



# MIT Open Access Articles

## *A Self-Aligned InGaAs HEMT Architecture for Logic Applications*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

<b>Citation</b>	Waldron, Niamh, Dae-Hyun Kim, and JesÚs A. del Alamo. "A Self-Aligned InGaAs HEMT Architecture for Logic Applications." IEEE Transactions on Electron Devices 57.1 (2010): 297–304. © Copyright 2012 IEEE
<b>As Published</b>	<a href="http://dx.doi.org/10.1109/ted.2009.2035031">http://dx.doi.org/10.1109/ted.2009.2035031</a>
<b>Publisher</b>	Institute of Electrical and Electronics Engineers (IEEE)
<b>Version</b>	Final published version
<b>Citable link</b>	<a href="http://hdl.handle.net/1721.1/71795">http://hdl.handle.net/1721.1/71795</a>
<b>Terms of Use</b>	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

# A Self-Aligned InGaAs HEMT Architecture for Logic Applications

Niamh Waldron, Dae-Hyun Kim, and Jesús A. del Alamo, *Fellow, IEEE*

**Abstract**—In this paper, we present a novel self-aligned process for future III–V logic FETs. Using this process, we have demonstrated enhancement-mode 90-nm-gate-length InGaAs HEMTs with excellent logic figures of merit. We have carried out a detailed analysis of this device architecture to determine its future scaling capabilities. We find that, as the insulator is scaled to achieve enhancement mode, the performance of the device is limited by degradation of the  $I_{ON}/I_{OFF}$  ratio due to gate leakage current. By use of TLM test structures, we have determined that the barrier resistance dominates the source resistance. We use a tri-layer TLM model to predict the expected evolution of the contact resistance as it is scaled to realistic VLSI dimensions and find that the current technology results in resistance values that are two orders of magnitude higher than the desired target for sub-22-nm nodes. Using the model, we explore different options for device redesign. Both  $I_{ON}/I_{OFF}$  and source-resistance limitations imply that the use of a high- $k$  gate dielectric will be required for future device implementations.

**Index Terms**—Contact resistance, FET logic devices, HEMT.

## I. INTRODUCTION

RECENT excellent results achieved by short-channel InGaAs HEMTs [1], [2] have demonstrated the potential for using III–V as a channel material in future CMOS generations. To date, InGaAs HEMTs have typically been used for fiber-optic front-end systems and millimeter-wave applications where the device footprint is not of utmost concern. However, reduction of device pitch is the main driving force for logic processes. The 2007 ITRS roadmap calls for a metal-1 device half pitch of 15 nm for high-volume microprocessors by 2020 [3]. In contrast to this, state-of-the-art HEMTs typically have a source–drain separation on the order of 1.5–2  $\mu\text{m}$ . Therefore, in order to fulfill the promise of III–V-based FETs, a self-aligned architecture that allows a VLSI device pitch to be achieved must be implemented.

The most obvious approach to self-alignment, i.e., that of self-aligning source and drain implants to a refractory gate, has been used in the fabrication of AlGaAs/GaAs-based HEMTs [4], [5]. However, in the case of InAlAs/InGaAs-based devices,

which are preferable because of their higher channel mobility, the low dopant activation levels achieved in the source/drain regions limit the usefulness of this approach. The most typical self-alignment technique for the InAlAs/InGaAs system is to use the head of a T-gate as a shadow mask for source/drain metal deposition [6]. Excellent results have been achieved with 50-nm devices using this approach, but as the degree of self-alignment is determined by the size of the gate head, the source/drain spacing is still 0.3  $\mu\text{m}$  [7]. The fringing capacitance introduced by the overlap of the T-gate with the source/drain regions also poses a disadvantage, and as this approach requires the use of a T-gate, it is not suited for VLSI.

In this paper, which is an updated and expanded version of an earlier presentation [8], we present a novel self-aligned architecture for InGaAs-based HEMTs. We have demonstrated an ohmic-gate separation of less than 30 nm, which is  $> 30\times$  improvement over that of conventional device designs. We employ a nonalloyed ohmic contact scheme, which is a key requirement for VLSI applications. To achieve enhancement-mode devices (another key requirement for logic applications), we have vertically scaled the barrier/insulator thickness ( $t_{\text{ins}}$ ) by means of a dry etch. We demonstrate 90-nm-gate-length enhancement-mode self-aligned devices with excellent logic figures of merit.

We use the results of devices to examine the potential of further scaling this device architecture. We find that as the device becomes enhancement mode, the  $I_{\text{on}}/I_{\text{off}}$  ratio seriously degrades. A detailed study of the source resistance ( $R_s$ ) through TLM test structures reveals that the resistance associated with the barrier layer in the extrinsic device limits  $R_s$ . By applying a trilevel TLM model to the source, we find that, as the device is scaled to true VLSI dimensions,  $R_s$  increases dramatically if the current cap design is maintained. We use this model to investigate alternative approaches to the source design and suggest that, by eliminating the barrier and grading the cap to pure InAs, an acceptable contact resistance could be achieved.

## II. DEVICE FABRICATION

The main aim of this paper is the development and implementation of a self-aligned HEMT architecture that is suited for VLSI applications. Based on the analysis of conventional HEMTs and considerations of previous efforts at self-alignment, as discussed in the Introduction, certain elements were determined to be of key importance to the design. These are as follows.

- 1) Nonalloyed ohmic contacts must be used to avoid reliability issues, to have a planar surface, and to achieve

Manuscript received January 21, 2009; revised September 21, 2009. First published November 17, 2009; current version published December 23, 2009. This work was supported by the MSD Focus Center, ARL, SRC and Intel Corp. The review of this paper was arranged by Editor S. Deleonibus.

N. Waldron was with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with the Interuniversity Microelectronics Center, 3001 Leuven, Belgium (e-mail: waldron@imec.be).

D.-H. Kim and J. A. del Alamo are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

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.2009.2035031

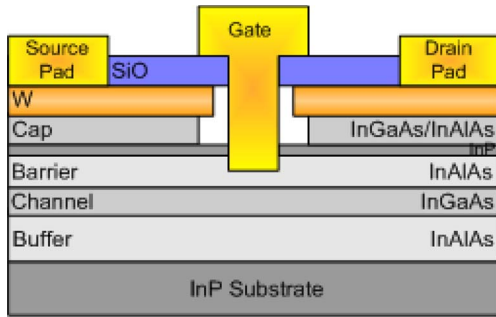


Fig. 1. Schematic cross section of the self-aligned device concept. W source/drain metallizations are separated from the gate by means of an air spacer.

the edge definition required for a very tight ohmic-gate spacing.

- 2) The use of the cap structure should be maintained in order to achieve the minimum possible value for the source and drain resistances.
- 3) The method of self-alignment should not be dependent on the physical structure of the gate. That is, the ohmic-gate separation should easily be varied, irrespective of the gate length or gate-head size.
- 4) The process should be well controlled to enable spacers in the few tens of nanometer range.
- 5) The devices fabricated should be of deep-submicrometer gate lengths to ensure that issues of relevance for future devices are studied.

Using the design space defined by these requirements, the self-aligned device architecture shown in Fig. 1 was conceived and implemented. The key feature of the structure is that a nonalloyed tungsten (W) source/drain metal is self-aligned to the gate by means of an “air spacer.” The self-alignment is achieved by means of a two-step e-beam lithography process. In the first step, the gate foot is defined by etching a SiO overlayer and the W layer. The air spacer is formed by pulling back W from the edge of the SiO layer by means of a highly selective isotropic plasma etch. During the subsequent liftoff process defined by the second lithography step, the deposited gate metal ends up being aligned to the edge of the SiO layer at a distance that is set by the W pull-back etch.

A description of the heterostructure used for device fabrication in this paper is given in [9]. The process flow (Fig. 2) begins with the deposition of 60 nm of tungsten that forms the source/drain ohmic metallization. Tungsten is RF sputter deposited in a homemade Advanced Engineering-based sputter system at a rate of 1 Å/s. Prior to deposition, the substrate is cleaned in a dilute 5%  $\text{NH}_4\text{OH}$  solution. Following W deposition, a 70-nm SiO layer is deposited by the PECVD system at 300 °C [Fig. 2(b)]. The contact resistance of W to the heavily doped cap was found to vary very little up to temperatures of 600 °C when annealed by RTA. The value of the W contact resistance measured by TLM test structures of  $60 \Omega \cdot \mu\text{m}$  is consistent with those obtained in [10]. The thermal stability of the contact resistance is the hallmark of an effective nonalloyed ohmic contact scheme, as alloyed contacts in InP-based HEMTs

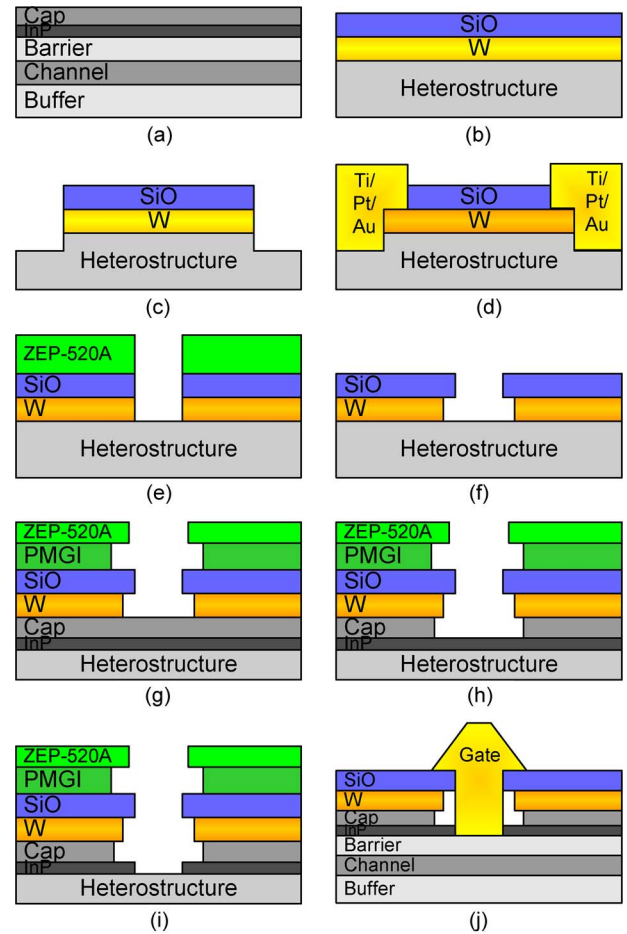


Fig. 2. Self-aligned HEMT process flow [(e) and onward show a zoomed-in version of (a)–(d)]. (a) Starting heterostructure. (b) W and SiO deposition. (c) Mesa isolation. (d) Source/drain pad deposition. (e) First e-beam lithography and W/SiO etch for gate-foot definition. (f) Spacer etch to pull W back from the edge of SiO. (g) Second e-beam lithography; a double-development procedure is used to create an overhang profile for liftoff. (h) Recess etch; the cap is etched under W. (i) Ar RIE etch to remove the InP etch stopper. (j) Gate metal deposition and liftoff.

are very sensitive to temperature and usually severely degrade at temperatures above 400 °C [11].

Device isolation in this process flow is achieved by means of mesa etching [Fig. 2(c)]. The SiO and W layers are both dry etched, and the underlying heterostructure is then wet etched. At this point, contact pads are deposited on W [Fig. 2(d)]. The pads are defined using an image reversal resist, which results in a 2- $\mu\text{m}$  source/drain pad spacing. The SiO layer covering the W is removed with the same  $\text{CF}_4$  dry etch as that used in mesa etching. The resist is still intact after this etch and is used for the liftoff of Ti/Pt/Au (25 nm/25 nm/300 nm) deposited by e-beam evaporation.

The gate of the device is formed by a two-step e-beam lithography process [Fig. 2(e)–(j)]. The first step defines the gate foot in the SiO/W layer and thus determines the gate length [Fig. 2(e)]. The spacer is also formed at this point ultimately, resulting in a self-aligned structure [Fig. 2(f)]. The second step defines the gate head and is used for the liftoff of the gate metal [Fig. 2(g)]. We use a T-gate process here to ensure the mechanical stability of the gate as it runs over the step formed by mesa isolation. In a planar isolation process such as STI, our

gate fabrication can be performed in a single lithography step, resulting in a trapezoidal, rather than a T-gate, shape. In this case, the oxide may also be removed to reduce the gate fringing capacitance.

The first step of the gate formation process involves gate-foot definition in the SiO/W bilayer system [Fig. 2(e)]. A high-resolution process is required to achieve short gate lengths. To this end, a 500-nm layer of ZEP-520A resist from Nippon Zeon Corporation was used. ZEP-520A offers a large etch resistance, as well as high resolution, and has been used to define gate lengths as short as 30 nm [12]. The gate-foot pattern is written in the ZEP layer by means of a Raith-150 e-beam lithography tool at an energy of 30 keV and a dose of  $1700 \mu\text{C}/\text{cm}^2$ . The SiO and W layers are dry etched using  $\text{CF}_4/\text{He}$  and  $\text{SF}_6/\text{O}_2$  chemistries, respectively. The W etch was adapted from the work of Pearton *et al.* [13], with the dc bias being set at  $-50 \text{ V}$  to avoid plasma-induced damage in the semiconductor. At this point, W is coincident with the edge of the SiO layer [Fig. 2(e)]. A lateral plasma  $\text{O}_2/\text{CF}_4$ -based etch is now performed, which selectively pulls W back from the SiO edge, thus creating an “air spacer” [Fig. 2(f)]. This spacer etch is the key enabling step of the process. The lateral etch rate of W is well controlled at about  $0.2 \text{ nm/s}$  and can thus be timed to set the desired distance to the gate. The selectivity of the etch to the SiO layer and heterostructure was measured within our metrology capabilities to be  $>100:1$ . The gate metal will subsequently be deposited and fill the opening defined by the SiO layer. Therefore, W that forms the source/drain metallization becomes self-aligned to the gate.

The spacer etch is followed by the second e-beam lithography step that is used for gate liftoff. The substrate is coated with a bilayer of resist that consists of  $1\text{-}\mu\text{m}$  PMGI and a 200-nm-layer of 1:1 ZEP-520A:anisole. The gate head is aligned back to the previously etched gate-foot pattern and is exposed at 30 keV with a dose of  $50 \mu\text{C}/\text{cm}^2$ . A double-development process is used to create an overhang profile that is suitable for liftoff [Fig. 2(g)]. ZEP-520A is developed with a 2:3 mixture of MEK:MIBK. The PMGI is then wet etched using the CD-26 developer that undercuts ZEP-520A. This second lithography step could be omitted if a trapezoidal-shaped gate is desired and a thicker ZEP-520A layer is used in the first step to define both gate-foot and liftoff patterns.

Gate-recess etching is performed following a two-step recess approach originally proposed by Suemitsu [14]. The cap layers of InGaAs and InAlAs are wet etched with a 20:1 citric acid: $\text{H}_2\text{O}_2$  mixture that selectively stops on the InP layer [Fig. 2(h)]. The cap is overetched so that there is sufficient lateral distance between the cap and the gate. From previous experience in fabricating devices on this heterostructure, the lateral etching distance ( $L_{\text{side}}$ ) is targeted to be 150 nm [15] to strike the best balance between performance and short-channel effects. The actual devices had an  $L_{\text{side}}$  of 200 nm. In a future optimized heterostructure, this distance should optimally be designed to be identical to the ohmic-gate spacing.

Following this, the InP etch stopper is removed using an Ar-based RIE step [Fig. 2(i)]. This Ar etch was reported by Suemitsu [16] to be somewhat selective to the InAlAs insulator layer with a selectivity on the order of 17:1. We have found

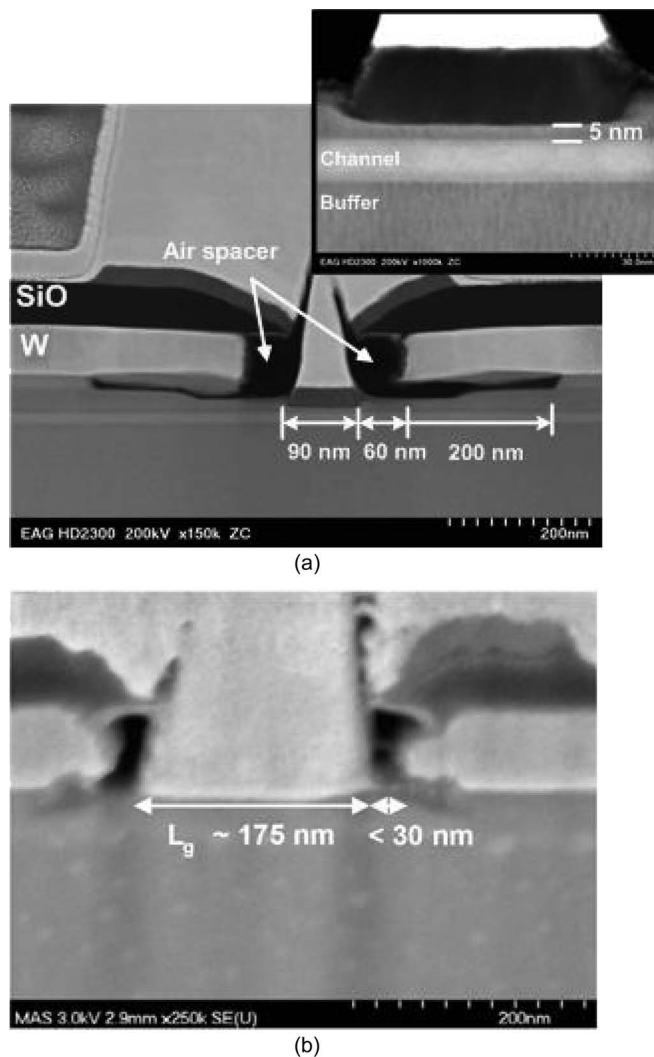


Fig. 3. (a) TEM of a completed 90-nm self-aligned device. The gate-to-source-ohmic metal distance is 60 nm.  $L_{\text{side}}$  is estimated to be 200 nm. Device optimization could reduce the need for such a large  $L_{\text{side}}$ . The inset shows that the barrier thickness has been thinned to 5 nm. (b) Cross-sectional SEM of a completed self-aligned 175-nm device where the ohmic-gate spacing is less than 30 nm. (This device is operational, but its characteristics are not reported in this paper.)

this etch to be not as selective, and by increasing the etching time, the InAlAs barrier layer could be thinned in a controlled manner. In this paper, we have fabricated devices with gate-to-channel distances of 11 and 5 nm. A feature of this process is that the Ar dry etch only thins the insulator layer directly under the gate, thus preserving the access resistance.

The device is completed by the deposition and liftoff of a Ti/Pt/Au (25 nm/25nm/600 nm) metal stack [Fig. 2(j)]. This thick stack provides adequate step coverage of the gate metal over the mesa edge.

Using the described process, we have fabricated devices with gate lengths between 90 and 240 nm. A cross-sectional TEM of a self-aligned 90-nm device is shown in Fig. 3(a). The InAlAs insulator layer of this device was thinned by the Ar RIE etch and has a gate-to-channel distance of 5 nm. We also fabricated devices with a nominal barrier thickness ( $t_{\text{ins}}$ ) of 11 nm. The ohmic-gate separation is about 60 nm. In a separate run, we succeeded in fabricating functional longer channel devices in



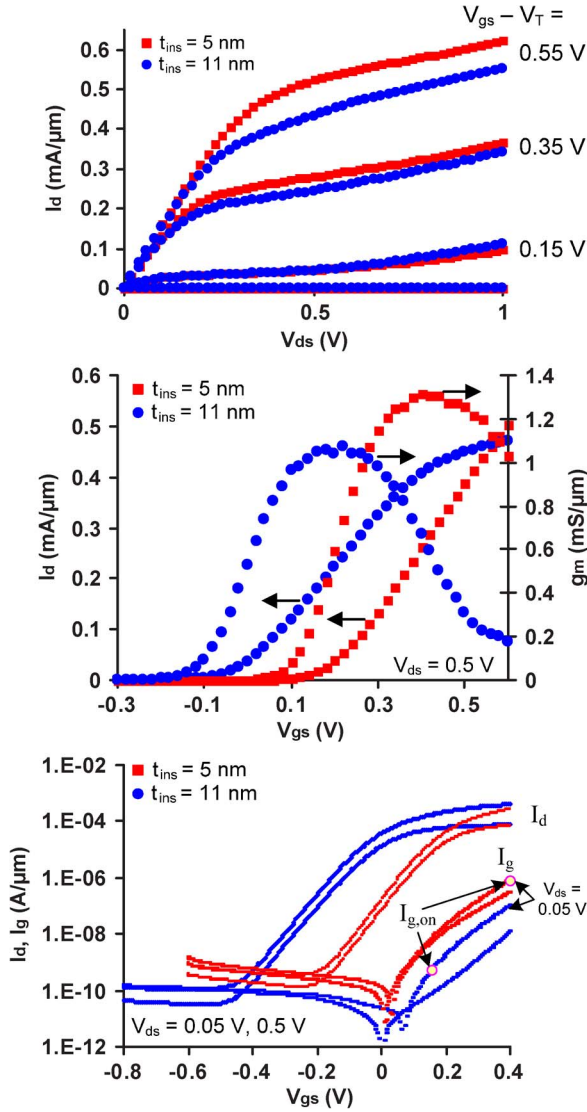


Fig. 4. Output, transfer, and subthreshold characteristics of 90-nm-gate-length devices with  $t_{\text{ins}} = 11$ - and 5-nm devices. The device performance improves as the insulator is scaled. The  $I_{g,\text{ON}}$  point (shown in the  $V_{\text{ds}} = 0.05$ -V curve to estimate the gate-ON-drain-OFF condition) of the enhancement-mode 5-nm- $t_{\text{ins}}$  device is further along the exponential gate leakage current curve, resulting in a reduced  $I_{\text{ON}}/I_{\text{LEAK}}$  ratio.

which the spacer distance was as small as 30 nm [Fig. 3(b)]. This is an improvement of  $> 30\times$  over conventional designs.

### III. RESULTS AND DISCUSSION

#### A. DC Characteristics

Both sets of devices with  $t_{\text{ins}}$ 's of 5 and 11 nm demonstrated excellent dc characteristics. These are shown in Fig. 4 for devices of 90-nm gate length. The thin-barrier devices are in enhancement mode with a  $V_T$  of 60 mV (defined at  $1 \mu\text{A}/\mu\text{m}$ ) for 90-nm gate length, a peak transconductance ( $g_m$ ) of  $1.32 \text{ mS}/\mu\text{m}$ , and drain-induced barrier lowering (DIBL) and subthreshold swing (SS) of  $55 \text{ mV}/\text{V}$  and  $70 \text{ mV}/\text{dec}$ , respectively. This represents a positive shift in  $V_T$  of 0.2 V over the 11-nm- $t_{\text{ins}}$  device and a significant increase in peak  $g_m$  from  $1.07 \text{ mS}/\mu\text{m}$  for the 11-nm- $t_{\text{ins}}$  device. The DIBL and SS val-

ues for  $t_{\text{ins}} = 11 \text{ nm}$  are also worse at  $88 \text{ mV}/\text{V}$  and  $80 \text{ mV}/\text{dec}$ , respectively. A tradeoff of barrier scaling is the increase in gate current of the 5-nm- $t_{\text{ins}}$  device due to enhanced tunneling.

The scaling behavior of the two sets of devices is shown in Fig. 5. The  $g_m$  of the 11-nm- $t_{\text{ins}}$  devices peaks at an  $L_g$  of 145 nm and then begins to decrease, which is consistent with the observations of Shinohara *et al.* [17].  $g_m$  continues to scale to at least 90 nm for the 5-nm- $t_{\text{ins}}$  samples [Fig. 5(a)].  $V_T$  roll-off is much suppressed in thinner insulator samples with a decrease in  $V_T$  of 52 mV for  $t_{\text{ins}} = 5 \text{ nm}$  compared to 132 mV for  $t_{\text{ins}} = 11 \text{ nm}$  as the device gate length scales from 220 to 90 nm [Fig. 5(b)]. DIBL and SS also improve as  $t_{\text{ins}}$  is decreased. All these figures of merit are defined as in [18]. The short-channel figures of merit achieved by  $t_{\text{ins}} = 5$ -nm devices are excellent and at least comparable to the best published data on Si CMOS devices from the literature of similar gate length [Fig. 5(c)]. These results demonstrate the merits of the dry-etch approach and the benefits that are achievable from vertical scaling.

We have also compared these devices in terms of their  $I_{\text{ON}}/I_{\text{OFF}}$  ratio. In III-V FETs, it is important to include the gate leakage current in the OFF current. We denote this combined current by  $I_{\text{LEAK}}$  [19]. We define  $I_{\text{LEAK}}$  to be the average of the current through the device in the two cases when the following are observed: 1) The gate is OFF and the drain ON ( $I_{\text{OFF}}$ ), and 2) the gate is ON and the drain OFF ( $I_{g,\text{ON}}$ )

$$I_{\text{LEAK}} = \frac{1}{2}(I_{\text{OFF}} + I_{g,\text{ON}}). \quad (1)$$

As in [20],  $I_{\text{OFF}}$  is defined at  $V_{\text{gs}} = V_T - 1/3V_{\text{dd}}$  ( $V_{\text{dd}} = 0.5 \text{ V}$ ), and  $I_{g,\text{ON}}$  and  $I_{\text{ON}}$  are defined at  $V_{\text{gs}} = V_T + 2/3V_{\text{dd}}$ , with  $V_{\text{dd}} = 0.5$  and  $0.05 \text{ V}$ , respectively. The  $I_{\text{ON}}/I_{\text{LEAK}}$  ratios for the 5- and 11-nm- $t_{\text{ins}}$  devices are then  $1.8 \times 10^3$  and  $6.3 \times 10^4$ , respectively. The 5-nm- $t_{\text{ins}}$  device is in enhancement mode, and at  $V_{\text{gs}} = V_T + 2/3V_{\text{dd}}$ , the gate is strongly forward biased.  $I_g$  increases exponentially in the forward-bias regime, resulting in a significantly higher value of  $I_{g,\text{ON}}$  for the 5-nm- $t_{\text{ins}}$  device compared to that for the 11-nm- $t_{\text{ins}}$  device (Fig. 4). This decrease in  $I_{\text{ON}}/I_{\text{LEAK}}$  ratio is a tradeoff that has to be taken into consideration when thinning the insulator to achieve enhancement-mode devices. Development of a gate-dielectric technology for III-V quantum-well MOSFET implementations is expected in the long run to address these gate leakage issues.

#### B. Source Resistance

The source resistance ( $R_s$ ) was measured directly from devices using the gate-current injection method [21], resulting in values of 220 and  $235 \Omega \cdot \mu\text{m}$  for the  $t_{\text{ins}} = 11$ - and 5-nm samples, respectively. The relatively small change in  $R_s$  as the barrier is thinned is in contrast to the results obtained by Kim and del Alamo [22], where  $R_s$  increased as the barrier was thinned by means of a wet-etch process. We attribute the better  $R_s$  results achieved in this paper to the fact that the dry etch used to thin the barrier is anisotropic in nature.

To further understand the constituent components of  $R_s$ , we adopt a simple model of the source resistance for our devices, as shown in Fig. 6, along with an estimate for their values, as

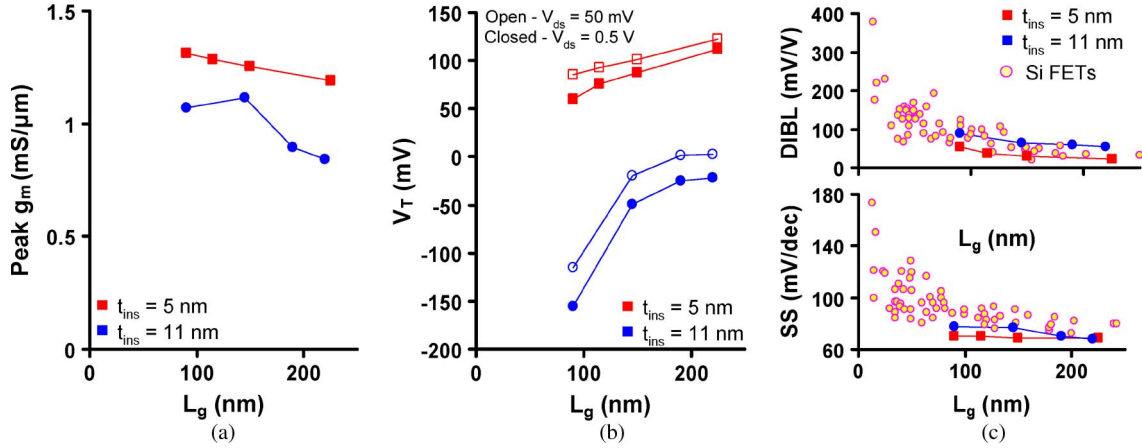


Fig. 5. Scaling behavior for the self-aligned devices with  $t_{\text{ins}} = 11$  and 5 nm. Thinning of the insulator results in an improvement in DIBL, SS, and  $V_T$  roll-off. The  $t_{\text{ins}} = 5$ -nm devices show outstanding short-channel effects when compared to Si CMOS devices.

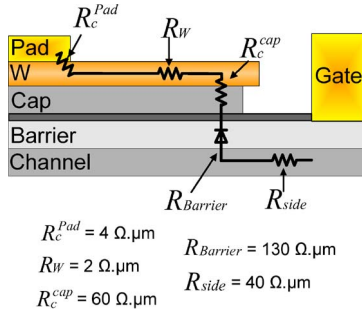


Fig. 6. Schematic of the component elements of the source resistance in the self-aligned device (with a nominal  $t_{\text{ins}}$  of 11 nm) and their extracted values from TLM measurements. As the contacts are nonalloyed, the current must flow from the contact through the heterobarrier and then into the channel. Based on a 1- $\mu\text{m}$  spacing between the gate and the source pad and 200-nm  $L_{\text{side}}$ , the total  $R_s$  is estimated to be 230  $\Omega \cdot \mu\text{m}$  from TLM measurements.

determined from the TLM measurements described in [8]. The sum of these components results in an estimated  $R_s$  of 230  $\Omega \cdot \mu\text{m}$ , which is consistent with the measured  $R_s$  values.

From this analysis, we find that  $R_{\text{Barrier}}$  is the dominant component of  $R_s$  (Fig. 6), accounting for more than 50% of the total value.  $R_c^{\text{cap}}$  is the next biggest component comprising 27%. As would be expected from a self-aligned design, the resistances of the W layer ( $R_W$ ) and the pad contact ( $R_c^{\text{pad}}$ ) are negligible. Known methods to reduce the barrier resistance, such as using a double-delta-doping layer in the barrier [23] or further engineering of the cap layers [24], should result in further reductions in  $R_s$ .

### C. Source-Resistance Scaling

In the analysis of devices and TLMs that was carried out previously, we assumed a semi-infinite contact length ( $\sim 10 \mu\text{m}$ ). For typical power and microwave applications of HEMTs, this is a valid assumption. However, for a III-V FET device to be used in a VLSI environment, this is not the case. If we assume that the maximum allowable contact length ( $L_c$ ) is the device half pitch that is less half the gate length, then, for the 15-nm node, a maximum available  $L_c$  of only about 10 nm must be assumed [3]. In the following, we project the value of

series resistance that is to be expected at these realistic contact dimensions.

To estimate how the  $R_s$  of self-aligned devices will scale as extrinsic device dimensions shrink, we must use a more sophisticated model for  $R_s$  than the simple lumped case previously discussed. To this end, we use a modified version of the distributed trilayer TLM model previously proposed by Reeves [25] (Fig. 7). Our structure technically comprises four layers—the extrinsic channel, the barrier, the cap, and the W metal. We can reduce it to the trilayer case by assuming that the first-order current conduction through the barrier layer is purely vertical. This is a good assumption because this is the highest resistive layer in our structure. The specific tunneling resistance from the extrinsic channel to the cap ( $\rho_t$ ) is then treated as a contact resistance between the two layers, with W being the third layer. We also assume that the sheet resistance of W (measured at 2  $\Omega/\text{sq}$ ) can be considered negligible. The key parameters required for the analysis are the sheet resistance of the cap layer ( $R_{\text{Scap}}$ ) and the extrinsic channel region under the cap ( $R_{\text{Sext}}$ ),  $\rho_t$ , the specific contact resistance between the cap and W ( $\rho_c$ ), and the contact length  $L_c$ . In the work of Reeves, the  $R_c$  of the structure is expressed as

$$R_c = K \cdot \left( \frac{f(R_{\text{Sext}} - \rho_t a^2) - (1-f)R_{\text{Scap}}}{b \tanh(bL_c)} - \frac{f(R_{\text{Sext}} - \rho_t b^2) - (1-f)R_{\text{Scap}}}{a \tanh(aL_c)} \right) \quad (2)$$

where

$$K = R_{\text{Sext}} / [\rho_t \cdot w \cdot (b^2 - a^2)] \quad (2.a)$$

$$a = \sqrt{\left[ \left\{ c - (c^2 - 4R_{\text{Sext}}R_{\text{Scap}}/(\rho_t\rho_c))^{\frac{1}{2}} \right\} / 2 \right]} \quad (2.b)$$

$$b = \sqrt{\left[ \left\{ c + (c^2 - 4R_{\text{Sext}}R_{\text{Scap}}/(\rho_t\rho_c))^{\frac{1}{2}} \right\} / 2 \right]} \quad (2.c)$$

$$c = [(R_{\text{Scap}} + R_{\text{Sext}})/\rho_t + R_{\text{Sext}}/\rho_c] \quad (2.d)$$

and  $w$  is the device width.

It should be noted that the resistance value in (2) contains only a portion of the source directly under the contact and does not include the  $R_{\text{Side}}$  portion of  $R_s$ .

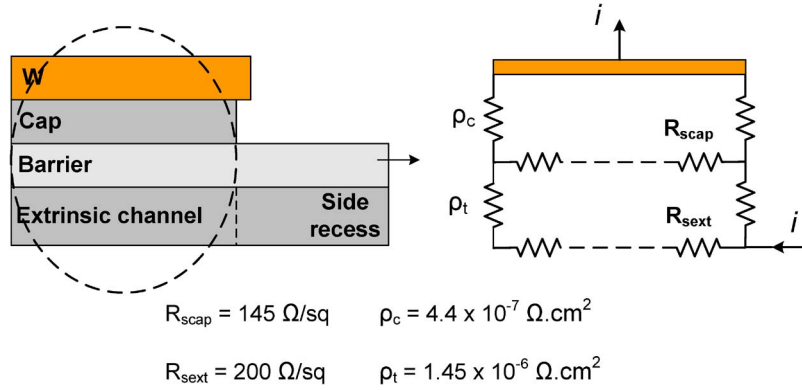


Fig. 7. Cross section of the nonalloyed ohmic contact and its corresponding trilayer TLM electrical network representation. We assume that current conduction in the barrier layer is only in the vertical direction and that the associated specific tunneling resistance  $\rho_t$  approximates a contact resistance between the extrinsic channel and cap layers. In this way, the nonalloyed contact reduces to a trilayer structure where all the current enters the contact through the extrinsic channel layer ( $f = 1$  [13]).

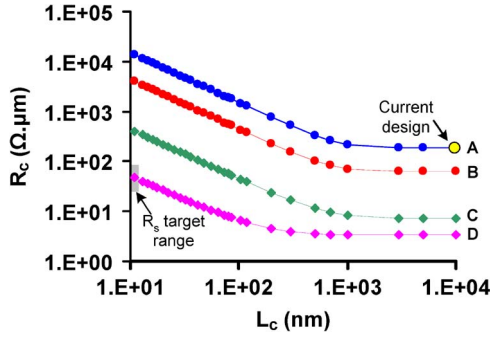


Fig. 8. Effect of contact-length scaling on  $R_c$ , as predicted by the use of trilayer approximation. (A) Current design. (B) Assumes that the barrier layer can be eliminated by use of a high- $k$  dielectric solution, i.e.,  $\rho_t \rightarrow 0$ . (C) Indium contact layer is increased to 70%.  $R_{scap} \rightarrow 14 \Omega/\text{sq}$  and  $\rho_c \rightarrow 4.4 \times 10^{-8} \Omega \cdot \text{cm}^2$ . (D) Cap is graded to InAs ( $\rho_c \rightarrow 5 \times 10^{-9} \Omega \cdot \text{cm}^2$ ).

The original analysis by Reeves considered an alloyed contact where current would enter the contact either through the alloyed layer sidewall or directly under the contact via  $R_{S\text{ext}}$ . The fraction of the current that enters in a conventional TLM manner through  $R_{S\text{ext}}$  is defined as  $f$  in (2). In the case of a nonalloyed ohmic contact, all the current enters the contact through  $R_{S\text{ext}}$  and  $f = 1$ . This results in simplification of (2) to

$$R_c = \frac{R_{s\text{ext}}}{w \cdot \rho_t (b^2 - a^2)} \left( \frac{R_{s\text{ext}} - \rho_t a^2}{b \tanh(bL_c)} - \frac{R_{s\text{ext}} - \rho_t b^2}{a \tanh(aL_c)} \right). \quad (3)$$

$R_{Scap}$  and  $\rho_c$  can be estimated directly from TLM test structures giving values of  $145 \Omega/\text{sq}$  and  $4.4 \times 10^{-7} \Omega \cdot \text{cm}^2$ , respectively. The sheet resistance of the extrinsic channel under the cap ( $R_{S\text{ext}}$ ) is different from that of  $R_{S\text{side}}$ , and we estimate it, by using a calibrated 1-D simulation of the heterostructure using nextnano, a self-consistent Schrödinger–Poisson solver [26]. This gives a value for  $R_{\text{ext}}$  of  $200 \Omega/\text{sq}$ . A value for  $\rho_t$  of  $1.45 \times 10^{-6} \Omega \cdot \text{cm}^2$  is estimated by forcing the modeled  $R_c$  to match the measured one from the TLM structures.

Using these values, we use the model to predict the expected evolution of  $R_c$  as  $L_c$  is scaled to more realistic values (Fig. 8). From the 2007 ITRS roadmap, the maximum allowable parasitic source-plus-drain resistance ( $R_{\text{sd}}$ ) at the 15-nm node is

145 or  $72.5 \Omega \cdot \mu\text{m}$  for  $R_s$ . This  $72.5\text{-}\Omega \cdot \mu\text{m}$  value could be considered an upper bound for the required  $R_s$ . For a lower bound, we require that the extrinsic device resistance results in no more than 20% decrease in  $I_{\text{on}}$ . To estimate the  $R_s$  needed to meet this requirement, we use the model of Antoniadis [27], where the width-normalized transistor current in saturation is defined as

$$I_D/W = C'_{\text{inv}}(V_{\text{GS}} - V_T)v \quad (4)$$

where  $C'_{\text{inv}}$  is the gate capacitance per unit area and  $v$  is the “effective” carrier velocity that can be expressed in terms of the virtual-source velocity ( $v_{x0}$ ), source resistance ( $R_s$ ), DIBL ( $\delta$ ), and  $C'_{\text{inv}}$  as

$$v = \frac{v_{x0}}{[1 + C'_{\text{inv}}R_sW(1 + 2\delta)v_{x0}]}. \quad (5)$$

Assuming an aggressive estimate for  $v_{x0}$  of  $2 \times 10^7$  cm/s,  $C'_{\text{inv}}$  of  $3 \times 10^{-6}$  F/cm corresponding to an effective oxide thickness (EOT) of about 1 nm, and DIBL of 100 mV/V,  $R_s$  should not exceed  $35 \Omega \cdot \mu\text{m}$  to limit the degradation of effective velocity ( $v$ ) (and, thus,  $I_{\text{on}}$ ) to less than 20%.

It is obvious from Fig. 8 curve A that, for our current technology, as  $L_c$  is scaled to 10 nm,  $R_c$  blows up and is over two orders of magnitude higher than the desired target range.

We have also used the model to evaluate different scenarios to obtain the desired contact resistance. We found that just changing  $R_{Scap}$  and  $\rho_c$  has very little impact on  $R_c$  as the barrier resistance represented by  $\rho_t$  dominates. We find that the only way to meet the desired target is to eliminate the barrier altogether. This implies the use of high- $k$ -based gate-dielectric solution. Eliminating the barrier has the effect of reducing  $R_c$  by about a factor of three, but it is still far away from the target region (line B in Fig. 8). Next, we assume the use of a cap with 70% InAs composition. This should theoretically pull the Fermi level into the conduction band, thus resulting in a very low sheet resistance and a nearly ideal ohmic contact. Shinohara *et al.* [28] have reported an  $R_{\text{Sheet}}$  of  $7 \Omega/\text{sq}$  using such a layer (about a factor-of-ten improvement over our design), and we similarly estimate a factor-of-ten reduction in  $\rho_c$ .



Again, a significant but not large-enough improvement in  $R_c$  is predicted (line C in Fig. 8).

Crook *et al.* [29] recently demonstrated a  $\rho_c$  of less than  $1 \times 10^{-8} \Omega \cdot \text{cm}^2$  by using a highly doped ( $3.5 \times 10^{19} \text{-cm}^{-3}$ )  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  cap and paying close attention to the surface preparation before contact formation. However, the lowest demonstrated value of  $\rho_c$  has been by Nittono *et al.* [30], who achieved a value of  $5 \times 10^{-9} \Omega \cdot \text{cm}^2$  by grading the contact layer to pure InAs. Grading the contact layer eliminates the internal potential barrier in the contact, thus reducing the contact resistance. This approach was originally for an HBT emitter, but such a cap design is possible in FETs. Assuming this demonstrated value of  $5 \times 10^{-9} \Omega \cdot \text{cm}^2$  for the contact resistance,  $R_c$  is reduced to about  $50 \Omega \cdot \mu\text{m}$  for an  $L_c$  of 10 nm (line D in Fig. 8), which falls within the target range. This analysis demonstrates that an aggressive approach to the extrinsic design of the device is needed to achieve the required performance of a future logic III–V FET with a self-aligned architecture.

#### IV. CONCLUSION

In conclusion, we have developed a self-aligned process that has allowed us to highlight the challenges and technical obstacles that need to be overcome in designing the extrinsic regions of future III–V devices. We have demonstrated 90-nm-gate-length InGaAs HEMTs with ohmic-gate–source separation of 60 nm, which is nearly a  $20\times$  reduction over that of conventional architectures. We have also obtained devices in which this distance is as small as 30 nm. Using this process, we have fabricated enhancement-mode devices with an InAlAs barrier that was thinned to 5 nm by means of a dry etch. The dry etch did not degrade the device performance, and  $R_s$  did not increase significantly while yielding devices with excellent electrostatic integrity. Enhancement-mode devices of 90-nm gate length show outstanding logic characteristics in terms of current,  $g_m$ , and short-channel effects. We have examined the future scaling potential of devices and determined that, for the intrinsic device, the  $I_{\text{on}}/I_{\text{off}}$  ratio as the insulator is thinned and, for the extrinsic device, the scaling of the source resistance with the contact length are the main limiting factors. Both issues can be addressed by introducing a high- $k$  gate-dielectric insulator in the intrinsic gate stack.

#### ACKNOWLEDGMENT

The authors would like to thank D. Antoniadis for the helpful discussions. Epitaxial heterostructures were grown by MBE Technology, Singapore. The devices were fabricated at the Microsystems Technology Laboratories and the Scanning-Electron-Beam Lithography Facility of MIT.

#### REFERENCES

- [1] D.-H. Kim and J. A. del Alamo, "Scaling behavior of  $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$  HEMTs for logic," in *IEDM Tech. Dig.*, 2006, pp. 837–840.
- [2] D.-H. Kim, J. A. del Alamo, J. H. Lee, and K. S. Seo, "Performance evaluation of 50 nm  $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$  HEMTs for beyond-CMOS logic applications," in *IEDM Tech. Dig.*, 2005, pp. 767–770.
- [3] International Technology Roadmap for Semiconductors. [Online]. Available: <http://public.itrs.net>
- [4] K. Nishii, M. Nishitsuji, T. Uda, T. Yokoyama, S. Yamamoto, T. Kunihisa, and A. Tamura, "A novel high-performance WSi-gate self-aligned N-AlGaAs/InGaAs/N-AlGaAs pseudomorphic double heterojunction MODFET by ion implantation," in *Proc. IEEE Int. Symp. Compound Semicond.*, 1998, pp. 487–490.
- [5] J. H. Huang, J. K. Abrokwah, and W. J. Ooms, "Nonalloyed InGaAs/GaAs ohmic contacts for self-aligned ion implanted GaAs heterostructure field effect transistors," *Appl. Phys. Lett.*, vol. 61, no. 20, pp. 2455–2457, Nov. 16, 1992.
- [6] U. K. Mishra, A. S. Brown, L. M. Jelloian, M. Thompson, L. D. Nguyen, and S. E. Rosenbaum, "Novel high performance self-aligned 0.15 micron long T-gate AlInAs–GaInAs HEMTs," in *IEDM Tech. Dig.*, 1989, pp. 101–104.
- [7] D. A. J. Moran, H. McLelland, K. Elgaid, G. Whyte, C. R. Stanley, and I. Thayne, "50-nm self-aligned and "standard" T-gate InP pHEMT comparison: The influence of parasitics on performance at the 50-nm node," *IEEE Trans. Electron Devices*, vol. 53, no. 12, pp. 2920–2925, Dec. 2006.
- [8] N. Waldron, D.-H. Kim, and J. A. del Alamo, "90 nm self-aligned enhancement-mode InGaAs HEMT for logic applications," in *IEDM Tech. Dig.*, 2007, pp. 633–636.
- [9] D.-H. Kim and J. A. del Alamo, "Beyond CMOS: Logic suitability of  $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$  HEMT," in *Proc. CS MANTECH*, 2006, pp. 251–254.
- [10] H. Matsuzaki, T. Maruyama, T. Koasugi, H. Takahashi, M. Tokumitsu, and T. Enoki, "Lateral scale down of InGaAs/InAs composite-channel HEMTs with tungsten-based tiered Ohmic structure for 2-S/mm  $g_m$  and 500-GHz  $f_T$ ," *IEEE Trans. Electron Devices*, vol. 54, no. 3, pp. 378–384, Mar. 2007.
- [11] N. Yoshida, Y. Yamamoto, H. Takano, T. Sonoda, S. Takamiya, and S. Mitsui, "Alloyed and non-alloyed ohmic contacts for AlInAs/InGaAs high electron mobility transistors," *Jpn. J. Appl. Phys.*, vol. 33, no. 6A, pp. 3373–3376, Jun. 1994.
- [12] T. Suemitsu, H. Yokoyama, T. Ishii, T. Enoki, G. Meneghesso, and E. Zanoni, "30-nm two-step recess gate InP-based InAlAs/InGaAs HEMTs," *IEEE Trans. Electron Devices*, vol. 49, no. 10, pp. 1694–1700, Oct. 2002.
- [13] S. J. Pearton, F. Ren, and C. R. Abernathy, "Low bias dry etching of tungsten and dielectric layers on GaAs," *Semicond. Sci. Technol.*, vol. 8, no. 10, pp. 1897–1903, Oct. 1993.
- [14] T. Suemitsu, H. Yokoyama, Y. Umeda, T. Enoki, and Y. Ishii, "High-performance 0.1  $\mu\text{m}$  gate enhancement-mode InAlAs/InGaAs HEMTs using two-step recessed gate technology," *IEEE Trans. Electron Devices*, vol. 46, no. 6, pp. 1074–1080, Jun. 1999.
- [15] D.-H. Kim, J. A. del Alamo, J.-H. Lee, and K.-K. Seo, "The impact of side-recess spacing on the logic performance of 50 nm  $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$  HEMTs," in *Proc. IPRM*, 2006, pp. 177–180.
- [16] T. Suemitsu, "Study of sub-0.1- $\mu\text{m}$  gate InP-based high electron mobility transistors." Ph.D. dissertation, Waseda Univ., Tokyo, Japan, 2000.
- [17] K. Shinohara, Y. Yamashita, A. Endoh, K. Hikosaka, T. Matsui, S. Hiyamizu, and T. Mimura, "Importance of gate-recess structure to the cutoff frequency of ultra-high-speed InGaAs/InAlAs HEMTs," in *Proc. IPRM*, 2002, pp. 451–454.
- [18] D.-H. Kim, J. A. del Alamo, J.-H. Lee, and K.-S. Seo, "Logic suitability of 50-nm  $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$  HEMTs for beyond-CMOS applications," *IEEE Trans. Electron Devices*, vol. 54, no. 10, pp. 2606–2613, Oct. 2007.
- [19] C.-H. Jan, P. Bai, J. Choi, G. Curello, S. Jacobs, J. Jeong, K. Johnson, D. Jones, S. Klopocic, J. Lin, N. Lindert, A. Lio, S. Natarajan, J. Neiryneck, P. Packan, J. Park, I. Post, M. Patel, S. Ramey, P. Reese, L. Rockford, A. Roskowski, G. Sacks, B. Turkot, Y. Wang, L. Wei, J. Yip, I. Young, K. Zhang, Y. Zhang, M. Bohr, and B. Holt, "A 65 nm ultra low power logic platform technology using uni-axial strained silicon transistors," in *IEDM Tech. Dig.*, 2005, pp. 60–63.
- [20] R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, A. Majumdar, M. Metz, and M. Radosavljevic, "Benchmarking nanotechnology for high performance and low-power logic transistor applications," *IEEE Trans. Nanotechnol.*, vol. 4, no. 2, pp. 153–158, Mar. 2005.
- [21] W. J. Azzam and J. A. del Alamo, "An all-electrical floating-gate transmission line model technique for measuring source resistance in heterostructure field-effect transistors," *IEEE Trans. Electron Devices*, vol. 37, no. 9, pp. 2105–2107, Sep. 1990.
- [22] D. H. Kim and J. A. del Alamo, "Lateral and vertical scaling of  $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$  HEMTs for post-Si-CMOS logic applications," *IEEE Trans. Electron Devices*, vol. 55, no. 10, pp. 2546–2553, Oct. 2008.
- [23] D. A. J. Moran, E. Boyd, K. Elgaid, F. McEwan, H. McLelland, C. R. Stanley, and I. G. Thayne, "Self-aligned T-gate InP HEMT realisation through double delta doping and a non-annealed ohmic process," *Microelectron. Eng.*, vol. 73/74, pp. 814–817, Jun. 2004.

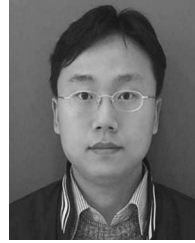


- [24] T. Enoki, T. Kobayashi, and Y. Ishii, "Device technologies for InP-based HEMTs and their application to ICs," in *GaAs IC Symp. Tech. Dig.*, 1994, pp. 337–340.
- [25] G. K. Reeves and B. Harrison, "An analytical model for alloyed ohmic contacts using a trilayer transmission line model," *IEEE Trans. Electron Devices*, vol. 42, no. 8, pp. 1536–1547, Aug. 1995.
- [26] Nextnano Semiconductor Nanodevice Simulation Tool. [Online]. Available: <http://www.wsi.tum.de/nextnano3>
- [27] A. Khakifirooz and D. A. Antoniadis, "MOSFET performance scaling. Part I. Historical trends," *IEEE Trans. Electron Devices*, vol. 55, no. 6, pp. 1391–1400, Jun. 2008.
- [28] K. Shinohara, Y. Yamashita, A. Endoh, I. Watanabe, K. Hikosaka, T. Matsui, T. Mimura, and S. Hiyamizu, "547-GHz ft  $\text{In}_{0.7}\text{Ga}_{0.3}\text{As} - \text{In}_{0.52}\text{Al}_{0.48}\text{As}$  HEMTs with reduced source and drain resistance," *IEEE Electron Device Lett.*, vol. 25, no. 5, pp. 241–243, May 2004.
- [29] A. M. Crook, E. Lind, Z. Griffith, M. J. W. Rodell, J. D. Zimmerman, A. C. Gossard, and S. R. Bank, "Low resistance, nonalloyed ohmic contacts to InGaAs," *Appl. Phys. Lett.*, vol. 91, no. 19, p. 192 114, Nov. 2007.
- [30] T. Nittono, I. Hiroshi, O. Nakajima, and T. Ishibashi, "Non-alloyed ohmic contacts to n-GaAs using compositionally graded  $\text{In}_x\text{Ga}_{1-x}\text{As}$  layers," *Jpn. J. Appl. Phys.*, vol. 27, no. 9, pp. 1718–1722, Sep. 1988.



**Niamh Waldron** received the B.Sc. degree in applied physics from the University of Limerick, Limerick, Ireland, in 1996 and the M.Sc. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2002 and 2007, respectively. Her M.S. thesis work was on the use of strained-Si technology for RF power devices, and her Ph.D. research was on the use of InGaAs HEMT technology for logic applications.

From 1996 to 2000, she was an Implant and RTA Process Engineer with Analog Devices, Ireland, where she worked on BiCMOS development and integration. From 2007 to 2008, she was a Postdoctoral Associate with MIT. In 2008, she joined the Interuniversity Microelectronics Center, Leuven, Belgium, as a Postdoctoral Researcher, where she is currently working on the heterogeneous integration of III–V-based channel devices for advanced CMOS nodes.



**Dae-Hyun Kim** was born in Korea on November 13, 1974. He received the B.S. degree in electronics from Kyungpook National University, Daegu, Korea, in 1997 and the M.S. and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 2000 and 2004, respectively.

From 2004 to 2005, he was a Postdoctoral Associate with the Inter-University Semiconductor Research Center, Seoul National University. In 2005, he joined the Massachusetts Institute of Technology, Cambridge, where he actively participated in the research program entitled "Potential of III–V HFETs for post-Si logic applications." He is currently with Teledyne Scientific Company, Thousand Oaks, CA. He has a particular interest in InGaAs PHEMTs for terahertz applications.

Dr. Kim was the recipient of the Best Student Paper Award at ISDRS in 2003, the Best Student Paper Award at the Korean Conference on Semiconductors in 2004, and the Young Researcher Award at EUMW in 2004. He is currently a Reviewer for the IEEE ELECTRON DEVICE LETTERS.



**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, in 1983 and 1985, respectively.

From 1985 to 1988, he was with NTT LSI Laboratories, Atsugi, Japan. He was an NSF Presidential Young Investigator. Since 1988, he has been with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, where he is currently the Donner Professor of Electrical Engineering and Computer Science and a MacVicar Faculty Fellow. His current research interests are in microelectronics technologies for communications and logic processing. He has a particular interest in Si LDMOS, CMOS, GaAs PHEMTs, and GaN HEMTs for RF power applications and in InGaAs HEMTs as a beyond-the-roadmap semiconductor logic technology. He is also active in online laboratories for science and engineering education.

Prof. del Alamo has received several teaching awards at MIT: the Baker Award, the Edgerton Junior Faculty Achievement Award, the Smullin Award, and the Bose Award. He is a member of the Royal Spanish Academy of Engineering. He is currently an Editor for the IEEE ELECTRON DEVICE LETTERS.