed on the electron density spectrum plot. Fig. 8(a) is plotted at $V_G = 0.4$ V and $V_D = 0.35$ V, which are the estimated intrinsic device voltages at the ON-state (after accounting for the effect of $R_{SD}$). From Fig. 8(a), the top-of-the-barrier can be identified to reside at 105 nm (5 nm inside the channel from the point where the gate electrode begins).

A second way to identify the top-of-the-barrier is by identifying the point of maximum gate control by locating the position where $dN_S(x)/dV_G$ is maximized (where $N_S(x)$ is the charge in the channel per square centimeter). This method places the top-of-the-barrier at 104 nm. Both approaches give very similar results, so we take the top-of-the-barrier to be at 104.5 nm. The corresponding charge and the velocity [defined as $I_{ON}/N_S(x)$] at the top-of-the-barrier are $N_S \approx 1.3 \times 10^{12}$ per cm$^2$ and $v_{\text{ave}} \approx 2.7 \times 10^5$ cm/s as shown in Fig. 8(b) and (c), respectively. The charge density and velocity are rather low for this light mass channel due to the fact that the source Fermi level is less than 0.1 eV above $E_C$ under ON-state conditions. Fig. 8 shows that these quantities are very sensitive to the precise location of the beginning of the channel. (Within a few nm away from that point, very different values can be extracted.) This information is available in our simulator, but it is not available when analyzing experimental data.

To answer the first question about how close the charge is to $Q = C_G(V_D - V_T)$, the simulated equilibrium carrier density versus gate voltage is shown in Fig. 8(d) (solid-blue) (this is read from the simulator results in the center of the channel at $V_{DS} = 0$ V conditions). The quantity $Q = C_{\text{ins}}(V_D - V_T)$ with $C_{\text{ins}} = 0.032$ F/m$^2$ (or $3.2 \times 10^{-6}$ F/cm$^2$) is shown as the solid-square-black line in Fig. 8(d). Assuming that $C_G = C_{\text{ins}}$ clearly overestimates the charge. From the slope of the $C_G$ versus $V_D$ plot (dashed-red line), we observe that $C_G$ is 2.5 times smaller than $C_{\text{ins}}$. From $C_G = C_{\text{ins}}C_S/(C_{\text{ins}} + C_S)$, we obtain a semiconductor capacitance of $C_S = 0.67C_{\text{ins}}$. A simple calculation of the quantum capacitance (defined as $C_Q = \partial Q/\partial E_f$ and being a measure of the density of states at the Fermi level), however, shows that $C_Q \sim 1.5C_{\text{ins}}$, which indicates that $C_S$ is a factor of $\sim 2$ less than $C_Q$. As discussed by Pal et al. [46], this occurs when the shape of the quantum well is bias dependent.

According to Fig. 8(d), at $V_{GS} = 0.4$ V, the charge at the top of the barrier under equilibrium conditions is $N_S \approx 1.5 \times 10^{12}$ per cm$^2$. The value found from the simulation under $V_{DS} = 0.35$ V is $N_S \approx 1.3 \times 10^{12}$ per cm$^2$, which is lower than the equilibrium value. At high $V_D$, it might be expected that DIBL would reduce $V_T$ and therefore increase the charge. Part of the reason for the lower charge under drain bias could be that only the positive velocity states are occupied at high $V_D$, whereas the negative going states, filled by the drain Fermi level, are now emptied. The quantum capacitance, therefore, being a measure of the occupied density of states at the Fermi level, decreases under large drain bias by a factor of two. The lower $C_Q$ lowers the semiconductor capacitance $C_S$, and as a consequence, the charge in the device offsets the DIBL. The result of these two counteracting mechanisms is that the charge at the top of the barrier under high $V_{DS}$ is somewhat less than its value at equilibrium.

The second question had to do with the value of the ballistic velocity from the numerical simulation as compared to the value expected from the bandstructure. For a given $E(k)$ and Fermi level, we can determine the corresponding $N_S$ and $(v) = v_{\text{ave}}$ under ON-state conditions where only $+k$ states are occupied. Fig. 9 shows the result for the parabolic effective mass (EMA) dispersion used in the quantum simulations (square-blue). For comparison, the InAs and GaAs velocities are shown,
calculated using dispersions extracted from an atomistic tight-binding model [30]. The weighted average of these two results is also shown in Fig. 9 (solid-brown). The weighted average bandstructure velocity for the dispersion used in the quantum simulation is shown in solid-circle-blue. The weighted average of the InAs and GaAs \((\text{In}_{0.7}\text{Ga}_{0.3}\text{As})\) velocity is shown in solid-brown.

Fig. 8. Intrinsic device parameters at ON-state. (a) The electron density spectrum. The density weighted \(E_C\) and \(E_V\) profiles are shown (dot-white lines). The top of the barrier is identified at 104.5 nm. (b) The charge density along the length of the channel. (c) The average velocity along the length of the channel. (d) The equilibrium \((V_D = 0 \text{ V})\) carrier density versus \(V_G\) (solid-blue) in the middle of the channel. The charge as \(C_{\text{ins}}(V_G - V_T)\) is shown in solid-square-black. The charge as \((C_{\text{ins}}/2.5)(V_G - V_T)\) is shown in dot-red.

Fig. 9. “Positive going” average bandstructure velocity versus inversion carrier density of a 15-nm-thick quantum well, using a simple semiclassical ballistic model. The velocities of InAs and GaAs are shown in solid-square-black. Their bandstructures are calculated using an atomistic tight-binding model. The EMA bandstructure velocity for the dispersion used in the quantum simulation is shown in solid-circle-blue. The weighted average of the InAs and GaAs \((\text{In}_{0.7}\text{Ga}_{0.3}\text{As})\) velocity is shown in solid-brown.

have to do with tunneling currents and quantum mechanical reflections around the top-of-the-barrier, which tend to reduce the average velocity. In support of this conjecture, we note that the Fermi level in the quantum model is almost a \(k_B T\) closer to \(E_C\) than in the semiclassical model at the same carrier density, which indicates a carrier population below the top-of-the-barrier and/or “negative” going state population in the quantum model. The standing wave pattern at the vicinity of the top-of-the-barrier is a sign of this effect. It is also evident in Fig. 9 that nonparabolicity can be important at this bias regime and can cause about 10% degradation in the average carrier velocity. Nonparabolicity is another possible contribution to the \(g_m\) degradation observed in the experimental data but not captured in the EMA treatment.

Our discussion so far has focused on the \(\ell_{\text{ins}} = 3 \text{ nm}\) and \(L_G = 60 \text{ nm}\) device. Experimental data are available for various InAlAs barrier thicknesses as well as gate lengths. These two issues are briefly discussed here. Fig. 10 shows how the...
barrier thickness affects the performance of the $L_G = 60$ nm device. The equilibrium carrier density in the channel is shown in solid-square-blue, extracted as in Fig. 8(d) for all devices at the same $V_G - V_T = 0.2$ V. The carrier density in the channel doubles as the barrier thickness decreases from $t_{ins} = 10$ nm to $t_{ins} = 3$ nm—as expected. Under a high drain bias of $V_D = 0.35$ V, however, the carrier density at the top of the barrier (dash/dot-diamond-black) shows a much slower variation with barrier thickness. This occurs because under high drain biases, $C_Q$ decreases by a factor of $\sim 2$, which drives the device toward the quantum capacitance limit in which variations in $C_{OX}$ are not as significant. Increasing DIBL with increasing barrier thickness lowers the $V_T$ and increases the charge in the channel. An interplay between these two effects reduces the charge variations as a function of $t_{ins}$. The increase in charge as $t_{ins}$ is scaled from 10 to 3 nm is only $\sim 30\%$. The velocity at the top of the barrier (dash-circle-red) shows an increase of $\sim 20\%$ with barrier thickness scaling. Scaling the barrier thickness down to 3 nm can therefore improve the performance. Further scaling of the barrier, however, increases the importance of $C_Q$ and might not offer additional advantage at the ON-state. Scaling, for example, the barrier from $t_{ins} = 3.8$ nm to $t_{ins} = 3.0$ nm (a $\sim 20\%$ increase in $C_{OX}$) increases the carrier velocity and charge density only slightly by 2\%–3\%, resulting only in a $\sim 5\%$ increase in the on-current.

We next investigate the gate length dependence of the $t_{ins} = 3$ nm HEMTs. Experimentally in [12], $L_G = 60$, 85, and 135 nm devices were reported. Significant gate length dependence was observed experimentally, with the on-current decreasing as the gate length increases. This trend is shown in Fig. 11. This figure is the same as in Fig. 4(a), with all the three gate length data included (for clarity, we have shown only the highest gate voltage $V_G = 0.5$ V in each case). The solid-circle-red lines present the experimental data for the different gate lengths. The solid-blue lines present the simulated results for the same devices after the series resistance was included. Other than changing the gate length of the device, no further adjustments have been made to any parameters ($V_T, V_G$, etc.) of the devices for this comparison. Although it is not shown in the plots, a fairly good match was observed between the simulated and measured data for lower gate biases. For the high gate bias case, the simulated results show little gate length dependence—as it is expected from a ballistic model. The small differences originate from the changes in the electrostatics. The measured high $V_G$ data, however, show a significant gate length dependence. The longest device is about 40\% below the ballistic simulation, while the shortest device is only $\sim 20\%$ below. These results indicate increased scattering in the $L_G = 85$ nm device and even stronger scattering in the $L_G = 135$ nm device.

Finally, we should mention once again some of the uncertainties and simplifications that affect our analysis. The first is the $\pm 1$-nm uncertainty in the etched AlInAs layer thickness, which, however, does not introduce considerable uncertainty at the ON-state. Second, the simplified device structure for the simulation had the source/drain regions that were only 60 nm long rather than 1 $\mu$m as in the experimental device, as well as the idealized assumed constants with a fixed value of $R_{SD}$. This simplification is likely to affect the high current region, where source design issues are expected to become important. Lattice distortions and the effect of strain in the channel were not considered and may have an impact on the effective mass of the channel.

V. Conclusion

The performance of recently demonstrated high-performance In$_{0.7}$Ga$_{0.3}$As HEMTs was investigated using a quantum ballistic model self-consistently coupled to a 2-D Poisson solver for electrostatics. With the addition of external series resistors consistent with experimental observations, reasonable agreement between the ballistic simulation and the experimental data was obtained for all of the 60-nm channel-length devices with barrier thicknesses of 3, 7, and 11 nm. Despite the assumptions in the model and the uncertainties in the exact values of the barrier thickness, series resistance, idealized contacts, and channel effective masses, these results suggest that 60-nm channel-length III–V HEMTs operate rather close to the ballistic limit. The on-current performance of longer channel length HEMTs, however, appears to be degraded by scattering, although they still operate at over one-half of the ballistic limit.

For operation near the ballistic limit, the ballistic injection velocity rather than bulk mobility becomes the parameter of interest. The ballistic injection velocity for this device was found to be relatively low for this light effective mass material, because of the relatively low inversion charge operating conditions, quantum tunneling and reflections, and conduction band nonparabolicity. The semiconductor capacitance also plays an important role by increasing the effective oxide thickness of the thinnest barrier device by 2.5 times. The results reported here strongly suggest that source design is an important factor for III–V FETs, as has also been recently pointed out by Fischetti et al. [26]. Increasing the doping in the barrier layer should improve the transconductance degradation. These simulations also identify key factors for improving the III–V HEMT performance as reduction of the parasitic series resistance, optimization of the source design, and reduction of the barrier thickness, which will be beneficial to the ON-state performance but have only a small effect on the OFF-state performance.
ACKNOWLEDGMENT

The authors would like to thank Dr. D.-H. Kim, Prof. J. del Alamo, and Prof. D. Antoniadis of the Massachusetts Institute of Technology for providing details about the experimental structures and for extensive discussions about the analysis presented here. The authors would also like to thank S. Koswatta, H. Pal, and Y. Liu for the helpful discussions and also Prof. G. Klimke for the advice on the tight-binding calculations. Computational resources for this work were provided through nanohub.org by the Network for Computational Nanotechnology (NCN).

REFERENCES


1386 IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 56, NO. 7, JULY 2009

Neophytos Neophytou received the B.S. degree in electrical and computer engineering and the M.S. and the Ph.D. degrees in the area of microelectronics and nanotechnology from Purdue University, West Lafayette, IN, in 2001, 2003, and 2008, respectively. He is currently a Postdoctoral Researcher with the Institute for Microelectronics, TU Wien, Vienna, Austria, working on the effects of bandstructure on the electronic properties of nanoscale devices as well as the effect of atomic variations and defects in their transport capabilities. His research interests include computational modeling of quantum mechanical electron transport through new channel materials such as carbon nanotubes, nanowires, graphene-based channels, and III–V materials.

Titash Rakshit was born in Kolkata, India. He received the Ph.D. degree in the area of nanoelectronics from Purdue University, West Lafayette, IN, in 2004. His Ph.D. thesis involved theoretical simulation of electronic conduction in silicon-based molecular systems. Since 2005, he has been with Intel Corporation, where he is currently a Staff Research Scientist. His work at Intel involves research on future generation logic devices and circuits.

Mark S. Lundstrom (F’94) received the B.E.E. and M.S.E.E. degrees from the University of Minnesota, Minneapolis, in 1973 and 1974, respectively, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1980. From 1974 to 1977, he was with the Hewlett-Packard Corporation, Loveland, CO, working on integrated circuit process development and manufacturing support. In 1980, he joined the School of Electrical Engineering, Purdue University, where he is currently the Don and Carol Scifres Distinguished Professor of electrical and computer engineering and the Founding Director of the Network for Computational Nanotechnology. He served as the Director of Purdue University’s Optoelectronics Research Center from 1989 to 1993 and as an Assistant Dean of Engineering from 1991 to 1994. His current research interests center on the physics of small electronic devices, particularly nanoscale transistors, on carrier transport in semiconductor devices, and on devices for energy conversion, storage, and conservation. Prof. Lundstrom currently serves as an IEEE Electron Devices Society Distinguished Lecturer and is an elected member of the IEEE Electron Devices Society administrative council. Over the course of his career, he has received a number of awards for his contributions to research and education, and in 2009, he was elected to the U.S. National Academy of Engineering.