InP/GaAsSb DHBTs for THz Applications and Improved Extraction of their Cutoff Frequencies

C.R. Bolognesi (*Fellow, IEEE*), R. Flückiger, M. Alexandrova, W. Quan, R. Lövblom, O. Ostinelli
Millimeter-Wave Electronics Group, ETH-Zürich, CH 8092 Zürich, Switzerland, email: colombo@ieee.org

Abstract**—**The development of InP/GaAsSb DHBTs is reviewed and contextualized with respect to other III-V high-speed technologies. Pertinent material properties and challenges in the proper assessment of f_{MAX} are discussed. An iterative de-embedding algorithm involving no additional test structures/measurements is shown to yield the correct f_{MAX} from unilateral gain data for both DHBTs and HEMTs.

I. Introduction

In the late 1990's, InP/GaInAs HBTs achieved cutoff frequencies of 150 GHz with a common-emitter breakdown voltages of $BV_{\mathrm{CEO}} \sim 4$ V, while the GaAs HBT literature reported 120 GHz with $BV_{\mathrm{CEO}} = 5$V, placing both at a product $f_{\mathrm{T}} \times BV_{\mathrm{CEO}} = 600$ GHz·V. It was a natural idea to insert an InP collector under the GaInAs base and form a double heterojunction bipolar transistor (DHBT) with the intent to exploit the superior high-field breakdown and transport properties of InP. This however complicated matters due to the blocking nature of the "Type-I" heterojunction with $\Delta E_{\mathrm{C}} = 0.25$ eV between GaInAs and InP. Various grading approaches were studied [1, 2] but they all share a maximum electric field just where the energy gap is minimal —right at the B/C junction.

The use of a GaAsSb base layer independently arose as an alternative for InP DHBTs in 3 laboratories in 1996 [3-5]. The first two groups [3, 4] did not pursue the idea. Starting work in 1997 [5], we reported the first InP DHBTs with $f_{\mathrm{T}}/f_{\mathrm{MAX}} = 300/300$ GHz and $BV_{\mathrm{CEO}} > 6$V [6]. The result was of significance as it was the fastest bipolar transistor ever built, in any material system. To contextualize, InP/GaInAs DHBTs of the time offered $f_{\mathrm{T}}/f_{\mathrm{MAX}} = 200/300$ GHz [2] and appeared confined to lower bandwidths than InP HEMTs which already reached $f_{\mathrm{T}} = 300$ GHz. In a swift transition to industry, a GaAsSb DHBT process was transferred to *Agilent* (now *Keysight*) [7] where it matured independently and began shipping in 2004.

II. Physical Characteristics

A. Band Diagram

We consider the band diagram (Fig. 1) and relate it to device performance. The $\mathrm{GaAs_{0.51}Sb_{0.49}}$ alloy is lattice-matched to InP, enabling the growth of a symmetrical transistor structure with InP emitter and collector layers. Junction symmetry reduces the collector offset voltage. As a subsequent refinement, it became advantageous to eliminate the $\Delta E_{\mathrm{C}} = -0.12$ eV at the E/B junction with a composite graded (Ga,In)P emitter to increase the transistor current gain (Fig. 2). Finally, the addition of a quasi-electric field helps drive electrons across the base. In Fig. 1 this is achieved by grading the group-V element, *i.e.* the As/Sb ratio. This band diagram leads to Gummel characteristics with a unity collector current ideality factor $n_{\mathrm{C}} = 1.0$ indicating optimal turn-on of the E/B junction. The Type-II alignment at the B/C junction next ensures that little reverse bias needs to be applied to the collector when compared to graded Type-I collectors. Consequently, InP/GaAsSb DHBTs operate at lower V_{CE} voltages than their GaInAs-based counterparts. InP/GaAsSb DHBTs also reach their peak RF performance at significantly lower collector current densities and power dissipation levels.

B. Pertinent Material Properties

The use of a pure InP collector is attractive because of its good thermal conductivity $\kappa_{\mathrm{th}} = 0.7$ W/cm·K. In contrast, ternary and quaternary III-V alloys show values of 0.1 and < 0.05 W/cm·K, respectively. For device reliability in ICs, it is critical to minimize junction temperature. Noting that thermal resistance roughly scales with the inverse of device area makes InP/GaAsSb DHBTs even more compelling.

Reliability studies on industrially produced InP/GaAsSb DHBTs are rare but the technology was deemed "very robust" [8]. Considering why this may be, we can think of Ga(As,Sb) and (Ga, In)As as 50:50 alloys between GaAs and GaSb or InAs, respectively. Device failures involve atomic motion in response to stress, regardless of mechanism. As the mechanical properties of III-V alloys vary linearly with composition [9], we compare GaSb and InAs. First, the critically resolved shear stress (CRSS, *i.e.* the stress needed to induce dislocation motion: dislocations form easily when the CRSS value is low) and the stacking fault energy are higher in GaSb than InAs [10]. As well, dislocation velocity is much lower in GaSb compared to InAs (by 4 orders of magnitude at 200ºC) [10]. Finally, the anion intermixing induced by Au on III-V surfaces is somewhat weaker on GaSb than on InAs [11]. Such integrated physical considerations do suggest that GaAsSb has little to envy GaInAs for reliability —it should even be *better.*

GaAsSb can easily be *p*-doped with C to the very high levels necessary for the thin-base layers for high-speed DHBTs. The surface Fermi level pinning energy on GaAsSb is close to the valence band edge, ensuring low base Ohmic contact resistances, and a weaker surface depletion of the exposed base surface between the emitter mesa and the base contact. Both provide clear advantages for GaAsSb with respect to GaInAs for DHBT lateral and vertical transistor scaling.

978-1-5090-3902-9/16/$31.00 ©2016 IEEE

III. Cutoff Frenquencies of THz Transistors

One key challenge with THz transistors is the accurate determination of cutoff frequency metrics. Whereas the current gain cutoff frequency f_{T} is straightforward to determine thanks to a generally well-behaved $h_{21}(f)$, or even from low-frequency measurements via Gummel's method, the maximum oscillation frequency f_{MAX} is far more delicate. On one hand, determining f_{MAX} from maximum available gain (MAG) data requires accurately measuring to very high frequencies, *i.e.* well above the stability factor transition to $k > 1$. With Mason's Unilateral power gain U, the most significant error source in f_{MAX} arises from multiple spikes appearing in $U(f)$ [12].

According to our findings, spikes in U are introduced by the usual de-embedding process of the SHORT/OPEN pads in probe tip calibrated measurements using off-wafer impedance standards [13, 14]. These multiple U spikes are strictly unrelated to device physics. The standard pad de-embedding procedure does not accurately model their distributed nature at higher frequencies. The idea was tested by modeling the de-embedding of a "known" 600 GHz transistor DUT in ADS. Pads were treated as lossy lines, and the SHORT/OPEN patterns measured by simulating the structures in Fig. 6. The de-embedding procedure yields fundamentally different results whether one removes the OPEN [13] or SHORT [14] first. This is a critical issue because authors often do not clearly specify how they handled de-embedding. Both methods yield incorrect f_{MAX} values because the spikes in U corrupt the frequency response leading to gross errors in the extrapolated f_{MAX}.

We developed an iterative extraction algorithm to overcome this problem (Fig. 7), and verified its validity in ADS. Spikes in U do remain around given frequencies, emerging at $\lambda/8$ (with $\lambda/4$ periodicity), but their extent decreases rapidly with the number of iterations n. Clearly, the iterative scheme allows one to recover the correct "known" $f_{\mathrm{MAX}} = 600$ GHz from the de-embedding, thus proving the method. Iterative de-embedding was next applied to a real measurements of a 600 GHz transistor. The standard SHORT/OPEN [14] extraction yields a massively optimistic $f_{\mathrm{MAX}} = 1$ THz, while the OPEN/SHORT yields an underestimated $f_{\mathrm{MAX}} = 500$ GHz, which is closer to the correct value. Regardless of whether one begins iterations with the SHORT or OPEN in Fig. 7, both versions of the iterative method converge to the same f_{MAX} which can be regarded as the true device f_{MAX}. Important trends are apparent: shorter pads shift U spikes to higher frequencies, and smaller devices are more susceptible to gross U "spikiness" and errors when extrapolating for f_{MAX}. From Figs. 8-10, the SHORT/OPEN usual de-embedding *characteristically* curls U up, corrupting the –3dB corner of any single-time constant fit through U. The OPEN/SHORT method acts conversely. These observations have forensic value.

We tested our iterative scheme on 50 nm gate mHEMTs with a 0.3 μm^2 gate area from the *IAF* (Fig. 12). Here too, the SHORT/OPEN method [14] severely overestimates f_{MAX} yielding values >1.5 THz whereas the iterative scheme yields 500 GHz, in agreement with *IAF* measurements of MSG/MAG up to 450 GHz using on-wafer calibration standards. Externally smoothing U to run a single-time constant fit through the resulting data set would mask the spikes but still severely overestimate f_{MAX} —this is to be prohibited. We hereby propose the present iterative scheme as a new standard de-embedding practice. Failing that, authors should at a strict minimum present data de-embedded using both the SHORT/OPEN and OPEN/SHORT schemes, understanding that device f_{MAX} is closer to the lower bound.

Beyond the correct extraction of f_{MAX} metrics, iterative de-embedding greatly improves the agreement between modeled and measured MMIC circuit results, as seen in Fig. 13.

IV. Cutoff Frenquencies of THz Transistors

The foregoing material suggests that f_{MAX} records based on the two common de-embedding approaches are to be treated with caution. So far, the only published result using iterative de-embedding is [15]. The device showed $f_{\mathrm{T}}/f_{\mathrm{MAX}} = 503/779$ GHz with a power density of 10 mW/μm^2 and a $BV_{\mathrm{CEO}} = 4.1$ V ($J_{\mathrm{C}} = 1$ kA/cm^2) for $f_{\mathrm{T}} \times BV_{\mathrm{CEO}} = 2.1$ THz·V. Using conventional de-embedding, Rode *et al.* reported $f_{\mathrm{T}}/f_{\mathrm{MAX}} = 404/901$ GHz at 42 mW/μm^2 with $BV_{\mathrm{CEO}} = 4.3$ V (10 kA/cm^2) [16], which must be derated to ~3.4 V (1 kA/cm^2) for comparison to our data (thus, $f_{\mathrm{T}} \times BV_{\mathrm{CEO}} = 1.4$ THz·V). A more aggressively processed 0.2 × 4.4 μm^2 InP/GaAsSb DHBT shows $f_{\mathrm{MAX}} = 1.2$ THz at 14.8 mW/μm^2 according to standard de-embedding procedures (Fig. 14). This compares favorably to similarly de-embedded 0.22 × 2.7 μm^2 GaInAs-based DHBTs operated at 33 mW/μm^2 despite the smaller area of the GaInAs DHBT [17]. Iterative de-embedding reduces f_{MAX} to 882 GHz for our THz device (Fig. 15). Considering larger devices, iterative de-embedding for a 0.2 × 10 μm^2 InP/GaAsSb DHBT yields $f_{\mathrm{MAX}} = 700$ GHz.

In terms of future device developments, the use of quaternary base layers could enable further breakthroughs in light of the $f_{\mathrm{T}} = 513$ GHz with $BV_{\mathrm{CEO}} = 5.2$ V (1 kA/cm^2) reported by *NTT* for an impressive $f_{\mathrm{T}} \times BV_{\mathrm{CEO}} = 2.7$ THz·V [18].

Acknowledgment

The authors gratefully acknowledge Prof. O. Ambacher and Dr. M. Ohlrogge from the *Faunhofer IAF*. ETH FIRST Laboratory staff support is acknowledged as well.

References

[1] C. Nguyen *et al.*, *IEEE Electron Dev. Lett.*, Vol. 17, p. 133, 1995.
[2] A. Fujihara *et al.*, *Proc. IEEE IEDM*, p. 772, 2001.
[3] R. Bhat *et al.*, *Appl. Phys. Lett.*, Vol. 68, p. 985, 1996.
[4] B. T. McDermott, *et al.*, *Appl. Phys. Lett.*, Vol. 68, p. 1386, 1996.
[5] C. R. Bolognesi, NSERC Database. Available online: http://www.nserc-crsng.gc.ca/ase-oro/index_eng.asp
[6] M. W. Dvorak *et al.*, *IEEE Electron Dev. Lett.*, Vol. 22, p. 361, 2001.
[7] N. J. Moll and C.R. Bolognesi, US Patent 6,761,480, 2004.
[8] G. A. Koné *et al.*, *Proc. IEEE IPRM*, p. 208, 2012.
[9] S. Adachi, *Properties of Semiconductor Alloys*, Wiley, 2009.
[10] O. Oda, *Compound Semiconductor Bulk Materials and Characterizations*, World Scientific, 2007.
[11] L.J. Brillson, *Thin Solid Films*, Vol. 89, p. 461, 1982.
[12] V. Teppati *et al.*, *IEEE Trans. Electron Devices*, Vol. 61, p. 984, 2014.
[13] M.C.A.M. Koolen *et al. Proc. BCTM*, p. 188, 1991.
[14] L. Tiemeijer *et al.*, *IEEE Trans. Electron Devices,* Vol. 50, p. 822, 2003.
[15] M. Alexandrova *et al.*, *IEEE Electron Dev. Lett.*, Vol. 35, p. 1218, 2014.
[16] J.C. Rode *et al.*, *IEEE J. Electron Dev. Soc.*, Vol. 3, p. 54, 2015.
[17] V. Jain *et al.*, *Proc. IEEE DRC*, p. 271, 2011.
[18] N. Kashio *et al.*, *IEEE Electron Dev. Lett.*, Vol. 36, p. 657, 2015.

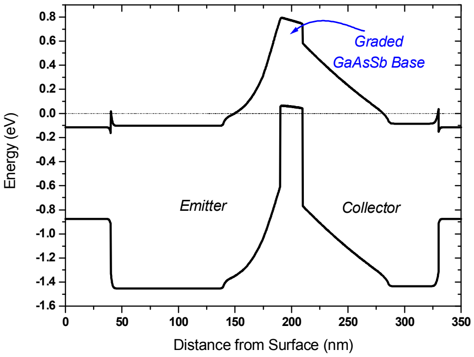

Fig. 1. Equilibrium band diagram of a modern high-speed InP/GaAsSb DHBT.

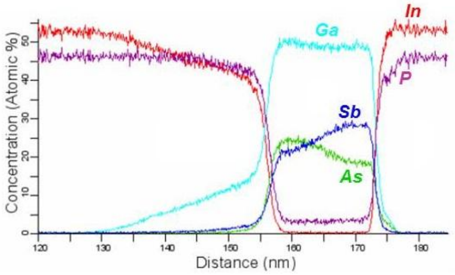

Fig. 2. Atom probe measurement of composition through a mixed-group-V graded base transistor produced at ETHZ.

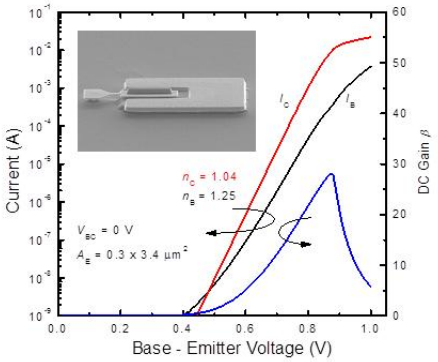

Fig. 3. Typical Gummel characteristics with n_C = 1.04.

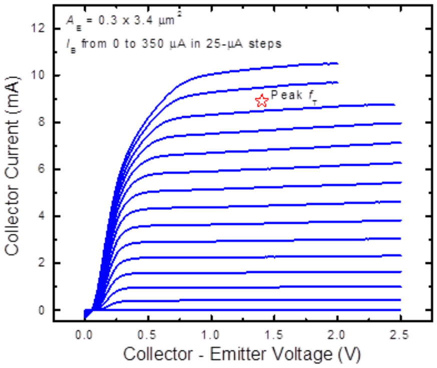

Fig. 4. Typical I_C-V_{CE} characteristics with a low offset voltage.

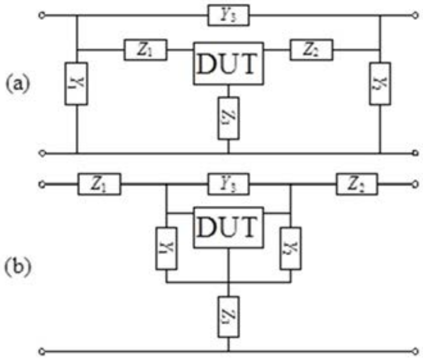

Fig. 5. Lumped equivalent circuits used in the traditional pad de-embedding models: (a) OPEN/SHORT and (b) SHORT/OPEN.

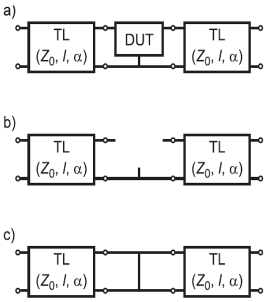

Fig. 6. ADS simulation of de-embedding process. b) OPEN, c) SHORT.

SHORT/OPEN	OPEN/SHORT
$Z^{\mathrm{SO}}_{\mathrm{TEMP},i} = \dfrac{(Y^{\mathrm{SO}}_{\mathrm{S},i\text{-}1})^{-1}}{n+1-i}$	$Y^{\mathrm{OS}}_{\mathrm{TEMP},i} = \dfrac{Y^{\mathrm{OS}}_{\mathrm{O},i\text{-}1}}{n+1-i}$
$Y^{\mathrm{SO}}_{\mathrm{TEMP},i} = \dfrac{((Y^{\mathrm{SO}}_{\mathrm{O},i\text{-}1})^{-1} - Z^{\mathrm{SO}}_{\mathrm{TEMP},i})^{-1}}{n+1-i}$	$Z^{\mathrm{OS}}_{\mathrm{TEMP},i} = \dfrac{(Y^{\mathrm{OS}}_{\mathrm{S},i\text{-}1} - Y^{\mathrm{OS}}_{\mathrm{TEMP},i})^{-1}}{n+1-i}$
$Y^{\mathrm{SO}}_{X,i} = ((Y^{\mathrm{SO}}_{X,i\text{-}1})^{-1} - Z^{\mathrm{SO}}_{\mathrm{TEMP},i})^{-1} - Y^{\mathrm{SO}}_{\mathrm{TEMP},i}$	$Y^{\mathrm{OS}}_{X,i} = ((Y^{\mathrm{OS}}_{X,i\text{-}1} - Y^{\mathrm{OS}}_{\mathrm{TEMP},i})^{-1} - Z^{\mathrm{OS}}_{\mathrm{TEMP},i})^{-1}$

Fig. 7. Iterative de-embedding algorithm. Measured admittance matrices are assigned the subscript “0” and are the same in both cases. The case n = 1 corresponds to traditional de-embedding.

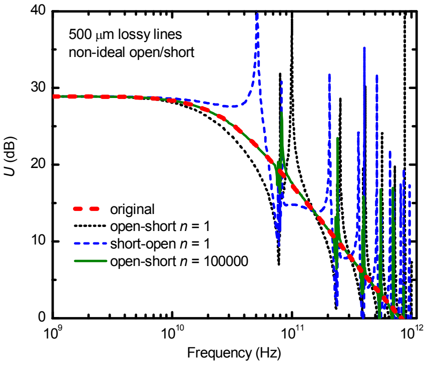

Fig. 8. Simulated de-embedding of a 600 GHz transistor with 500 μm pads. The usual SHORT/OPEN approach (blue, dashed) results in huge overestimate for f_{MAX}. The iterative method exactly recovers the known “original” DUT.

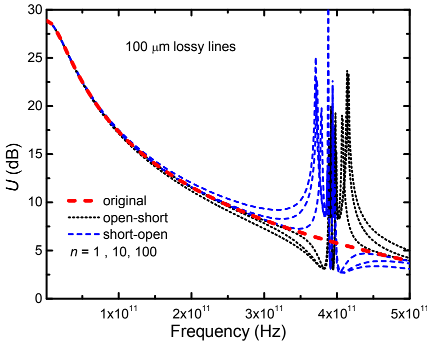

Fig. 9. De-embedding results with 100 µm pads for $n = 1$, 10 and 100 iterations. Fractional dB errors in U at 100 GHz extrapolate to significant errors on f_{MAX}.

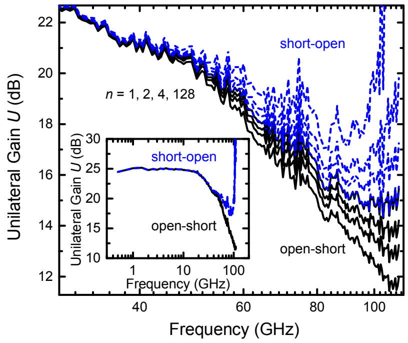

Fig. 10. De-embedding of a real 600 GHz DHBT for 1, 2, 4 and 128 iterations. Note linear frequency scale. Results converge quickly with n.

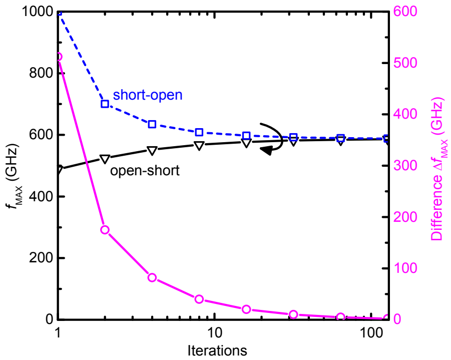

Fig. 11. Estimated f_{MAX} *vs.* iteration number for the device measured in Fig. 10.

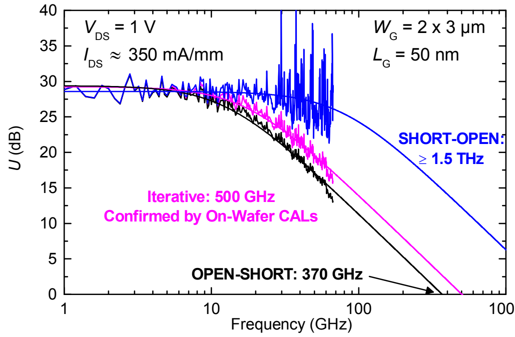

Fig. 12. De-embedding of a 50 nm gate 2 × 3 µm mHEMT from the *Fraunhofer IAF*. Iterative scheme gives $f_{\mathrm{MAX}} = 500$ GHz in agreement with MAG measurements up to 450 GHz at *IAF*.

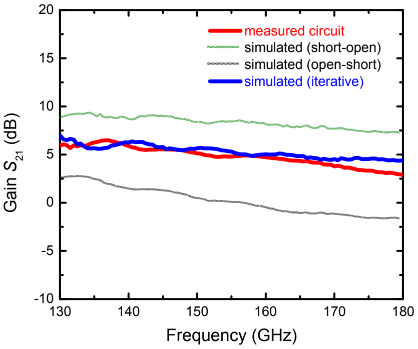

Fig. 13. Measured MMIC gain. The iterative scheme results in good agreement with the measured data. Substrate is un-thinned and parasitic modes induce a discrepancy above 170 GHz.

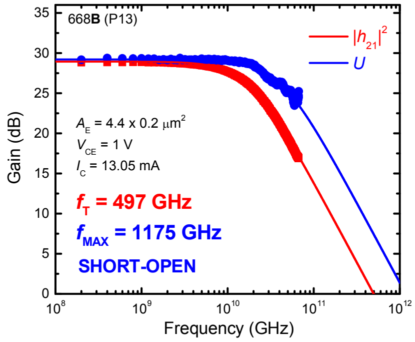

Fig. 14. Conventional SHORT/OPEN [14] de-embedding yields $f_{\mathrm{MAX}} = 1.2$ THz according to standard practice.

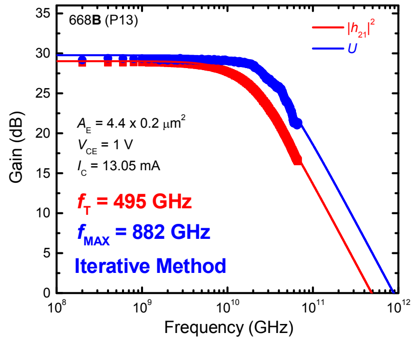

Fig. 15. Iterative method applied to device of Fig. 14 yields $f_{\mathrm{MAX}} = 882$ GHz. OPEN/SHORT [13] leads to $f_{\mathrm{MAX}} = 775$ GHz.