InP DHBT Analytical Modeling: Toward THz Transistors

Nil Davy, *Member, IEEE*, Virginie Nodjiadjim, *Member, IEEE*, Muriel Riet, Colin Mismer, Marina Deng, *Member, IEEE*, Chhandak Mukherjee, *Member, IEEE*, Bertrand Ardouin, and Cristell Maneux, *Member, IEEE*

***Abstract*—InP double heterojunction bipolar transistors (InP DHBTs) are one of the key technologies considered for terahertz (THz) applications. The improvement of their frequency performance is challenging and strongly dependent on various parameters (manufacturing process, geometry, and epitaxial structure). In this article, a novel method is developed to take into account these parameters and predict the frequency performance of the technology. This approach consists of rebuilding the S-parameter matrix of the small-signal model. Elements of the small signal model are identified, and their assessment is described in detail. Once calibrated with the present state-of-the-art device features, the model shows a good agreement with the measurements. Based on this result, analysis of the emitter and base technological features are performed along with optimizations of the vertical structure. Finally, the necessary optimizations for developing a THz transistor are detailed. This works provides guidelines for technological improvement and opens the way for designing transistors operating at frequencies above a THz.**

***Index Terms*—Double heterojunction bipolar transistor (DHBT), indium phosphide, InP/InGaAs, modeling, terahertz (THz).**

I. Introduction

HIGH-FREQUENCY systems with terahertz (THz) operating frequencies are of growing interest, be it for wireless communication, optical transmission or for imaging, and various ultrahigh-speed transistor technologies have emerged to respond to the challenge. Among the potential candidates, InP-based technologies have demonstrated impressive cut-off frequencies above 1 THz for both high electron mobility transistors (HEMTs) [1] and double heterojunction bipolar transistors (DHBTs) [2], [3]. Due to their higher breakdown voltages as well as a better power handling capability, DHBTs are a prime candidate for the design of power amplifiers (PAs) operating above 100 GHz [4] as well as for 1.6-Tb/s optical transceivers [5], [6].

Manuscript received 4 October 2022; revised 6 January 2023 and 27 February 2023; accepted 3 March 2023. Date of publication 15 March 2023; date of current version 20 October 2023. This work was supported in part by ANR-FNS through ULTIMATE project under Grant ANR-16-CE93-0007, and in part by the European Commission through the Photonics Public Private Partnership Initiatives (Twilight Project) under Grant H2020-ICT-2019-2. This article was recommended by Associate Editor E. R. Keiter. *(Corresponding author: Nil Davy.)*

Nil Davy, Virginie Nodjiadjim, Muriel Riet, Colin Mismer, and Bertrand Ardouin are with III-V Lab, 91767 Palaiseau, France (e-mail: nil.davy@3-5lab.fr).

Marina Deng, Chhandak Mukherjee, and Cristell Maneux are with the Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400 Talence, France.

Digital Object Identifier 10.1109/TCAD.2023.3257706

The major challenge this technology faces is that one needs to improve the transition frequency f_T and the maximal oscillation frequency f_{MAX} simultaneously. These performance figures of merit are related to multiple parameters: the epitaxial structure (thickness, doping, and material), the geometry of the transistor and the technological features related to the process (contact resistivity, undercuts dimensions, etc.) Generally speaking, two different approaches are mainly used to study these technologies: technology computer-aided design (TCAD) and compact modeling (sometimes even a combination of both).

TCAD device simulations come in many forms, from the simple drift-diffusion model to the nonequilibrium Green's function formulation [7]. The choice of the simulation type is motivated by different factors: calculation time, desired precision, charge transport, mechanism under study, etc. TCAD simulation has the advantage of capturing the physical behavior of the device in-depth and, thus, helps to study different physical phenomena in detail (tunneling, recombination, etc.) For example, the main feature to capture when simulating III-V DHBTs is the presence of multiple energy levels of conduction bands leading to quite different electron mobility values [8], [9], [10].

On the other hand, compact models are used in device libraries for circuit simulation and need to accurately model the device behavior at low computational cost. Most of the compact models developed are based on physical equations [11]. Their use in conjunction with scaling rules have already been demonstrated for the optimization of InP DHBT [12]. Even though they capture a vast range of mechanisms, these models are not very well suited for device optimization. In fact, they have been developed to capture the electrical behavior of various bipolar technologies and do not take into account the impact of the technological architecture of a device on its performances. This limits the compact models in terms of the number of geometrical parameters that can be studied for optimizing a technology (e.g., L_E and W_E in [12]).

In order to simplify the design flow, in this work we present an analytical geometry-dependent model, developed following long-term evolution of research on the topic [13], to predict frequency performance of InP DHBTs. This model has been developed in order to obtain the best frequency performance for a given epitaxial structure and then to provide quick feedback on the influence of geometrical parameters related to device

architecture for process improvement. Consequently, a small signal approach at the optimal bias point is used. Using this approach, the number of parameters to be adjusted can be reduced to a minimum. The operating principle of the model is the following: first the small-signal model is constructed based on the information on the device geometry, vertical structure, and typical process parameter (contact resistivity, undercut ...) values obtained at the end of the fabrication process and then frequency performances are estimated. With the help of this model, it is possible to study the frequency performance through a detailed analysis of both RC products and transit times for various geometries and technological features. However, some mechanisms, such as self-heating or Kirk-effect were not incorporated into the model to avoid additional complexity and will be addressed in future iterations. The remainder of this article is organized as follows. Section II details the studied InP DHBT technology. The model equations are detailed in Section III. In Section IV, model validation against the measurements obtained with the present technology featuring various geometries is demonstrated. From these results, the influence of technological features such as undercut, contact resistivity is analyzed. The optimization of the vertical structure is also addressed. Next, a roadmap of the optimizations needed to reach THz frequencies is presented with attainable technological features. Finally, the conclusion follows.

II. Transistor Design and Technology

The model is based on the InP DHBT technology presented in [14]. The DHBTs are a self-aligned triple mesa wet etched technology. The emitter is passivated with SiN and the planarization is performed with polyimide. The intrinsic emitter is 40-nm thick. The GaInAs base is 28-nm thick, highly C-doped (8×10^{19} cm^{-3}) with a gradual composition from $Ga_{0.53}In_{0.47}As$ on the emitter side to $Ga_{0.47}In_{0.53}As$ on the collector side. The 130-nm thick composite collector contained an InGaAs spacer, an InP doping plane and a lightly doped InP part. Its Kirk current density is 6 mA/μm^2. Several transistor geometries are available: the emitter length L_E can vary between 3 and 10 μm, the emitter width W_E between 0.3 and 0.7 μm and the base contact width W_{EB} between 0.2 and 0.3 μm. Fig. 1 shows a 0.4×4 μm^2 InP DHBT before planarization. The epitaxial structure, for which the model was developed, is summarized in Fig. 2. The emitter contains one layer of InGaAs (not uniformly doped) and two layers of InP (a highly doped part and a lightly doped part which is the intrinsic emitter layer). The base is made of InGaAs and is compositionally graded. The composite collector is made of an InGaAs spacer as well as two InP layers (a doping plane and a lightly doped part). The subcollector is made of an InGaAs layer sandwiched between two InP layers.

III. Model Description

The predictive model consists of rebuilding the small-signal model of the device by calculating the values of all of its elements. To do this, the developed model needs the following information: vertical structure (thickness and doping), device geometry (emitter width and length, and base contact width)

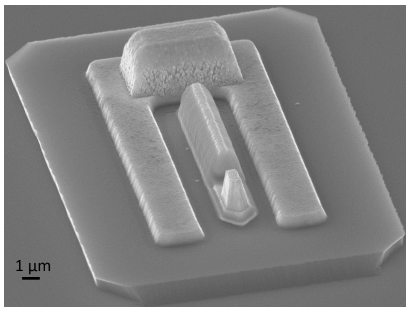

Fig. 1. SEM view of a 0.4×4 μm^2 InP DHBT before planarization.

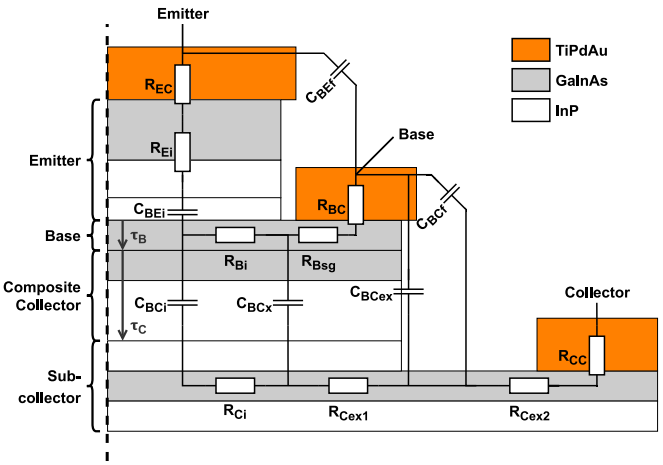

Fig. 2. Epitaxial structure and small signal elements (without the active part).

and relevant process parameter values (undercut and contact resistivity). In order to identify and understand the elements that need to be precisely modeled, the following equations should be looked at in detail:

$$f_T = \frac{1}{2\pi\left(\tau_B + \tau_C + \frac{nkT}{qI_C}(C_{BE} + C_{BC}) + (R_E + R_C)C_{BC}\right)} \quad (1)$$

$$f_{MAX} = \sqrt{\frac{f_T}{8\pi\,(R_{Bx} \cdot C_{BC} + R_{Bi} \cdot C_{BCi})}}. \quad (2)$$

Each of these elements is, in turn, made up of several contributions. Figs. 2 and 3 show these resistive and capacitive components schematically superimposed on the device structure. The relation between these contributions and (1) and (2) is detailed in the following paragraphs.

A. Emitter

The total emitter resistance R_E consists of two contributions: 1) the resistance due to the metal–semiconductor interface R_{EC} and 2) the resistance of the semiconductor itself R_{Ei}.

The contact resistance R_{EC} is directly linked to the contact resistivity through the following equation:

$$R_{EC} = \frac{\rho_E}{A_E} \quad (3)$$

where ρ_E is the contact resistivity and A_E is the area of the emitter semiconductor. The semiconductor resistance is given by the sum of the resistances of each undepleted semiconductor layer as well as the part of the intrinsic emitter which is

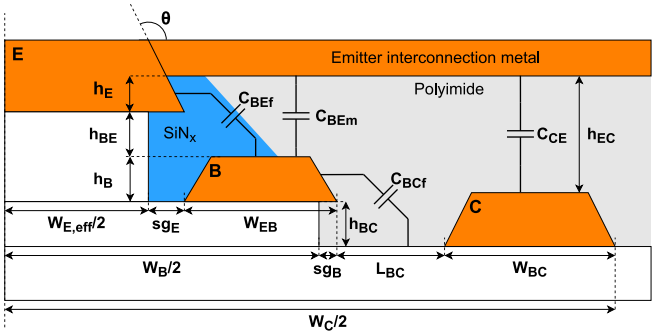

Fig. 3. Schematic semi-cross-sectional view with geometrical parameters and parasitic capacitances.

undepleted

$$R_{Ei} = \sum_{n=0}^{k} \frac{\rho_i \cdot e_i}{A_E} + \frac{\rho_{k+1}(e_{k+1} - T_{EB})}{A_E} \tag{4}$$

with ρ_i (resp., e_i) the resistivity (resp., the thickness) of the layer i and T_{EB} the depletion width of the emitter–base junction.

In addition to the resistance, the emitter contains three capacitances that make up C_{BE}: 1) the junction capacitance C_{BEi}; 2) a fringe capacitance C_{BEf}; and 3) a parasitic capacitance C_{BEm} induced by interconnections. The junction capacitance at $V_{BE} = 0$ V is expressed as

$$C_{BEj0} = \frac{\varepsilon_{InP} \cdot A_E}{\sqrt{\frac{2 \cdot \varepsilon_{InP} \cdot V_{DE}}{q \cdot N_D}}} \tag{5}$$

with V_{DE} being the diffusion voltage of the emitter–base junction. The denominator in (5) corresponds to the depletion depth in the emitter. In our case, the depletion depth is limited by the heavily doped InP layer, and, hence, the value in the denominator can be taken equal to the intrinsic emitter thickness. This capacitance is modulated by the voltage applied across the junction

$$C_{BEi}(V) = C_{BEj0}\left(1 - \frac{V}{V_{DE}}\right)^{-m_{je}}. \tag{6}$$

The grading coefficient m_{je} is set to 0.5 since the base–emitter junction is abrupt. As the theoretical expression tends to infinity when V_{BE} tends to the diffusion voltage V_{DE}, a coefficient FC (with values between 0 and 1) is introduced. For base–emitter voltage greater than $FC \times V_{DE}$, the new expression, taken from the Gummel–Poon model [15], has a linear trend

$$C_{BEi}(V) = C_{BEj0} \cdot (1 - FC)^{-m_{je}} \cdot \left[1 + \frac{m_{je}}{V_{DE} \cdot (1 - FC)} \cdot (V - FC \cdot V_{DE})\right]. \tag{7}$$

The FC coefficient depends on the emitter–base junction. For our technology, the typical values at $J_C = 6$ mA/μm^2 are: $FC = 0.78$, for $V_{DE} = 1$ V and $V_{BE} = 0.9 - 1$ V. The formulation of the fringe capacitance has been introduced in [16], in which the permittivity of SiN_x, the emitter passivation, is used and it is calculated as (see Fig. 3)

$$C_{BEf} = 2 \cdot L_E \frac{\varepsilon_{SiN_x}}{\theta} \ln\left(1 + \frac{h_E}{h_{BE}}\right) \tag{8}$$

where L_E is the emitter length and θ is the angle formed by the metallization. Finally, the capacitance C_{BEm} induced by interconnections is calculated as follows:

$$C_{BEm} = 2\frac{\varepsilon_{\text{polyimide}} \cdot L_E \cdot W_{EB}}{h_E + h_{BE}} \tag{9}$$

where symbols are the geometrical parameters shown in Fig. 3

B. Base

The base resistance contains three contributions: 1) the contact resistance R_{BC}; 2) the access resistance R_{Bsg} due to the distance between the base contact and the emitter that constitute the extrinsic base resistance R_{Bx}; and 3) the intrinsic base resistance R_{Bi}. The contact resistance R_{BC} is distributed along two directions: across the width and along the length of the emitter. An expression taking into account its distributed nature across the width has been introduced in [17]. Without the distribution of the resistance along the length of the emitter, R_{BC} would be inversely proportional to L_E which does not correspond to the observations. The distributed nature along the length of the emitter induces the consideration of the resistivity of the metal stack. It is then possible to write R_{BC} as

$$R_{BC} = \frac{1}{2}\sqrt{\frac{R_{Bm}}{G_B}} \coth\left(\sqrt{R_{Bm} G_B} L_E\right) \tag{10}$$

$$\text{with: } R_{Bm} = \frac{\rho_m}{W_{EB} \cdot h_B} \tag{11}$$

$$\text{and } \frac{1}{G_B} = R_{SB} \cdot L_{TB} \cdot \coth\frac{W_{EB} - sg_B}{L_{TB}}. \tag{12}$$

Here, ρ_m is the metal stack resistivity, R_{SB} is the sheet resistance of the base, and L_{TB} is the transfer length of the contact which is dependent of the contact resistivity ρ_{CB} and the sheet resistance, R_{SB}, defined as

$$L_{TB} = \sqrt{\frac{\rho_{CB}}{R_{SB}}}. \tag{13}$$

The above equation has been compared with the one introduced in [18] and a difference of less than 5% has been observed for transistors with $L_E = 5$ μm (and less than 8% for $L_E = 7$ μm) for the parameter values studied in the next section. The access resistance R_{Bsg} is the resistance induced by the distance between the base contact and the emitter. This resistance is expressed simply as follows:

$$R_{Bsg} = R_{SB} \cdot \frac{sg_E}{2 \cdot L_E}. \tag{14}$$

The third contribution is the intrinsic base resistance. The expression of this resistance at zero bias is developed in [19]

$$R_{Bi0} = R_{SB} \cdot \frac{W_E}{12 \cdot L_E}. \tag{15}$$

This resistance is modulated by the base current and a formulation of this effect is presented in [20]

$$R_{Bi} = R_{Bi0} \cdot \frac{\tan(z) - z}{z\tan^2(z)} \tag{16}$$

where the z coefficient is given by

$$z = \frac{\pi^2}{24}\frac{\sqrt{1+144\frac{I_B}{\pi^2 I_{RB}}}-1}{\sqrt{\frac{I_B}{I_{RB}}}} \tag{17}$$

with I_{RB} being the current at which the intrinsic base resistance becomes half of its zero bias value, R_{Bi0}.

The modeling of the base region also includes the computation of the transit time. The transit time calculation in the case of a linearly graded base has been introduced in [21]. A formula that takes into account the finite exit velocity is given in [22]

$$\tau_B = \frac{T_B}{v_{\text{exit}}}\left(\frac{kT}{\Delta E_C}\right)\left(1-e^{-\Delta E_C/kT}\right) + \frac{{T_B}^2}{D_n}\left(\frac{kT}{\Delta E_C}\right) - \frac{{T_B}^2}{D_n}\left(\frac{kT}{\Delta E_C}\right)^2\left(1-e^{-\Delta E_C/kT}\right) \tag{18}$$

with T_B being the base thickness, v_{exit} being the finite exit velocity, and D_n being the electron diffusion coefficient. The exit velocity is considered to be equal to the thermal speed by the following expression:

$$v_{\text{exit}} = \sqrt{\frac{2kT}{\pi m_e}}. \tag{19}$$

This transit time depends on the conduction band offset ΔE_C in the base. When ΔE_C is induced by a gradual doping and composition, it can be expressed as follows:

$$\Delta E_C = \Delta E_{C,\text{comp}} + \Delta E_{C,\text{dop}} - \Delta E_{C,BGN} \tag{20}$$

with $\Delta E_{C,\text{comp}}$ being the offset induced by composition grading, $\Delta E_{C,\text{dop}}$ the offset induced by doping grading, and $\Delta E_{C,BGN}$ the offset due to band gap narrowing. $\Delta E_{C,\text{comp}}$ is simply computed from the electron affinity as follows:

$$\Delta E_{C,\text{comp}} = \chi_C - \chi_E \tag{21}$$

with χ_C (resp., χ_E) being the electron affinity on the collector (resp., emitter) side. If the base is too heavily doped, the semiconductor becomes degenerate. To take this into account, the Joyce–Dixon approximation is used in the calculation of $\Delta E_{C,\text{dop}}$

$$\Delta E_{C,\text{dop}} = k_B T\left[\ln\frac{N_{ac}}{N_{ae}} + \frac{1}{\sqrt{8}}\left(\frac{N_{ac}-N_{ae}}{N_v}\right)\right]. \tag{22}$$

As doping varies along the base, band gap narrowing also varies. This induces $\Delta E_{C,BGN}$, which opposes $\Delta E_{C,\text{dop}}$. A model presented in [23] describes the distribution of the band gap narrowing between E_V and E_C. $\Delta E_{C,BGN}$

$$\Delta E_{C,BGN} = C_3\left(\frac{N_{ac}-N_{ae}}{10^{18}}\right)^{\frac{1}{4}} + C_4\left(\frac{N_{ac}-N_{ae}}{10^{18}}\right)^{\frac{1}{2}}. \tag{23}$$

In the case of GaInAs, $C_3 = 0.0113$ and $C_4 = 2.2988\text{e-}4$.

C. Collector and Subcollector

Resistances of the collector and subcollector are expressed using formulas analogous to those used for the base. We, therefore, obtain the following expressions for the resistances that constitute R_C:

$$R_{Ci} = R_{SC}\frac{W_E}{12\cdot L_E} \tag{24}$$

$$R_{Cex1} = R_{SC}\frac{sg_E + W_{EB} - sg_B}{2\cdot L_E} \tag{25}$$

$$R_{Cex2} = R_{SC}\frac{sg_B + L_{BC}}{2\cdot L_E} \tag{26}$$

$$R_{CC} = \frac{1}{2}\sqrt{\frac{R_{Cm}}{G_C}}\coth\left(\sqrt{R_{Cm}G_C}L_E\right). \tag{27}$$

Four capacitances are associated with the collector and the subcollector. The capacitance C_{BC} is made up of three contributions: 1) the junction capacitance C_{BCj} which splits into C_{BCi} and C_{BCx}; 2) the capacitance C_{BCex} due to the mesa undercut; and 3) the fringe capacitance C_{BCf}. The last capacitance is the parasitic capacitance C_{CE} due to interconnections. As the base–collector junction is made of different layers with different materials, the computation of the junction capacitance is more complex than the base–emitter junction capacitance. A method has been introduced in [24] to find the depletion width and compute the capacitance C_{BCj}. The junction capacitance C_{BCx} is then computed as

$$C_{BCx} = C_{BCj}\left(1-\frac{A_E}{A_C}\right). \tag{28}$$

The intrinsic base–collector junction capacitance C_{BCi} is modulated by the collector current due to the fact that the transit time is dependent of the current I_C. An equation has been introduced in [25] to take this effect into account as

$$C_{BCi} = \frac{A_E}{A_C}C_{BCj} - k_1\frac{I_C}{2}\left(1-\frac{I_C}{I_{TC}}\right) \tag{29}$$

where k_1 and I_{TC} are the coefficients that are extracted from measurement. The capacitance C_{BCex} is formed due to the undercut of the base mesa and is simply expressed as follows:

$$C_{BCex} = 2\cdot\varepsilon_{\text{polyimide}}\frac{sg_B\cdot L_E}{h_{EC}}. \tag{30}$$

The fringe capacitance C_{BCf} is computed with an equation similar to (8) and the parasitic capacitance C_{CE} is computed similar to C_{BEm}.

Unlike the base transit time, there is no analytical equation to compute the collector transit time. As collector transit time simulations are quite complex [10], a simple approximation is used in the model. As the model focuses on maximal frequency performance, the electron velocity is assumed to be constant and optimal in the collector. Thus, the collector transit time depends only on the collector thickness

$$\tau_C = \frac{T_C}{2\cdot v_{\text{avg}}} \tag{31}$$

with T_C being the collector thickness and v_{avg} the average electron velocity. This velocity is determined from the transit

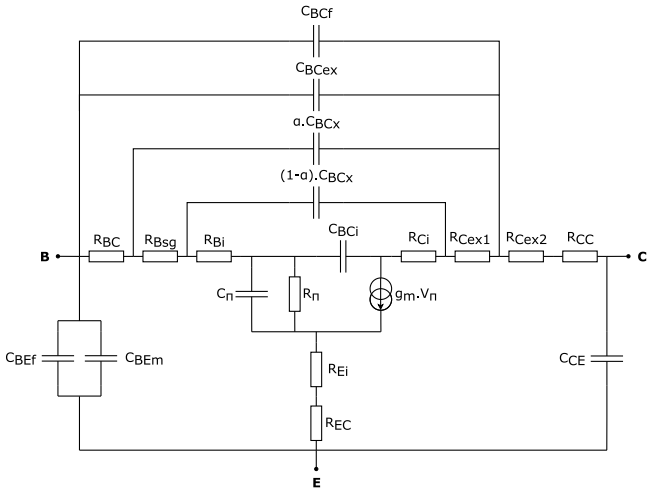

Fig. 4. Small signal model with computed elements.

time corresponding to the maximum f_T obtained from the measurement. Typically, the order of magnitude of this velocity is 3×10^7 cm/s.

D. Small Signal Model

In order to evaluate the frequency performance of the DHBTs under test, a small signal model using the elements detailed above is built. More specifically, the S-parameter matrix of the small signal model, shown in Fig. 4, is computed.

In addition to previously computed elements, which are only passive elements, four other parameters are calculated in order to take into account the behavior of the active transistor region

$$g_{m0} = \frac{qI_C}{nk_BT} \tag{32}$$

$$g_m = g_{m0} \cdot \exp(-j\omega\tau) \tag{33}$$

$$R_\pi = \frac{\beta}{g_{m0}} \tag{34}$$

$$C_\pi = C_{BE} + g_{m0} \cdot \tau \tag{35}$$

where n is the ideality coefficient of the base–emitter junction, τ is the total transit time ($\tau_B + \tau_C$) and β is the static current gain. An additional fitting parameter α, adjusted on measurements, has been introduced in order to take the distributed behavior of C_{BCx} into account and, thus, model f_{MAX} more accurately. The ideality coefficient n is directly obtained from measurements and it is set to a typical value obtained across several wafers. The static current gain β is determined from the following equation:

$$\beta = \frac{\tau_B}{\tau_{nB}} \tag{36}$$

with τ_{nB} being the electron lifetime in the base, determined from typical values obtained from measurements across multiple wafers.

IV. Results and Discussion

A. Comparison With Measurements

In order to validate our developed model, comparisons with measurements have been performed for different emitter widths and lengths. S-parameter measurements were performed using an Anritsu Vectorstar analyzer up to 110 GHz using on-wafer ALRM calibration and open-short de-embedding. Modeling and measurement are performed at $J_C = 6$ mA/μm^2 and $V_{CE} = 1.6$ V as this bias point allows the best compromise between bias points corresponding to the maximum f_T and f_{MAX}. In order to best fit the measurement the model needs to be calibrated with accurate parameter values. The real dimensions of the fabricated devices are extracted from SEM measurements while contact resistivities are extracted from TLM measurements. In our case the contact resistivity values are: $\rho_E = 4\ \Omega/\mu\mathrm{m}^2$, $\rho_B = 10\ \Omega/\mu\mathrm{m}^2$, and $\rho_C = 10\ \Omega/\mu\mathrm{m}^2$. Also, the following six parameters are tuned to further improved the model fitting: FC, k_1, I_{TC}, I_{RB}, v_{avg}, and α.

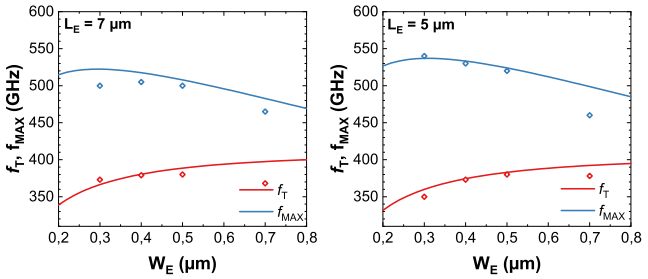

Fig. 5. Comparison of cutoff frequencies between model (line) and measurements (symbols), for different geometries.

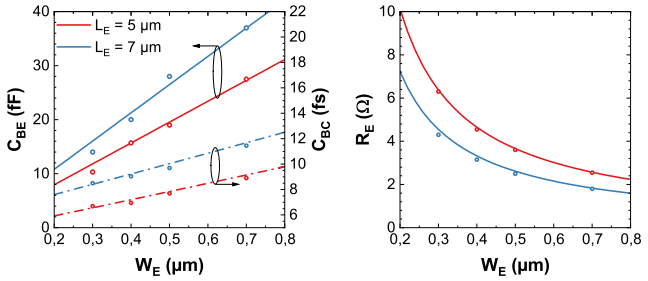

Fig. 6. Comparison between model (line) and measurements (symbol) for the base–emitter and the base–collector capacitances (left) as well as the emitter resistance (right), for different geometries.

Fig. 5 summarizes the comparison of the measured and simulated transition frequencies f_T and the maximum oscillation frequencies f_{MAX} for two emitter lengths L_E of 5 and 7 μm with a base contact width $W_{EB} = 0.3$ μm. For the transition frequency f_T, a good agreement between the model and the measurement is obtained for $W_E \leq 0.5$ μm. Similarly, a good agreement is also obtained for f_{MAX} for the narrower emitters. However, these predicted f_{MAX} values need to be interpreted carefully as they are quite difficult to extract reliably from measurements.

A more detailed analysis is required to understand the deviation between the model and the measurement, especially, for $W_E = 0.7$ μm. Fig. 6 shows the base–collector capacitance, the base–emitter capacitance, and the emitter resistance. A good agreement is observed for all emitter widths W_E and lengths L_E. Moreover, the error observed in Fig. 5 cannot either be due to the transit time (as τ_C is adjusted to fit to

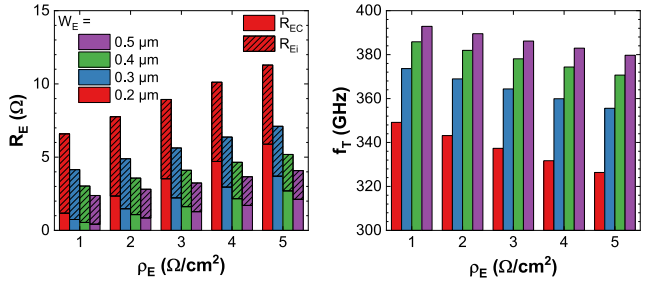

Fig. 7. Simulated emitter resistance R_E (the striped part corresponds to the semiconductor resistance R_{Ei}) at different contact resistivity for different emitter widths (left) and associated f_T (right).

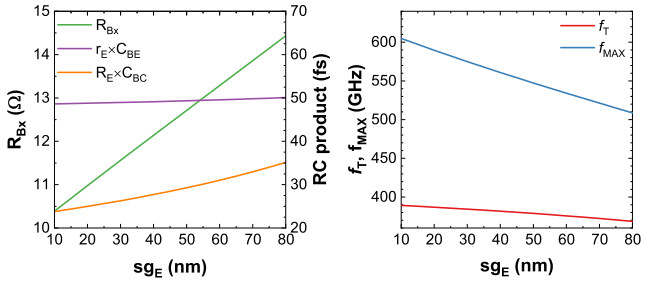

Fig. 8. Influence of the emitter undercut sg_E on R_{Bx}, $r_E \times C_{BE}$, and $R_E \times C_{BC}$ (left), and on the frequency performance f_T and f_{MAX}.

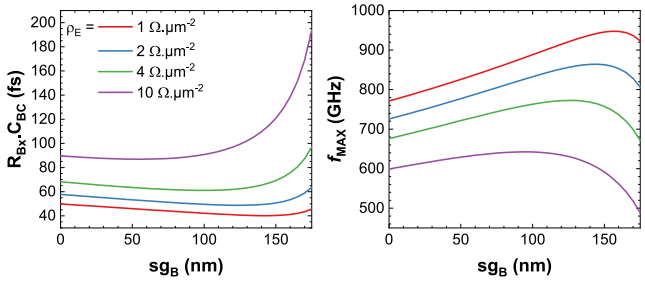

Fig. 9. Influence of the base undercut sg_B on the $R_{BX} \times C_{BC}$ product (left) and on the maximum oscillation frequency f_{MAX} (right).

the measurements) or the ideality factor n which is also accurately extracted from measurements. Hence, we infer that the discrepancies observed come from an underestimation of the total collector resistance R_C. For small geometries R_C is negligible compared to R_E, which is why the error is visible for $W_E = 0.7\ \mu$m and is more significant for $L_E = 7\ \mu$m.

B. Analysis of Technological Features

The optimization of cut-off frequencies is mainly done by reducing the dimensions of the transistor. This must be accompanied by improvement in technological features through process optimization as the parasitic elements of the fabricated device become non-negligible. In particular, the transistor operation is limited by contact resistances and parasitic capacitances. Hence, a good control over the emitter and base mesa fabrication is crucial. The importance of these two elements is analyzed using the model presented above for a transistor with an emitter width $W_E = 0.4\ \mu$m (unless otherwise specified) and an emitter length $L_E = 5\ \mu$m.

1) Emitter Features: Technological features of the emitter may lead to two limitations. First, the metal–semiconductor resistance adds a significant contribution to the total emitter resistance. This resistance partly explains why the f_T decreases with the emitter width (the other reason is that at constant current density the contribution (nkT/qI_C) increases). The reduction of the contact resistance is a key factor for supporting the miniaturization required by the frequency increase. Fig. 7 shows the total emitter resistance as well as the transition frequency f_T for different emitter contact resistivity ρ_E. The first point to note is that the reduction of ρ_E is more beneficial for small emitter widths. A decrease of ρ_E from 5 to 1 $\Omega/\mu\mathrm{m}^2$ increases f_T by 7% for transistor with $W_E = 0.2\ \mu$m while it increases f_T by 3.4% for transistor with $W_E = 0.4\ \mu$m. The second point to note is that this reduction of ρ_E to 1 $\Omega/\mu\mathrm{m}^2$ makes the contribution of the contact resistivity relatively minor. Therefore, further reduction of R_E must be achieved through the reduction of the semiconductor contribution R_{Ei}, as it has been done in [2] and [3].

Another key element for the emitter design is the undercut depth induced by the wet-etch process. This undercut increases the distance sg_E between the base contact and the emitter edge. Fig. 8 shows the influence of sg_E on different parameters such as the frequency performances. It appears that this undercut has to be reduced for multiple reasons. First, this undercut is directly related to the emitter area, this leads R_E to decrease and the C_{BE} to increase naturally while sg_E decreases. As shown in Fig. 8, this reduces the product $R_E \times C_{BC}$ as well as the product $r_E \times C_{BE}$, and in turn increases the transition frequency f_T. The reduction of $r_E \times C_{BE}$ with sg_E can be explained by the fact that the simulation is performed at a constant current density which then leads to the reduction of r_E as the emitter section increases. This undercut is also linked to R_{Bsg} (therefore, to R_{Bx}) which explains why f_{MAX} is impacted by sg_E. The undercut sg_E, therefore, needs to be minimized in order to maximize both the maximum oscillation frequency f_{MAX} and the transition frequency f_T.

2) Base Features: The base mesa fabrication strongly influences the maximum oscillation frequency f_{MAX} according to (2) as it impacts both R_{Bx} and C_{BCx}. Fig. 9 shows the impact of the base undercut on both $R_{Bx} \times C_{BC}$ and f_{MAX}. It is well known that reducing the base contact resistivity ρ_B increases f_{MAX}. However, more than that, through this decrease in ρ_B, the undercut sg_B can be increased and further maximize the frequency performance. Fig. 9 shows that the $R_{Bx} \times C_{BC}$ product reaches a minimum at a given sg_B which in turn depends on ρ_B. Optimizing only the contact resistivity prevents from achieving substantial gains on f_{MAX}. It is a difficult optimization as it requires to have a good control over the undercut. Moreover, this undercut slightly increases the transition frequency f_T by lowering the base–collector capacitance. Another method to achieve an optimal $R_{Bx} \times C_{BC}$ product could be to reduce the base contact width. This method has the disadvantage of increasing the contribution of the metal stack resistance, R_{Bm}, to the resistance R_{BC}.

C. Epitaxial Structure Optimization

As seen previously, technological feature optimization in the emitter and the base leads to significant gain on the maximum

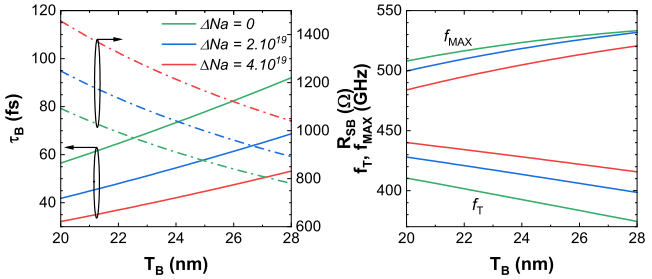

Fig. 10. Base transit time τ_B and base sheet resistance R_{SB} as function of the base thickness for different base doping graduality (left). Associated frequency performance (right).

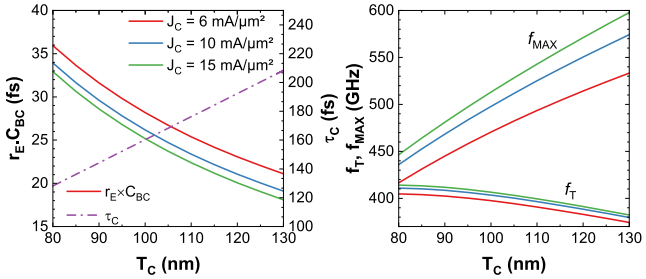

Fig. 11. Collector transit time τ_C and base–collector capacitance C_{BC} as function of collector thickness (left) and the associated frequency performance (right).

oscillation frequency f_{MAX}. The improvement of the transition frequency f_T depends, for the most part, on the vertical structure optimization. These optimizations are mainly focused on the base and the collector for which the transit times are described in (1).

1) Base Optimization: The optimization of the base is compulsory in order to reduce the base transit time and to maximize f_{MAX}. It is possible to decrease the transit time by creating an offset ΔE_C in the conduction band governed by (20). This offset can be created in two ways: 1) gradual composition or 2) gradual doping. Gradual composition is limited by the lattice mismatch it induces, however, it has the advantage of maintaining the base sheet resistance constant. On the other hand, gradual doping can lead to an increase of the base sheet resistance. The doping graduality is, thus, a tradeoff between base transit time and base sheet resistance. Another method of decreasing the base transit time is to reduce the base thickness. This solution also increases the base sheet resistance. Doping graduality and reduction of the base thickness, however, has the advantage of increasing the transistor static current gain β. Fig. 10 shows the frequency performance for different base thicknesses as well as different doping graduality (in each case there is a gradual composition from $Ga_{0.53}In_{0.47}As$ on the emitter side to $Ga_{0.47}In_{0.53}As$ on the collector side). The doping on the emitter side is set at 8×10^{19} cm^{-3}. Interestingly, a low graduality in doping marginally reduces f_{MAX} although it increases the base sheet resistivity. The significant increase in f_T compensates for it. The choice of the base epitaxial structure is, thus, a tradeoff between increasing f_T and decreasing f_{MAX}.

2) Collector Optimization: The optimization of the collector plays an important role in increasing frequency performances through to reduction of the collector transit time τ_C. In the case of a type-I InP DHBT, a major part of the optimization consists of optimizing the base–collector junction to reduce the effect of the heterojunction barrier. The model does not take this aspect into account as the electron velocity in the collector is considered to be constant. With this hypothesis, the study focuses on the impact of collector thickness on frequency performance as well as the Kirk current density (which is linked to the collector doping). The higher the Kirk current density, the higher the optimal collector current density (J_C at which f_T is maximal). This results in a lower dynamic emitter resistance r_E ($= (nkT/qI_C)$). Fig. 11 shows the $r_E \times C_{BC}$ product, the collector transit time and the

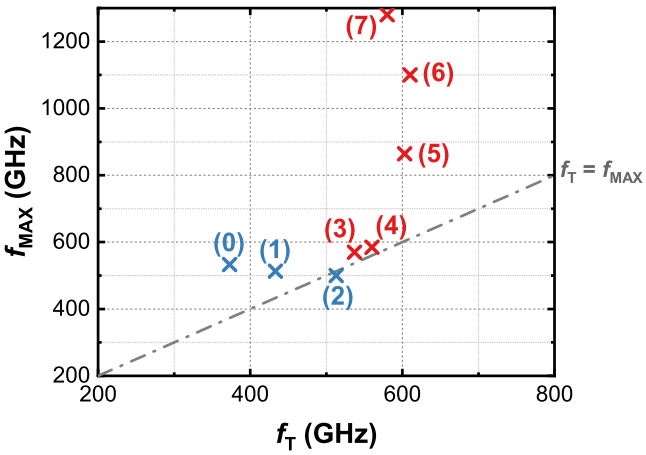

Fig. 12. f_T and f_{MAX} evolution with vertical structure (blue) and process (red) improvement.

frequency performances as a function of the collector thickness for three collector current densities.

As expected from our assumption, the transit time τ_C is proportional to the collector thickness T_C. It is, therefore, obvious for f_T to increase while T_C decreases. On the other hand, the reduction of T_C increases C_{BC} which then decreases f_{MAX}. As observed for the base, the choice of collector thickness is also a tradeoff between increasing f_T and decreasing f_{MAX}. Moreover, it is necessary to keep in mind that the collector thickness is linked to the breakdown voltage BV_{CE0}. Another way to improve the frequency performance is to increase the Kirk current density (by increasing the average collector doping). Due to this, it is possible to increase f_T, as the $r_E \times C_{BC}$ product is reduced. As the base–collector capacitance remains constant, this also increases f_{MAX}. The main drawback is that the device operates at higher temperature which can limit its performances due to self-heating.

D. Toward Ultimate Frequency Performance

As discussed previously, there are several means to improve the frequency performance of InP DHBTs. In this part, all these improvements are summarized to assess the ultimate achievable performances. Fig. 12 shows the expected frequency performance f_T and f_{MAX} after different optimizations. The optimizations related to the vertical structure are shown by blue symbols while the ones related to the technological improvement are presented using red.

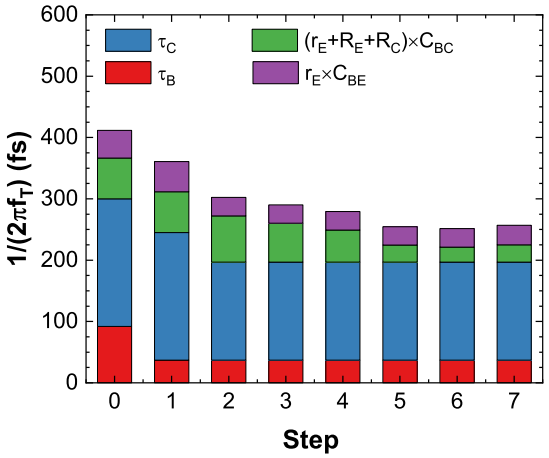

Fig. 13. RC product and transit times that limit f_T at each step of improvement.

On this figure, (0) corresponds to the initial measured performance of a $0.4 \times 5\ \mu\text{m}^2$ transistor with $W_B = 0.3\ \mu$m. The optimization performed at each of the steps, in Fig. 12, is as follows.

1) Base thickness is reduced to 20 nm, a gradual doping from 9×10^{19} to 6×10^{19} cm^{-3} is introduced. It reduces the base transit time and slightly increases the base sheet resistivity.
2) The collector thickness is reduced down to 100 nm. At the same time the doping is adjusted to push the Kirk current to 10 mA/μm^2. Owing to this the collector transit time as well as the dynamic emitter resistance are reduced. This leads to an increase in the transition frequency f_T.
3) Emitter undercut is reduced to 20 nm leading to an increase both in f_T and f_{MAX} as the emitter resistance and the base access resistance are reduced.
4) Emitter contact resistivity is reduced to 1 Ω/μm^2 leading to a slight increase of both f_T and f_{MAX}.
5) Base contact resistivity is reduced to 5 Ω/μm^2 and the base undercut to 130 nm. The transition frequency f_T increases as C_{BC} decreases. Combined with the reduction of R_{Bx} this leads to a significant increase of f_{MAX}.
6) Base contact resistivity is reduced to 1 Ω/μm^2.
7) The emitter width is reduced to $W_E = 0.2\ \mu m$. It decreases the intrinsic part of C_{BC} and leads to higher f_{MAX}. However, f_T is negatively affected by the increase in R_E.

With these optimizations, the expected frequency performances show the promise of InP DHBTs for THz applications. These predictions must be interpreted with caution as some effects are neglected in the model. For example, the *FC* coefficient introduced to compute the C_{BEj} capacitance may change with the layout. Likewise, the α coefficient introduced to take into account the distributed aspect of the base–collector capacitance changes when the base contact resistivity and undercut are modified. However, it is clear that, eventually, the maximum oscillation frequency f_{MAX} will be far higher than the transition frequency f_T. To understand this limitation, it is possible to analyze the elements contributing to f_T. Fig. 13 shows the different contributions that limit the transition frequency at each improvement step.

As expected, the first two steps related to the base and collector transit time reduction induce an important impact on f_T. The optimal base undercut that leads to the reduction of the base contact resistivity also has a relatively significant impact on the transition frequency. At the last step, despite significant technological optimizations, the transition frequency is still limited. The collector transit time represents more than 60% of the total contribution. As the possible gains through an optimization of technological features are low, the only way to improve the transition frequency would be to reduce the collector thickness which would result in a lower breakdown voltage.

V. Conclusion

We have developed an analytical geometric model to predict frequency performance of InP/InGaAs DHBTs. This model uses an approach which consists of reconstructing the small signal model. To this end, the computation of all the intrinsic and parasitic elements has been detailed. This analytical model shows good agreement with measurements obtained on various transistor geometries. Next, the influences of the emitter features as well as that of the base on the frequency performance have been studied. The optimization of the vertical structure was also addressed. Finally, the combination of both technological features and vertical epitaxial structure optimizations have been studied to understand the ultimate frequency performance that can be achievable. This work paves the way to transistors featuring $f_{\text{MAX}} > 1$ THz and $f_T > 500$ GHz.

This work also opens up the possibilities for even more accurate modeling, in particular by combining this model with compact modeling approaches taking into account additional phenomena, such as the Kirk effect and self-heating.

References

[1] X. Mei et al., "First demonstration of amplification at 1 THz using 25-nm InP high electron mobility transistor process," *IEEE Electron Device Lett.*, vol. 36, no. 4, pp. 327–329, Apr. 2015.
[2] J. C. Rode et al., "Indium phosphide Heterobipolar transistor technology beyond 1-THz bandwidth," *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 2779–2785, Sep. 2015.
[3] A. M. Arabhavi et al., "InP/GaAsSb double Heterojunction bipolar transistor emitter-fin technology with $f_{\text{MAX}} = 1.2$ THz," *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 2779–2785, Sep. 2015. [Online]. Available: https://ieeexplore.ieee.org/document/9672313/
[4] J. F. Buckwalter et al., "Prospects for high-efficiency silicon and III-V power amplifiers and transmitters in 100-300 GHz bands," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Apr. 2021, pp. 1–7.
[5] "Twilight-EU Project—Towards the New Era of 1.6 TB/s System-in-Package Transceivers for Datacenter Applications Exploiting Wafer-Scale Co-Integration of InP Membranes and InP-HBT Electronics." Accessed: Nov. 22, 2022. [Online]. Available: https://ict-twilight.eu/
[6] R. Hersent et al., "160-GSa/s-and-beyond 108-GHz-bandwidth over-2-V_{ppd} output-swing 0.5-um InP DHBT 2:1 AMUX-driver for next-generation optical communications," *IEEE Microw. Wireless Compon. Lett.*, vol. 32, no. 6, pp. 752–755, Jun. 2022.
[7] X. Wen et al., "Performance prediction of InP/GaAsSb double heterojunction bipolar transistors for THz applications," *J. Appl. Phys.*, vol. 130, no. 3, Jul. 2021, Art. no. 034502. [Online]. Available: http://aip.scitation.org/doi/10.1063/5.0054197
[8] V. Palankovski, R. Schultheis, and S. Selberherr, "Simulation of power heterojunction bipolar transistors on gallium arsenide," *IEEE Trans. Electron Devices*, vol. 48, no. 6, pp. 1264–1269, Jun. 2001.

[9] M. Müller, P. Dollfus, and M. Schröter, "1-D drift-diffusion simulation of two-valley semiconductors and devices," *IEEE Trans. Electron Devices*, vol. 68, no. 3, pp. 1221–1227, Mar. 2021.
[10] X. Wen et al., "A multiscale TCAD approach for the simulation of InP DHBTs and the extraction of their transit times," *IEEE Trans. Electron Devices*, vol. 66, no. 12, pp. 5084–5090, Dec. 2019.
[11] M. Schroter, S. Lehmann, S. Fregonese, and T. Zimmer, "A computationally efficient physics-based compact bipolar transistor model for circuit design—Part I: Model formulation," *IEEE Trans. Electron Devices*, vol. 53, no. 2, pp. 279–286, Feb. 2006.
[12] C. Mukherjee et al., "Scalable compact modeling of III-V DHBTs: Prospective figures of merit toward Terahertz operation," *IEEE Trans. Electron Devices*, vol. 65, no. 12, pp. 5357–5364, Dec. 2018.
[13] S. Blayac et al., "InP HBT self-aligned technology for 40 Gbit/s ICs: Fabrication and CAD geometric model," in *Proc. Conf. 11th Int. Conf. Indium Phosphide Related Mater. (IPRM)*, May 1999, pp. 483–486.
[14] N. Davy et al., "0.4-um InP/InGaAs DHBT with a 380-GHz f_T, >600-GHz f_{MAX} and BVCE0 >4.5 V," in *Proc. IEEE BiCMOS Compd. Semicond. Integr. Circuits Technol. Symp. (BCICTS)*, Dec. 2021, pp. 1–4.
[15] H. K. Gummel and H. C. Poon, "An integral charge control model of bipolar transistors," *Bell Syst. Tech. J.*, vol. 49, no. 5, pp. 827–852, May 1970.
[16] R. Shrivastava and K. Fitzpatrick, "A simple model for the overlap capacitance of a VLSI MOS device," *IEEE Trans. Electron Devices*, vol. 29, no. 12, pp. 1870–1875, Dec. 1982.
[17] G. Reeves and H. Harrison, "Obtaining the specific contact resistance from transmission line model measurements," *IEEE Electron Device Lett.*, vol. 3, no. 5, pp. 111–113, May 1982.
[18] W. Liu, "Emitter-length design for microwave power heterojunction bipolar transistors," *Solid-State Electron.*, vol. 36, no. 6, pp. 885–890, 1993. [Online]. Available: https://www.sciencedirect.com/science/article/pii/003811019390011E
[19] F. Ali and A. K. Gupta, *HEMTs and HBTs: Devices, Fabrication, and Circuits*. London, U.K.: Artech House, 1991.
[20] J. Hauser, "The effects of distributed base potential on emitter-current injection density and effective base resistance for stripe transistor geometries," *IEEE Trans. Electron Devices*, vol. 11, no. 5, pp. 238–242, May 1964.
[21] H. Kroemer, "Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region," *Solid-State Electron.*, vol. 28, no. 11, pp. 1101–1103, Nov. 1985. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/003811018590190X
[22] M. J. W. Rodwell et al., "Submicron scaling of HBTs," *IEEE Trans. Electron Devices*, vol. 48, no. 11, pp. 2606–2624, Nov. 2001.
[23] J. Lopez-Gonzalez and L. Prat, "The importance of bandgap narrowing distribution between the conduction and valence bands in abrupt HBTs," *IEEE Trans. Electron Devices*, vol. 44, no. 7, pp. 1046–1051, Jul. 1997. [Online]. Available: http://ieeexplore.ieee.org/document/595930/
[24] T. Nardmann, M. Schroter, and P. Sakalas, "A multiregion approach to modeling the base-collector junction capacitance," *IEEE Trans. Electron Devices*, vol. 63, no. 9, pp. 3808–3811, Sep. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7529154/
[25] T. K. Johansen, V. Krozer, V. Nodjiadjim, A. Konczykowska, J. Dupuy, and M. Riet, "Improved external base resistance extraction for Submicrometer InP/InGaAs DHBT models," *IEEE Trans. Electron Devices*, vol. 58, no. 9, pp. 3004–3011, Sep. 2011.

Nil Davy (Member, IEEE) received the engineering degree in electrical engineering from Phelma, Grenoble Institute of Technology, Grenoble, France, in 2020. He is currently pursuing the Ph.D. degree with the Circuit for Analog/Digital Interface Department, III-V Lab, Palaiseau, France, in partnership with IMS Bordeaux, Bordeaux, France.

His current research interests include the design, characterization, and modeling of InP double heterojuction bipolar transistors.

Virginie Nodjiadjim (Member, IEEE) received the Ph.D. degree in electronic engineering from the University of Lille I, Lille, France, in 2009.

She joined III-V Lab, Palaiseau, France, as a Research Engineer in 2009, where she has studied compound semiconductor heterojunction bipolar transistors, and is currently in charge of the InP double heterojunction bipolar transistor development through the optimization of the structure, the design of the device, characterization, and modeling. She has authored or coauthored more than 70 peer-reviewed scientific publications.

Muriel Riet was born in Choisy-le-Roi, France, in 1958. She received the Ph.D. degree in electronic engineering from the University of Paris XI, Bures-sur-Yvette, France, in 1985.

She joined CNET, Reserch Center of France TELECOM, Rennes, France, in 1985, where she has been in charge of HBT technology for high bit-rate optical communications up to 40 Gb/s. She has been in charge of InP HBT technology for high-bite-rate optical commutations up to 40 Gb/s with the III-V Lab, Palaiseau, France, since 1998.

Colin Mismer received the M.Sc. (engineering) degree in chemical industry from the ENSICAEN, Caen, France, in 2011.

He joined Nokia Bell Laboratories and III-V Lab, Palaiseau, France, in 2017, where he is currently a Microelectronics Process Engineer. His work focuses on the fabrication of InP-DHBT integrated circuit and the development of new technological building blocks.

Marina Deng (Member, IEEE) received the Ph.D. degree in electronics from the University of Lille, Villeneuve-d'Ascq, France, in 2014.

During her Ph.D. research with IEMN, Villeneuve-d'Ascq, she worked on SiGe HBT small-signal and high-frequency noise characterization and modeling in the sub-THz range. Since November 2015, she has been an Associate Professor with the University of Bordeaux, Bordeaux, France. She is currently the Head of the Model 4 Circuit team, within the IC Design research group. Her current research focuses on millimeter-wave and sub-millimeter-wave characterization of HF transistors for THz electronics applications, including on-wafer calibration and noise measurements.

Chhandak Mukherjee (Member, IEEE) received the M.Tech. and Ph.D. degrees from the Indian Institute of Technology Kharagpur, Kharagpur, India, in 2010 and 2013, respectively.

He had been with the IMS laboratory, University of Bordeaux, Bordeaux, France, and the Leibniz Institute for Microelectronics, IHP, Frankfurt, Germany, as a Postdoctoral Researcher. Since 2019, he has been working as a CNRS Researcher with the IMS Laboratory focusing on advanced characterization methods, reliability assessment, and physics-based compact modeling of advanced and emerging semiconductor technologies. He is currently involved as a Workpackage Leader of a European project for developing 3-D integrated computing architectures with embedded neural networks based on vertical nanowire transistors as AI accelerators.

Bertrand Ardouin received the M.S. and Ph.D. degrees in electrical engineering from the University of Bordeaux, Bordeaux, France, in 1998 and 2001, respectively.

In 2002, he has been the Co-Founder of XMOD Technologies, Bordeaux, a startup he has led as a CEO and an R&D Director until 2019. After the acquisition of XMOD technologies by SERMA Group, he has been a Business Unit Manager with SERMA Technologies, Pessac, France, in charge of electrical expertise and environmental tests and responsible compact modeling activities. He is currently responsible of the High-Speed Analogue Digital Interfaces Group, III-V Lab, Palaiseau, France. His areas of interest cover CMOS, SiGe and InP HBT devices, harsh-environment electronics, reliability, RF and mmWave characterization, modeling & PDK, process technology, statistical modeling, and high-speed IC design.

Cristell Maneux (Member, IEEE) received the M.Sc. degree in electronics engineering and the Ph.D. degree in electronics from the University of Bordeaux, Bordeaux, France, in 1994 and 1998, respectively.

She has been Associate Professor with IMS Laboratory, Department of Sciences and Engineering, University of Bordeaux, from 1998 to 2012, where she has been a Professor since 2012, and the Director since January 2022. She currently leads the French ANR LEGO Project, and the European research project FVLLMONTI, call H2020 FETPROACT-09-2020. She has also been the Director of the IMS-CEA LETI Common Lab, Grenoble, France, since 2021 and is a member of the steering committee of the IMS-ST Microelectronics Common Lab, Grenoble, since 2006. She has coauthored more than 200 publications on high-impact journals and conferences. Her research interests focus on compact modeling of advanced and emerging devices: InP HBT, SiGe HBT, carbon nanotube transistors, graphene transistors, nanowire transistors; device electrical characterization: DC, RF, pulsed, low-frequency noise, RTS noise; device failure mechanisms and integrated circuit reliability; THz-integrated devices for beyond 5G communications; and unconventional nanoelectronics.

Dr. Maneux serves as a Technical Program Committee Member and/or a Reviewer for DATE, EuMW, Euro-SOI, ISCAS, NEWCAS, and ESSDERC. Since 2022, she has been an Associate Editor for the IEEE TRANSACTIONS OF COMPUTER AIDED DESIGN and since 2020, she has been a member of the Editorial Advisory Board of Solid State Electronics.