Physics and Mitigation of Excess OFF-State Current in InGaAs Quantum-Well MOSFETs

Jianqiang Lin, *Student Member, IEEE*, Dimitri A. Antoniadis, *Life Fellow, IEEE*, and Jesús A. del Alamo, *Fellow, IEEE*

***Abstract*— A number of recent reports have noted excess OFF-state leakage current (I_{OFF}) in scaled InGaAs quantum-well nMOSFETs. There is growing evidence that a combination of band-to-band tunneling (BTBT) and a floating-body bipolar gain effect is responsible for this. Unless this issue is effectively addressed, the scaling potential of this transistor structure will be compromised. This paper presents a detailed study of the physics of I_{OFF} and explores I_{OFF} reduction strategies through 2-D device simulations that have been calibrated with experiments. In essence, under OFF conditions at even moderate values of V_{ds}, a BTBT process at the drain-end generates holes in the channel and thereby reduces the source–channel potential barrier. This results in injection of electrons into the channel that contribute to enhanced I_{OFF} while the holes are injected into the source where they recombine. In a nanoscale device, the bipolar effect that is at play here can have a very large current gain and amplify many fold even a small BTBT current. A study of approaches to mitigating this effect is analyzed here. It is concluded that the most effective strategy is to minimize the bipolar current gain rather than BTBT in scaled transistors.**

***Index Terms*— III–V, band-to-band tunneling (BTBT), bipolar effect, floating body, MOSFETs, quantum-well (QW), self-aligned.**

I. Introduction

InAs-RICH InGaAs is a promising channel material for future CMOS applications due to its superior electron transport properties [1], [2]. In recently demonstrated scaled InGaAs quantum-well (QW) MOSFETs with tight source and drain (SD) spacing around the gate, excess drain-to-source leakage current has been observed that prevents the effective turn-OFF of the transistor [3]–[6]. This is a serious problem that must be addressed before the logic potential of this material system can be realized. As a reference, the International Technology Roadmap for Semiconductors (ITRS) requires I_{OFF} of 10 pA/μm for low-standby-power applications [7].

Recently, there has been mounting evidence that excess OFF-state current has its origin in band-to-band tunneling (BTBT) at the drain-end of the channel [3]–[6]. In [3], I_{OFF} was found to feature the characteristic voltage and temperature signatures of gate-induced drain leakage. It was postulated that I_{OFF} arises from BTBT at the drain-end of the channel [3]. A more detailed study has recently shown that while BTBT indeed is the likely triggering mechanism, a parasitic bipolar transistor action in the floating QW-channel is responsible for significant amplification of the BTBT current [6]. The simulations in [6] indicate that the bipolar current gain in nanoscale devices can be rather large, easily above 100, and it rises rapidly with channel length reduction. This greatly magnifies the BTBT current and makes the excess I_{OFF} problem much worse.

It has been well known for some time that in floating-body silicon-on-insulator (SOI) MOSFETs, the presence of a lateral bipolar junction transistor (BJT) with substantial current gain yields a dramatic enhancement of BTBT and impact-ionization current [8], [9]. A number of anomalies have been traced to this combination of mechanisms, such as the kink effect for analog device applications [10], jitter [11], and premature breakdown [12], among others [10]. In InGaAs high-electron-mobility transistors, floating body effects associated with hole generation through impact ionization have also been widely reported [13]–[15]. These have been found to result in the kink effect, excess gate leakage, excess and frequency-dependent output conductance, and premature breakdown and burnout [16].

Multiple factors can influence the magnitude of BTBT and bipolar gain in InGaAs QW-MOSFETs, and a systematic study of this is yet to be performed. The purpose of this paper is to fill this gap through a simulation study that has been calibrated with experiments. As compared with [6] whose focus was the experimental verification of the coupling of BTBT and the bipolar gain effect, this paper targets detailed physical understanding and exploration of mitigation strategies.

This paper is organized as follows. Section II summarizes the physics of BTBT and bipolar amplification in QW-MOSFETs. In Sections III and IV, we present the simulations of the OFF-state current that illuminate BTBT-induced source–channel barrier lowering and the bipolar gain amplification and its L_g dependence. Section V describes a study of device design approaches to mitigate the OFF-state current problem. In particular, we study the role of: 1) doping of the channel region; 2) InAs composition of channel and cap; 3) carrier lifetime; and 4) geometrical design of the access regions. Finally, Section VI provides the conclusions of this paper.

Manuscript received January 6, 2015; revised February 27, 2015; accepted March 3, 2015. Date of publication March 20, 2015; date of current version April 20, 2015. This work was supported in part by the Defense Threat Reduction Agency under Contract HDTRA 1-14-1-0057, in part by the National Science Foundation under Award 0939514 through the E3S STC Project, and in part by the Singapore-MIT Alliance for Research and Technology Centre. The review of this paper was arranged by Editor K. J. Chen.

The authors are with the Microsystems Technology Laboratories, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: linjq@mit.edu; daa@mit.edu; alamo@mit.edu).

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.2015.2410292

0018-9383 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

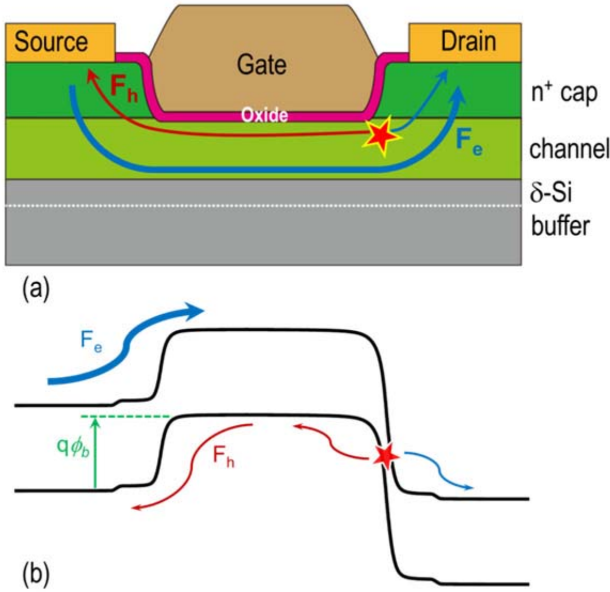

Fig. 1. (a) Schematic of BTBT and bipolar gain action in an InGaAs QW-MOSFET biased in the OFF-state. (b) Corresponding energy band diagram along the channel from source to drain. The source–channel barrier height ($q\phi_b$) determines the OFF-state drain current.

This paper should be instrumental in identifying effective OFF-state current reduction strategies for future nanoscale InGaAs QW MOSFETs.

II. Physics of Excess I_{OFF} in Scaled InGaAs QW-MOSFETs

Experimental evidence of bipolar amplification of BTBT in floating-body QW-MOSFETs was recently presented in [6]. The basic concept is summarized in Fig. 1, which shows a schematic cross section of a simplified InGaAs QW-MOSFET biased in the OFF-state with high V_{ds} and a corresponding energy band diagram; in the rest of this paper, we will refer to this bias condition simply as the OFF-state. Overlaid are arrows that indicate electron and hole flow.

In the OFF-state, BTBT takes place in the high-field region around the drain-end of the channel. As sketched in Fig. 1 by the thin arrows, valence electrons from the channel tunnel across this region and are collected by the drain. The holes that are left behind accumulate in the channel leading to the reduction of the source–channel energy barrier, which under steady-state conditions allows the BTBT-generated holes to be injected into the source and recombine in the heavily doped cap or at the source–contact interface.

An important consequence of the reduction of the source–channel energy barrier is electron injection from the source into the channel. These electrons diffuse through the channel and contribute an additional drain current component. This is indicated in Fig. 1 by the thick blue arrow. In essence, this is a bipolar gain effect similar to what has been observed in floating-body SOI MOSFETs [8], [9]. In a nanoscale device, the gain of the parasitic bipolar transistor can be quite large and, as a result, the total drain current can be many times higher than the BTBT electron–hole generation current.

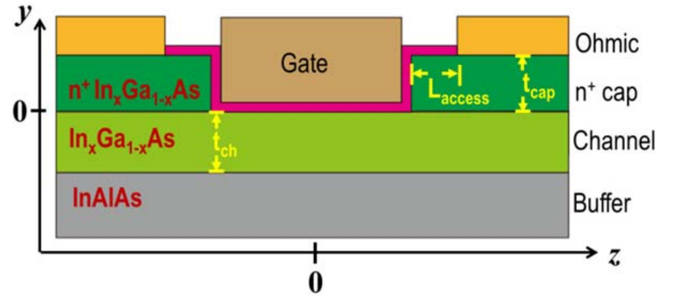

Fig. 2. Cross-sectional device structure used in simulations.

III. Simulation Setup

A cross-sectional schematic of the simplified InGaAs QW-MOSFET modeled in this paper is shown in Fig. 2. This is a self-aligned device with a raised SD architecture similar to devices in which excess drain current has been observed in the OFF-state [3]–[6]. In our model device, the gate dielectric is characterized by equivalent oxide thickness (EOT) $= 1$ nm. The channel is an 8-nm thick $In_xGa_{1-x}As$ with InAs composition of x_{ch} on an InAlAs buffer layer lattice matched to InP. The cap that forms the raised SD is also $In_xGa_{1-x}As$ but with an InAs composition of x_{cap}. The doping level is 2×10^{19} cm^{-3}, and no doping or δ-doping is present elsewhere. The thickness of the cap and channel is denoted as t_{cap} and t_{ch}, respectively. The space between the gate edge and the ohmic contacts is known as the access region and has a length L_{access}. Our baseline structure uses $x_{ch} = 0.7$, $x_{cap} = 0.53$, $t_{cap} = 30$ nm, $t_{ch} = 8$ nm, $L_{access} = 0$, and $L_g = 40$ nm.

The 2-D device simulator Sentaurus Device by Synopsys [17] was used in these simulations. The coupled Poisson, electron and hole continuity equations are solved self-consistently with a nonlocal-path BTBT model. Quantum effects are not included in these simulations due to convergence difficulties. We expect some minor corrections when quantum effects are accounted for as a result of the presence of a quantum dark region that pushes the charge centroid away from the front and back interfaces of the channel. In the baseline simulation setup, only recombination at the SD metal contacts, which are characterized by an infinite surface recombination velocity, is allowed. The contacts are long enough for their length not to be relevant. Recombination in the body of the channel and the cap is neglected because the carrier diffusion lengths in these two regions are estimated to be much longer than the gate length and the cap thickness, respectively. This estimation is based on the experimental carrier lifetimes of $In_{0.53}Ga_{0.47}As$/InP structures [18] and reasonable mobilities [19], [20], and yields a diffusion length that is over 10 μm in the channel and $\sim$0.2 μm in the n^+ cap. Nevertheless, we explore the effect of carrier recombination in the body of the semiconductor in Section V.

For this paper, we apply the TCAD models and parameters in [6] as they match the experimental results reasonably well.

IV. OFF-State Current Modeling

This section gives the results of simulations of devices with different L_g values, where BTBT is turned ON and OFF.

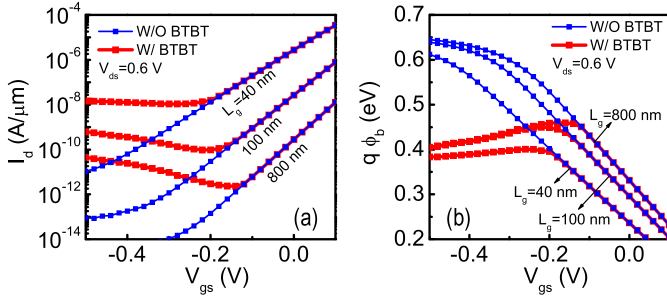

Fig. 3. (a) Simulated subthreshold I_d–V_{gs} characteristics with and without BTBT for MOSFETs with gate lengths of 40, 100, and 800 nm. (b) Corresponding source–channel energy barrier ($q\phi_b$) as a function of V_{gs}. In all cases, $V_{ds} = 0.6$ V.

The goal is to provide physical understanding of the relevant physics.

A. BTBT-Induced Source–Channel Barrier Lowering

Fig. 3(a) shows the simulated subthreshold I_d–V_{gs} characteristics at $V_{ds} = 0.6$ V for $L_g = 40$, 100, and 800 nm MOSFETs with and without BTBT. For all devices, I_d decreases monotonically as V_{gs} is reduced when the BTBT model is not included, as expected in an ideal MOSFET. The subthreshold swing degradation with decreasing L_g is due to short-channel effects (SCEs). In contrast, when the BTBT model is included, the OFF-state drain current, I_d ($V_{gs} < -0.2$ V, $V_{ds} = 0.6$ V), reaches a minimum value and then increases as V_{gs} decreases far below threshold. This level of I_d is set by the coupling between BTBT, bipolar gain, and gate voltage through the modulation of the height of the potential barrier, ϕ_b, that exists between the source and the channel under the gate (Fig. 1).

To better understand this, Fig. 4 plots the conduction band edge (E_c) from source to drain in the longitudinal direction at a depth of 2 nm into the channel ($y = -2$ nm in Fig. 2) for two transistors with $L_g = 800$ and 40 nm. With the BTBT model turned OFF (Fig. 4, left panels), as V_{gs} decreases, $q\phi_b$ increases until normal hole accumulation in the channel becomes prevalent. In contrast, with the BTBT model present (right panels), $q\phi_b$ increases only initially with decreasing V_{gs} and then becomes independent of V_{gs}. This is what causes I_d in the OFF-state to essentially saturate. Making V_{gs} more negative flattens the top of the inverse-U-shaped band profile in the channel. This is because a significant concentration of holes accumulates in the channel leading to a more effective screening of the longitudinal charge distribution.

The evolution of $q\phi_b$ with V_{gs} for the simulations of Fig. 3(a) is more clearly visible in Fig. 3(b). In the absence of BTBT, $q\phi_b$ increases as V_{gs} becomes more negative. Eventually, saturation takes place when the Fermi level at the surface approaches the valence band edge and the channel goes into hole accumulation. Including BTBT introduces increased accumulation of holes that limits the rise of $q\phi_b$ at higher values of V_{gs} as follows. Reduction in V_{gs} gives rise to higher V_{dg} and hence higher electric field at the drain edge of the channel [3]; this enhances the BTBT generated hole current

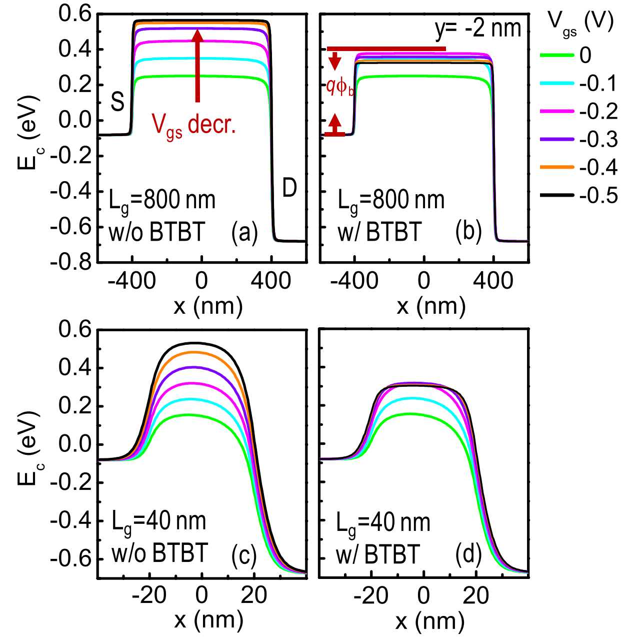

Fig. 4. Simulated conduction band edge along the channel from source to drain (E_c) at various gate voltages for InGaAs MOSFETs with (a) and (b) $L_g = 800$ nm and (c) and (d) $L_g = 40$ nm. In all cases, V_{ds} is 0.6 V. (a) and (c) BTBT model is turned OFF. (b) and (d) BTBT model is turned ON.

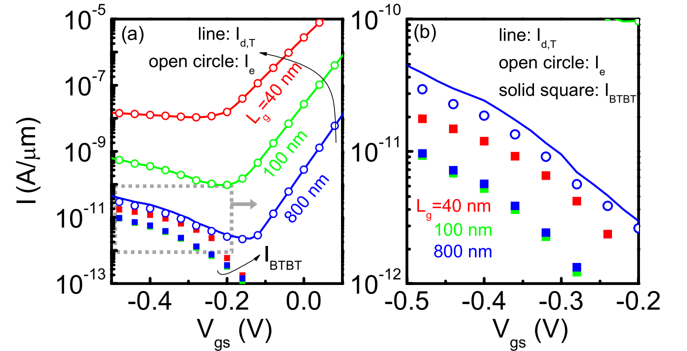

Fig. 5. (a) Electron (I_e) and hole (I_{BTBT}) channel currents and total drain current ($I_{d,T}$) characteristics versus V_{gs} for MOSFETs with gate lengths of 40, 100, and 800 nm at $V_{ds} = 0.6$ V. (b) Enlarged view for the box region in (a).

injected into the channel ($I_{h,\text{BTBT}}$) which in turn brings about a reduction in $q\phi_b$ counterbalancing the effect of V_{gs}.

B. Bipolar Gain Amplification and L_g Dependence

As shown in Fig. 3(b), the source–channel barrier height in the OFF-state in the presence of BTBT is essentially the same for the $L_g = 100$ and 800 nm devices, which is understandable because the BTBT current ($I_{\text{BTBT}} = |I_{e,\text{BTBT}}| = |I_{h,\text{BTBT}}|$) only depends on the E-field at the drain edge and is independent of gate length. This is indeed verified in Fig. 5 that plots I_{BTBT} corresponding to the three gate lengths. Compared with the difference between total currents $I_{d,T}$, the I_{BTBT} currents are very similar—almost the same for the two longer devices and within a factor of 3, including the shortest one.

The bipolar current gain, β, on the other hand, does depend on L_g. β can be defined as

$$\beta = \frac{I_e}{I_{\text{BTBT}}}.$$

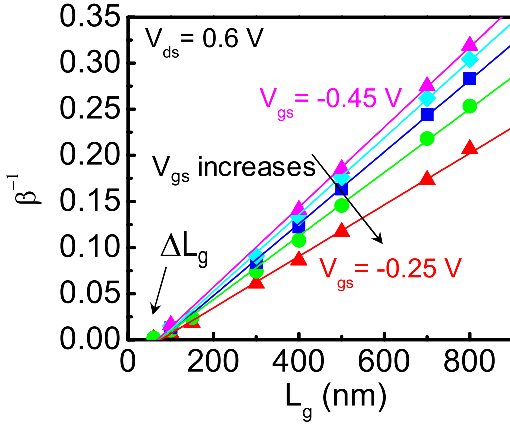

Fig. 6. β^{-1} versus L_g at $V_{\rm ds} = 0.6$ V for $V_{\rm gs}$ values between -0.45 and -0.25 V.

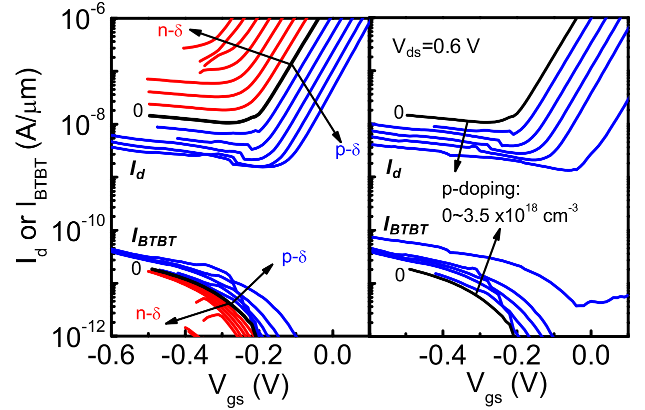

Fig. 7. Subthreshold I_d–$V_{\rm gs}$ and $I_{\rm BTBT}$–$V_{\rm gs}$ characteristics for $L_g = 40$ nm for different (a) backside δ-doping type and doping level and (b) channel uniform p-type doping level. For both simulations, $V_{\rm ds}$ is 0.6 V.

In the subthreshold regime, electron transport through the channel takes place by diffusion and a shorter L_g leads to a steeper electron gradient and higher electron current. Fig. 5(a) also plots the electron channel current (I_e) and the terminal drain current ($I_{d,T}$). $I_{d,T}$ denotes the total current by source electron injection and BTBT, $I_{d,T} = I_e + I_{\rm BTBT}$. It is clear that in spite of a similar $I_{\rm BTBT}$, the shorter devices have significantly higher I_e and, therefore, $I_{d,T}$. For short devices, Since β is much larger than unity ($I_e \gg I_{\rm BTBT}$), it is expected that $I_{d,T} \simeq I_e$. The gain reduces for longer devices. This is indeed the case for the longer device, $L_g = 800$ nm, as shown in Fig. 5(b).

Our simple physical arguments lead us to postulate an inverse linear relationship between β and L_g. Indeed, this is what we observe in the simulations at all values of $V_{\rm gs}$ in Fig. 6, which also shows that, for a given L_g, β is reduced as the gate voltage becomes more negative. The physical origin is discussed as follows.

SCEs complicate this idealized picture in a number of ways. In a BJT, the quasi-neutral base is the portion of the base with nearly zero longitudinal E-field, through which injected carriers flow by diffusion. Due to the presence of the depletion regions associated with the p-n junctions, the effective electrical thickness of the quasi-neutral base is smaller than that of the metallurgical base. In analogy with this, in a QW-MOSFET in the subthreshold regime, the portion of the channel that is under negligible lateral field is shorter than the gate length. This does not appreciably affect long-channel devices, but it becomes significant in short-channel transistors. As Fig. 4 shows, for $L_g = 800$ nm, the conduction band under the gate is rather flat, while for $L_g = 40$ nm, it exhibits an inverse-U shape. The effective channel length in this case is substantially shorter than L_g leading to an increase in current gain.

We find in our simulations that for a given layer design, the difference between physical gate length and effective channel length is weakly dependent of $V_{\rm gs}$. To the first order, we neglect this dependency. Then this difference can be seen in Fig. 6 where the extrapolation of β^{-1} to $\beta^{-1} = 0$ yields $\Delta L_g \simeq 70$ nm. ΔL_g is related to the penetration of the electric field lines that emanate from the SD into the channel. This effectively shortens the effective channel length that electrons diffuse through on their way from source to drain.

Fig. 6 reveals that β decreases as $V_{\rm gs}$ is made more negative. As in an n-p-n bipolar transistor, β is related to the barrier that holes must overcome to get injected into the emitter (source, in this case). The lower the barrier, the easier the injection and the lower β is. In this regime, the reduction of β as $V_{\rm gs}$ becomes more negative, is associated with the reduction of ϕ_b for negative $V_{\rm gs}$, as shown in Fig. 3(b).

V. Transistor Design for OFF-State Leakage Reduction

In this section, we discuss the device design approaches to reduce BTBT and bipolar gain. The degree to which such approaches might be employed to mitigate excess OFF-state I_d in InGaAs QW-MOSFETs would depend on the considerations of the overall device performance and specifications. A detailed study of this is outside the scope of this paper.

A. Delta Doping and Channel Doping

Different doping profiles have been implemented in experimental InGaAs QW-MOSFETs. N-type δ-doping is typically added beneath the channel (indicated by dashed line in Fig. 1) to reduce ON-state access resistance and to maintain high channel mobility [3], [21]. Uniform p-type doping of the channel region has also been used [22]. We have found that the doping type and profile strongly influence the BTBT rate and the bipolar gain characteristics, as well as many other important transistor parameters, and should be an important consideration in transistor design.

Fig. 7 shows the simulated subthreshold I_d–$V_{\rm gs}$ and $I_{\rm BTBT}$–$V_{\rm gs}$ characteristics at $V_{\rm ds} = 0.6$ V as a function of backside δ-doping type and level (left), and channel uniform p-type doping level (right) for $L_g = 40$ nm transistors. The backside δ-doping is located in the InAlAs buffer 5-nm beneath the channel. In Fig. 7(a), the black line shows the case for zero δ-doping (the baseline design) and the arrows indicate increasing magnitude of δ-doping in 0.25×10^{12} cm^{-2} increments. Red color with a left-pointing arrow refers to

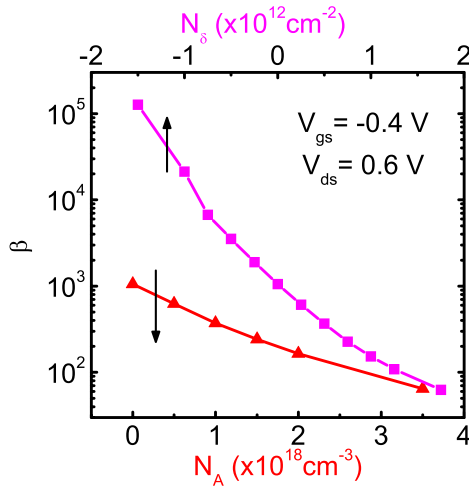

Fig. 8. β versus channel uniform p-doping (N_A at the bottom x-axis) and backside δ-doping level (N_δ at the top x-axis) at $V_{gs} = -0.4$ V and $V_{ds} = 0.6$ V. L_g for this device is 40 nm. Negative δ-doping values indicate n-type doping.

n-type and blue with a right-pointing arrow for p-type. In Fig. 7(b), the arrows indicate channel p-doping level changes from 0 to 3.5×10^{18} cm^{-3}.

Changing doping level and type affects the threshold voltage, which shifts positively with increasing p-type doping and negatively with increasing n-type doping. In addition, the saturated I_d (in the OFF-state) increases as the n-type doping increases and it decreases as p-type doping increases though eventually this reduction saturates. Interestingly, the behavior of I_{BTBT} is counter to this, it decreases as n-type doping increases, and it increases as p-type doping increases.

Fig. 7 reveals the important role that β plays in OFF-state I_d and its dependence on doping. This can be seen more clearly in Fig. 8 that plots β at $V_{gs} = -0.4$ V and $V_{ds} = 0.6$ V against the doping level in the δ-doped layer (top x-axis) and channel p-type doping N_A (bottom x-axis).

The large impact of channel doping on the bipolar gain can be understood by considering the role of the vertical electric field across the channel. In the channel of a QW-MOSFET biased in the subthreshold regime, an electric field exists in the transverse direction (perpendicular to the surface). Its presence can significantly influence the bipolar action that is partially responsible for the excess OFF-state current. To illustrate this, Fig. 9 shows the energy band diagram in the transverse direction under the gate in a device biased in the OFF-state with some degree of hole accumulation in the channel. The presence of a transverse field splits the electron and hole distributions spatially with holes being located closer to the surface and the electrons placed against the channel–buffer interface. Effectively this also implies an energy separation between the electron and the hole populations that, neglecting quantization, is narrower than the bandgap by an amount roughly equal to the band bending in the channel or $q\phi_c$ in Fig. 9. In analogy to the impact on β of bandgap narrowing in the base of a BJT, this should increase the current gain by an amount roughly given by $\exp(q\phi_c/kT)$. Similar behavior in the bipolar action of a Si MOSFET has been reported [23].

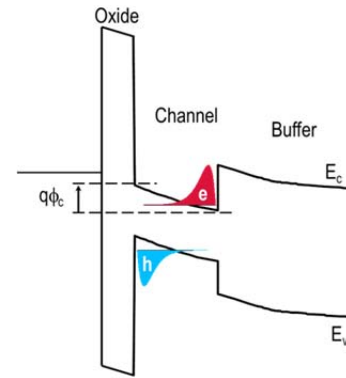

Fig. 9. Energy band diagram below the gate around the source in the direction perpendicular to the surface.

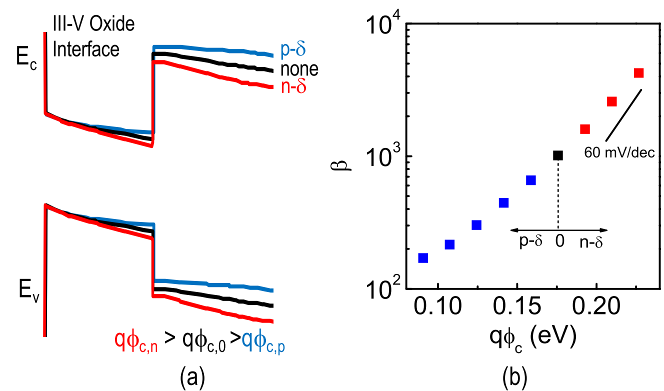

Fig. 10. (a) Simulated energy band diagrams in the perpendicular direction in the channel under the gate for three different types of backside δ-doping (at the same $V_{gs} - V_t$). For clarity, the band diagrams are matched at the oxide–semiconductor interface. (b) β versus $q\phi_c$ at fixed $V_{gs} - V_t = -0.35$ V and $V_{ds} = 0.6$ V in semilog scale indicates the exponential dependence of β on $q\phi_c$. $q\phi_c$ is extracted in the center of the channel ($z = 0$).

In addition, in Si SOI MOSFETs, an increase in OFF-state I_d is observed after irradiation [24], [25] that generates positive fixed charge in the buried oxide beneath the channel and leads to electron accumulation at the back interface.

The impact of channel doping on the bipolar gain shown in Fig. 8 can be easily understood in these terms. Fig. 10(a) shows energy band diagrams in the vertical direction in the channel under the gate at the same $V_{gs} - V_t$ ($V_{gs} - V_t < 0$) for three different cases: 1) p-type δ-doping; 2) no δ-doping; and 3) n-type δ-doping. For clarity, the band diagrams are matched at the oxide–semiconductor interface. With an n-type δ-doping in the buffer, there is steeper band bending in the channel that results in increased current gain. Conversely, with increased p-type delta doping, ϕ_c is reduced and β also shrinks. Similar reasoning applies for the uniform p-type channel doping design.

To confirm this understanding, Fig. 10(b) shows β versus ϕ_c as extracted from the simulator for different δ-doping designs. Those values are extracted at fixed $V_{gs} - V_t = -0.35$ V (V_t is defined at $I_d = 1$ μA/μm) and $V_{ds} = 0.6$ V from Fig. 7(a). The expected exponential dependence of β on ϕ_c is indeed obtained.

This discussion reveals that the optimum design is the one that minimizes band bending in the channel in the transverse

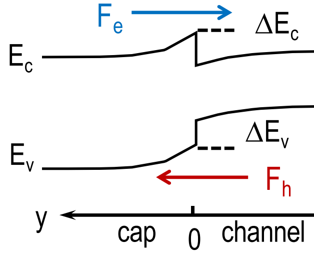

Fig. 11. Energy band diagram under the contact along the perpendicular direction for $x_{\mathrm{cap}} < x_{\mathrm{ch}}$, ($E_{g,\mathrm{cap}} > E_{g,\mathrm{ch}}$).

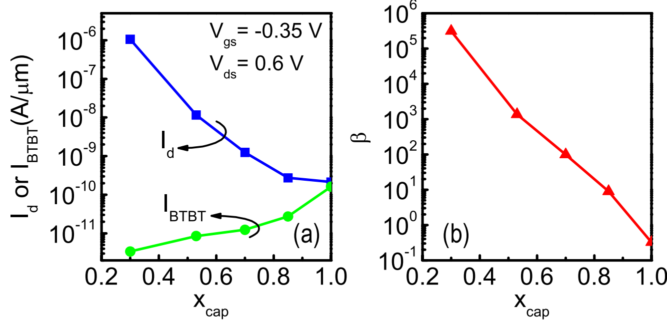

Fig. 12. (a) Extracted I_d and I_{BTBT} and (b) β versus x_{cap} at $V_{\mathrm{gs}} = -0.35$ V and $V_{\mathrm{ds}} = 0.6$ V. L_g for this device is 40 nm and $x_{\mathrm{ch}} = 0.7$.

direction in the subthreshold regime. Consequently, the device architectures that improve the gate control over the channel potential in the subthreshold regime, such as ultrathin-body QW-MOSFET, double-gate FinFET, trigate MOSFET and gate-all-around nanowire design will be an advantage.

B. InAs Composition in Channel and Cap

So far, we have studied a device design in which the InAs composition in the channel is $x_{\mathrm{ch}} = 0.7$ and that of the n^+-cap is $x_{\mathrm{cap}} = 0.53$. The use of a different InAs composition in the cap and channel results in a heterojunction at the cap/channel interface. This introduces an additional barrier to hole injection from the channel into the cap that greatly increases the bipolar current gain. In essence, this behaves as a heterojunction bipolar transistor [26]. Fig. 11 shows this case by sketching an energy band diagram under the contacts in the vertical direction.

We have studied the effect of varying x_{cap} while keeping x_{ch} constant. Fig. 12 graphs I_d, I_{BTBT}, and β versus x_{cap} at a fixed $V_{\mathrm{gs}} = -0.35$ V and $L_g = 40$ nm for $x_{\mathrm{ch}} = 0.7$. It is clear that OFF-state I_d decreases when x_{cap} increases from 0.3 to 1. This is because β decreases in a very significant way as the hole barrier at the channel–cap interface is reduced and eventually eliminated. In contrast, the BTBT current increases as x_{cap} increases.

BTBT is inherently a nonlocal process. At the drain edge of the channel, electron generation mostly happens in the cap, and hole generation in the channel [6]. Hence, the InAs composition of both the cap and the channel, through their respective bandgaps, affects the tunneling barrier. As the bandgap in the cap is reduced, the tunneling barrier for BTBT also shrinks and BTBT at the drain edge of the channel is enhanced. This is consistent with experimental observations in [5]. The increase in I_{BTBT}, however, is significantly weaker than the bipolar gain reduction and OFF-state leakage follows the trend of the latter. When x_{cap} approaches 1, β vanishes, and OFF-state leakage approaches I_{BTBT}.

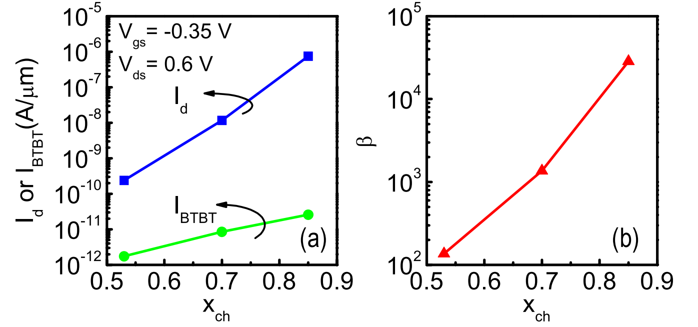

Fig. 13. (a) I_d and I_{BTBT} and (b) β versus x_{ch} at $V_{\mathrm{gs}} = -0.35$ V and $V_{\mathrm{ds}} = 0.6$ V. L_g for this device is 40 nm and $x_{\mathrm{cap}} = 0.53$.

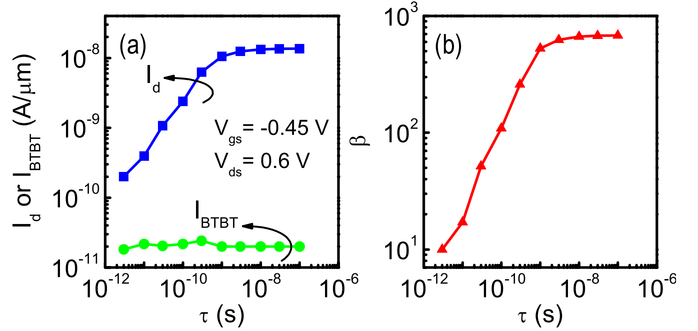

Fig. 14. (a) Extracted I_d and I_{BTBT} and (b) β versus τ at $V_{\mathrm{gs}} = -0.45$ V and $V_{\mathrm{ds}} = 0.6$ V. L_g for this device is 40 nm.

Fig. 13 shows the impact of x_{ch} for a fixed $x_{\mathrm{cap}} = 0.53$. OFF-state I_d increases as x_{ch} increases. In this case, β and I_{BTBT} are both enhanced. β enhancement can be understood by the heterojunction effect and the BTBT current enhancement as a result of bandgap reduction in the channel.

This paper reveals that the InAs composition in channel and cap both play a very important role in determining OFF-state I_d. All other things being equal, it is clear that reducing the valence band discontinuity ($\Delta E_v = E_{v,\mathrm{ch}} - E_{v,\mathrm{cap}}$) at the channel–cap interface constitutes an effective way to suppress OFF-state current.

C. Carrier Lifetime

In SOI MOSFET, an effective approach to mitigate the parasitic bipolar effect is to introduce recombination centers that reduce the carrier lifetime [27]. Recombination of the BTBT-generated holes in the channel brings down their concentration and the need to inject them across the source–channel barrier.

This effect is studied by introducing a Shockley–Read–Hall-type recombination process in the cap and channel that is characterized by a carrier lifetime, τ [18]. Fig. 14 shows the impact of τ on OFF-state I_d, I_{BTBT}, and β at $V_{\mathrm{gs}} = -0.45$ V and

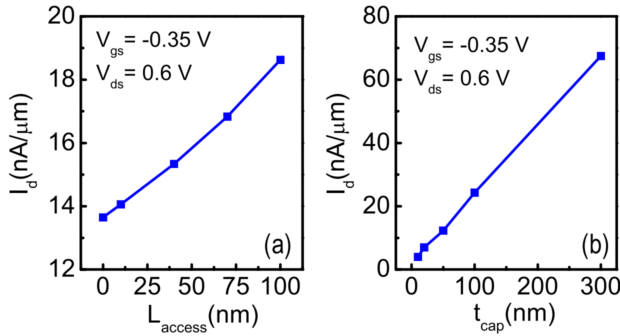

Fig. 15. I_d extracted at $V_{gs} = -0.35$ V and $V_{ds} = 0.6$ V for (a) varying L_{access} and (b) varying t_{cap} in, otherwise, a baseline device design.

$V_{ds} = 0.6$ V for the baseline device. As expected, the BTBT current is independent of τ. The I_d behavior then follows the evolution of the bipolar gain. For $\tau > 1$ ns, I_d and β are almost independent of τ. In this regime, the carrier diffusion length is much longer than the device dimensions, so hole recombination mainly takes place at the contacts. As τ falls below 1 ns, I_d and β decrease rapidly as a result of hole recombination in the channel and cap.

In Si MOSFETs, carrier lifetime reduction can be achieved by the implantation of various impurities, e.g., germanium [27]. For InGaAs MOSFETs, lifetime reduction could be implemented by introducing recombination centers during SD regrowth or in the channel recess process. Of course, a concern in doing this is its impact on the transport characteristics.

D. Dimensional Design of Access Region

Fig. 15 shows the impact of access region design on OFF-state I_d. Increasing the access region length and the cap thickness both result in higher I_d. In both cases, I_{BTBT} is independent of L_{access} and t_{cap} as the access region is heavily doped and the electron concentration in the channel directly underneath is unaffected by changes in either of these two parameters. I_d then follows the bipolar gain β, which exhibits a linear trend with both L_{access} and t_{cap}. This is expected from classic BJT behavior as in both cases hole injection is suppressed as the path that holes must diffuse through in the cap region before they recombine at the contact is lengthened [26]. The linear dependence of I_d on both L_{access} and t_{cap} is weaker than other dependencies studied above and, therefore, engineering the access regions offers limited opportunities for significant suppression of excess leakage in comparison with their relevance in the overall operation of the device.

VI. Conclusion

This paper has explored the physics behind excess OFF-state drain leakage current in InGaAs QW-MOSFETs that has its origin in the combination of BTBT and a bipolar gain effect. We have also discussed mitigating approaches to excess OFF-state leakage current in scaled InGaAs QW MOSFETs. The most effective approach consists of minimizing the current gain of the parasitic bipolar transistor in these devices. We have studied four different ways of accomplishing this: 1) reduction of the vertical electric field in the channel by doping engineering; 2) reduction of the heterojunction effect by engineering the InAs compositions in the cap and channel; 3) reduction of lifetime of excess carriers generated by BTBT; and 4) dimensional control of the extrinsic device.

References

[1] J. A. del Alamo, "Nanometre-scale electronics with III–V compound semiconductors," *Nature*, vol. 479, no. 7373, pp. 317–323, 2011.

[2] M. J. W. Rodwell *et al.*, "Nanometer InP electron devices for VLSI and THz applications," in *Proc. 72nd Device Res. Conf.*, Jun. 2014, pp. 215–216.

[3] J. Lin, X. Zhao, T. Yu, D. A. Antoniadis, and J. A. del Alamo, "A new self-aligned quantum-well MOSFET architecture fabricated by a scalable tight-pitch process," in *Proc. IEEE IEDM*, Dec. 2013, pp. 16.2.1–16.2.4.

[4] S. Lee *et al.*, "High performance raised source/drain InAs/$In_{0.53}Ga_{0.47}As$ channel metal-oxide-semiconductor field-effect-transistors with reduced leakage using a vertical spacer," *Appl. Phys. Lett.*, vol. 103, no. 23, p. 233503, 2013.

[5] J. Mo, E. Lind, G. Roll, and L.-E. Wernersson, "Reduction of off-state drain leakage in InGaAs-based metal-oxide-semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. 105, p. 033516, Jul. 2014.

[6] J. Lin, D. A. Antoniadis, and J. A. del Alamo, "Off-state leakage induced by band-to-band tunneling and floating-body bipolar effect in InGaAs quantum-well MOSFETs," *IEEE Electron Device Lett.*, vol. 35, no. 12, pp. 1203–1205, Dec. 2014.

[7] *International Technology Roadmap for Semiconductors (ITRS)*, 2013. [Online]. Available: http://public.itrs.net

[8] J. Chen, F. Assaderaghi, P.-K. Ko, and C. Hu, "The enhancement of gate-induced-drain-leakage (GIDL) current in short-channel SOI MOSFET and its application in measuring lateral bipolar current gain beta," *IEEE Electron Device Lett.*, vol. 13, no. 11, pp. 572–574, Nov. 1992.

[9] J.-Y. Choi and J. G. Fossum, "Analysis and control of floating-body bipolar effects in fully depleted submicrometer SOI MOSFET's," *IEEE Trans. Electron Devices*, vol. 38, no. 6, pp. 1384–1391, Jun. 1991.

[10] J.-P. Colinge, "The SOI MOSFET," in *Silicon-on-Insulator Technology: Materials to VLSI*, 3rd ed. Norwell, MA, USA: Kluwer, 2004, ch. 5, p. 208.

[11] Y.-C. Tseng, T. M. Huang, T. Monk, T. Diaz, T. M. Ford, and J. C. S. Woo, "Comprehensive study on AC characteristics in SOI MOSFETs for analog applications," in *Proc. IEEE VLSI Technol.*, Jun. 1998, pp. 112–113.

[12] J. M. Hwang, H. Lu, Y. D. Sheu, W. Bailey, P. Mei, and G. Pollack, "Premature breakdown in non-fully depleted SOI/MOSFETs with body-tied-to-source structure," in *Proc. IEEE SOI Conf.*, Oct. 1991, pp. 34–35.

[13] M. H. Somerville, A. Ernst, and J. A. del Alamo, "A physical model for the kink effect in InAlAs/InGaAs HEMTs," *IEEE Trans. Electron Devices*, vol. 47, no. 5, pp. 922–930, May 2000.

[14] A. N. Ernst, M. H. Somerville, and J. A. del Alamo, "Dynamics of the kink effect in InAlAs/InGaAs HEMTs," *IEEE Electron Device Lett.*, vol. 18, no. 12, pp. 613–615, Dec. 1997.

[15] M. H. Somerville, J. A. del Alamo, and W. Hoke, "Direct correlation between impact ionization and the kink effect in InAlAs/InGaAs HEMTs," *IEEE Electron Device Lett.*, vol. 17, no. 10, pp. 473–475, Oct. 1996.

[16] J. A. del Alamo and M. H. Somerville, "Breakdown in millimeter-wave power InP HEMTs: A comparison with GaAs PHEMT's," *IEEE J. Solid-State Circuits*, vol. 34, no. 9, pp. 1204–1211, Sep. 1999.

[17] *Synopsys Sentaurus Device User Guide Version H-2013.03*, Synopsis Inc., Mountain View, CA, USA, 2013.

[18] R. K. Ahrenkiel, R. Ellingson, S. Johnston, and M. Wanlass, "Recombination lifetime of $In_{0.53}Ga_{0.47}As$ as a function of doping density," *Appl. Phys. Lett.*, vol. 72, no. 26, pp. 3470–3472, 1998.

[19] T. Matsuoka, E. Kobayashi, K. Taniguchi, C. Hamaguchi, and S. Sasa, "Temperature dependence of electron mobility in InGaAs/InAlAs heterostructures," *Jpn. J. Appl. Phys.*, vol. 29, no. 10, p. 2017, 1990.

[20] I. J. Fritz, T. J. Drummond, G. C. Osbourn, J. E. Schirber, and E. D. Jones, "Electrical transport of holes in GaAs/InGaAs/GaAs single strained quantum wells," *Appl. Phys. Lett.*, vol. 48, no. 24, pp. 1678–1680, 1986.

[21] J. Lin, D. A. Antoniadis, and J. A. del Alamo, "Sub-30 nm InAs quantum-well MOSFETs with self-aligned metal contacts and sub-1 nm EOT HfO_2 insulator," in *Proc. IEEE IEDM*, Dec. 2012, pp. 32.1.1–32.1.4.
[22] A. Majumdar *et al.*, "CMOS-compatible self-aligned $In_{0.53}Ga_{0.47}As$ MOSFETs with gate lengths down to 30 nm," *IEEE Trans. Electron Devices*, vol. 61, no. 10, pp. 3399–3404, Oct. 2014.
[23] S. Verdonckt-Vandebroek, S. S. Wong, J. C. S. Woo, and P. K. Ko, "High-gain lateral bipolar action in a MOSFET structure," *IEEE Trans. Electron Devices*, vol. 38, no. 11, pp. 2487–2496, Nov. 1991.
[24] F. E. Mamouni *et al.*, "Gate-length and drain-bias dependence of band-to-band tunneling-induced drain leakage in irradiated fully depleted SOI devices," *IEEE Trans. Nucl. Sci.*, vol. 55, no. 6, pp. 3259–3264, Dec. 2008.
[25] V. Ferlet-Cavrois *et al.*, "Total dose induced latch in short channel NMOS/SOI transistors," *IEEE Trans. Nucl. Sci.*, vol. 45, no. 6, pp. 2458–2466, Dec. 1998.
[26] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. New York, NY, USA: Wiley, 1981.
[27] K. R. Mistry *et al.*, "Parasitic bipolar gain reduction and the optimization of 0.25-μm partially depleted SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 46, no. 11, pp. 2201–2209, Nov. 1999.

Jianqiang Lin (S'08) received the B.Eng. (Hons.) and M.Eng. degrees in electrical engineering from the National University of Singapore, Singapore, in 2007 and 2009, respectively. He is currently pursuing the Ph.D. degree with the Massachusetts Institute of Technology, Cambridge, MA, USA.

He was a Research Intern with the IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, in 2012 and 2013, respectively.

Dimitri A. Antoniadis (M'79–SM'83–F'90–LF'14) received the B.S. degree in physics from the National University of Athens, Athens, Greece, in 1970, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1972 and 1976, respectively.

He joined Massachusetts Institute of Technology, Cambridge, MA, USA, in 1978, where he is currently the Ray and Maria Stata Professor of Electrical Engineering.

Jesús A. del Alamo (S'79–M'85–SM'92–F'06) received the Telecommunications Engineer degree from the Polytechnic University of Madrid, Madrid, Spain, in 1980, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1983 and 1985, respectively.

He has been with the Massachusetts Institute of Technology, Cambridge, MA, USA, since 1988, where he is currently a Donner Professor and a MacVicar Faculty Fellow.