Importance of Probe Choice for Extracting Figures of Merit of Advanced mmW Transistors

Sebastien Fregonese, Magali de Matos, Marina Deng, Didier Celi, Nicolas Derrier, Thomas Zimmer

Importance of probe choice for extracting figures of merit of advanced mmW transistors

Sebastien Fregonese, Magali De Matos, Marina Deng, Didier Céli, Nicolas Derrier, and Thomas Zimmer, Senior *Member, IEEE*

***Abstract*—The measurement of the figures of merit (FOM) of an advanced and miniaturized transistor becomes a challenge when its f_{MAX} goes above many hundreds of GHz. In fact, the quantities to be measured become smaller and smaller and thus the influence of the measurement environment becomes less and less negligible. Indeed, when measuring the same test structures using two different renowned commercial probes having a different topology, a "signature" of each probe can been observed, in particular when plotting $\sqrt{U} \times freq$ (U is the Mason gain) as a function of frequency, which is usually carried out for f_{MAX} estimation. For mmW technologies, f_{MAX} is the key figure of merit for benchmarking technologies, thus it becomes urgent to clarify this measurement. In this work, we give a proof that measurement becomes probe dependent. We constructed an accurate EM model of each probe using X ray tomography and thus simulated the measurement environment at close proximity of the wafer using EM and SPICE simulation. Hence, each signature of the probe is clearly reproduced by the simulation highlighting that the unexpected result is not due to an inaccuracy done by the user such as probe positioning or to a limitation of the VNA but is the result of the unwanted coupling between probe and substrate or between probes and the inability of the SOLT and on-wafer TRL calibration algorithms to completely remove these couplings.**

***Index Terms*— RF probes, on-wafer measurement, S-parameters, mmW, f_{MAX} determination, SiGe HBT, RF inductors, RF MOSFET, TRL calibration, SOLT calibration**

I. Introduction

A very strong competition is taking place between the different semiconductor foundries to offer technologies that are more and more efficient to meet the demand for millimeter wave applications. One of the key components for a given BiCMOS technology is the SiGe HBT transistor having a very interesting cost-performance ratio. A state of the art of the BiCMOS technologies is given in [1], in which we can note the performance record in pure bipolar technology with a f_{MAX} of 700 GHz. We can also highlight the STMicroelectronics technology with the highest integration node with 55 nm and having a f_{MAX} of 370 GHz. On the table given in [1], the most observed figure of merit is the f_{MAX} value. Indeed, it is an image of the frequency up to which the transistor still amplifies power. Usually, f_{MAX} is deduced from the Mason gain U which is calculated from measured y-parameters. Unfortunately, the f_{MAX} value is extremely difficult to determine as it is mentioned in [2] due to different reasons: invalidity of the single pole model used to calculate f_{MAX}, measurement noise due to small quantities that needs to be measured and especially measurement inaccuracy due to calibration algorithm and measurement environment.

The most prominent example is the measurement of the world-record SiGe HBT given in [3] showing f_{MAX} determination of the same device measured at two different laboratories giving similar results below 40 GHz but having different trends above 40 GHz. In fact, high frequency measurement on silicon technology is still a challenge, even between 40 and 110 GHz when analyzing sensitive parameters such as f_{MAX}.

At this point, we can also mention that the same difficulties are encountered to extract f_{MAX} of RF MOS technologies such as FD SOI technologies [4]. Recently a study has been carried out [5] showing the uncertainty in extracting the f_{MAX} in the lower frequency regime (below 80 GHz) for the two most advanced CMOS processes.

To go further, we can also reveal the difficulties related to the characterization of certain passive elements with, for example, the extraction of the quality factor of an inductor which can be extremely challenging to measure and which is however what is put forward by the foundries. Finally, this inherent difficulty to extract precisely the FOM also hampers the development of accurate compact models and design kits.

All the S-parameter measurements are usually performed on-wafer directly under ground-signal-ground (GSG) probes with a network analyzer operating at up to 110 GHz. Beyond the choice of the equipment, i.e. probe station and network analyzer, the choice of the probe topology is a major concern. In fact, different couplings between probes and between probe and substrate that can be correlated to the architecture of the probes and layout design [6]–[12].

In this paper, we compare measurements carried out with two industrials probes: FormFactor Infinity and GGB-Picoprobes. Both probes have very different topologies. We apply two types of calibrations: SOLT on alumina calibration kit and TRL on-wafer followed by a standard de-embedding

Manuscript received May 25, 2021. This work is partly funded by the French Nouvelle-Aquitaine Authorities through the FAST project and NANO2022 IPCEI project. The authors also acknowledge financial support from the EU under Project Taranto (No. 737454). The authors would like to thank STMicroelectronics for supplying the Silicon wafer.

Sebastien Fregonese, Magali De Matos, Marina Deng, , and Thomas Zimmer are with the IMS Laboratory, CNRS and University of Bordeaux France (e-mail: sebastien.fregonese@ims-bordeaux.fr).

Didier Céli, Nicolas Derrier are with STMicroelectronics, Crolles, France.

procedure (open-short for SOLT: after SOLT the reference plane is at the probe tip; and short-open for TRL: after TRL, the reference plane is after the pads). These results are validated through a simulation procedure in which an EM simulation is performed for each test structure (pad-open, pad-short, pad-load, thru, lines, transistor-open and transistor-short). For the transistor, an EM-SPICE co-simulation is carried out. Finally, the simulation results are treated in the same way as the measurement data are processed (calibration and de-embedding). Moreover, compared to the probe models proposed in [1], we have improved our models by using tomographic images. Finally, this paper aims to explain the unexpected trend of the $\sqrt{U} \times freq$ which is used to determine f_{MAX}.

The paper is organized as follows: I) the layout, the probes and their model are presented; II) the measurement and simulation results are presented and explained, then III) results are correlated to the probe topology.

II. Layout Design and Probes Topologies

A. Layout

The fabricated test structures are dedicated to on-wafer TRL calibration. They have a large space in-between the structures and a staggered arrangement of the test structures is implemented to reduce the influence of adjacent structures. A detailed description can be found in [12], [13]. Different research labs [10], [14], [15] [7] have shown that measurement results can be influenced by the adjacent structures, resulting in erroneous S-parameters. On the other hand, advanced technologies have a high cost per square millimeter which imposes to increase the density of the test structures used for device compact modelling. Nonetheless, in our specific case the space between DUTs is very large (Δx=207 µm, Δy=133 µm) which reduces the impact of neighbors [16].

Also, a continuous ground plane was used between structures as recommended in [17], which suppresses the penetration of the EM wave into the substrate.

Eventually, the pad layout design combined with the probe geometry can influence the probe-to-substrate coupling as demonstrated in [11] and may modify the FOM determination. A backside ground-wall behind the signal pad has been introduced. Also, the pad geometry has been optimized to be compatible with 100 and 50 µm pitch for millimeter-wave and sub-millimeter-wave measurements (see Fig.1 and Fig.2).

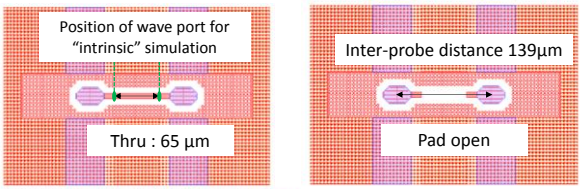

Fig.1. Layout of the principal test-structures highlighting distance between probes

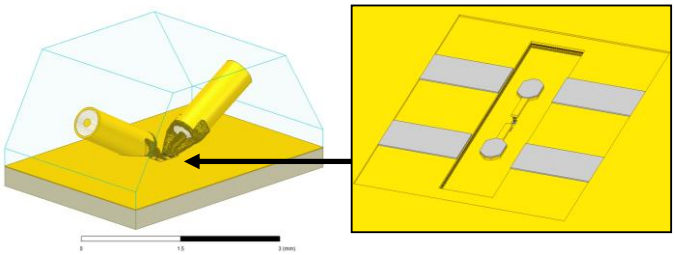

Fig.2. EM simulation set-up with Picoprobe probes and 3D structure of the LOAD showing the pad design (yellow is copper and grey is aluminum, dielectric is removed for clarity of figure)

B. Probes topologies

When choosing the probes, the three main characteristics that are observed are the insertion loss, the return loss and the DC contact characteristics (resistance, stability over time). But there is one characteristic that plays a major role when analyzing crosstalk and probe to substrate coupling: it concerns the ability of the probe to focus the EM field coming out of the micro-coaxial cable towards the device to be measured while limiting the stray field.

To build the EM models, we used an X-Ray tomography equipment from Carl Zeiss Versa 500. This equipment can measure the 3D structure with a 0.15µm accuracy that can be imported in HFSS EM tools after few steps of post data processing. The 3D image coming out from the X-ray tomography tool and imported in HFSS is given in Fig.3. For comparison, the X-Ray tomography-based EM model for the Picoprobe is also given in Fig.3, with the same scale. The tomography imaging technique is exactly depicting the geometry of the probe. The only limit concerns the uncertainty of the materials surrounding the metal.

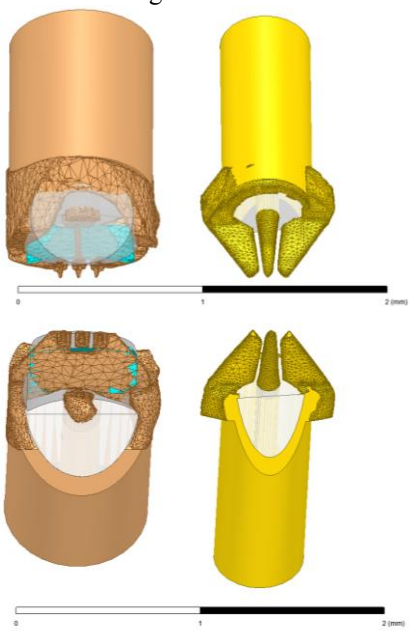

Fig.3. CAD model based on tomography image of Infinity probe (left) and Picoprobe (right) , top view and bottom view– The scale is the same in both cases. Left image: brown is copper, blue is polyimide, transparent is epoxy glue and white is Teflon; right image: yellow is Beryllium-Copper BeCu, white is teflon

III. EM Simulation and Measurement Setup

As described in [7], [9], [19] and in order to mimic the measurement environment, the probes are simulated together with the pads, access lines and intrinsic structures (see Fig. 2). An air-box is applied to completely cover this set up. The absorbing radiations type boundary condition is assigned on the surface of the air-box to emulate the infinite space like environment. The "wave port" is used to provide the excitation at each port on probe side. Two internal lumped ports are used in the case on the transistor to plug the SPICE model [19]. The simulation is performed for thru (95µm including access lines), reflect, line 1 (625µm), load, transistor-open and transistor-short. The set of lines allows an accurate TRL calibration from 20 GHz to 110 GHz. The whole set of simulation data is computed using exactly the same procedure as done in the measurements and applying a TRL calibration with impedance correction using the load. Moreover, in order to validate the accuracy of the calibration procedure, the intrinsic structures are also simulated (EM) with a reference plane placed after the pad (see Fig. 1). This is the same reference plane than the TRL procedure or SOLT procedure with de-embedding. Hence, the accuracy of the calibration procedure is validated if the calibrated data match with the intrinsic data. In the whole paper, the intrinsic data is used as a reference and is plotted in black dashed line, the measurement with calibration (and eventually de-embedding) are in symbols and the emulation of the calibration procedure (and eventually de-embedding) using EM(-SPICE) simulation including probes is in solid line.

Concerning the measurement part, a E8361A Vector Network Analyzer (VNA) from Agilent was used working up to 110 GHz and using extenders (N5260-60003) above 67 GHz. The intermediate frequency (IF) was set to 10 Hz and the power was adjusted at about -32 dBm up to 67 GHz and maintained lower than -30dBm for above frequencies, using mechanical attenuators. This power was chosen to conserve linearity and avoiding self-biasing since these test structures are dedicated to HBT measurement. Then, with the two probe types (Picobrobe 110 GHz and FormFactor Infinity 110 GHz) measurements were performed on the 65µm thru, the 589µm line and the reflect (for calibration purpose). The transistor-open, a transistor-short and a transistor were chosen as DUT (devices under test). For SOLT calibration, the calibration kit CS15 from GGB-industries is selected with its default parameter for the Picoprobe as well as the Infinity probe because an accurate EM model was built for the CS15 calibration substrate [9]. When comparing SOLT and TRL, exactly the same input data are employed.

IV. Results

A. Passives elements

For all the passives structures (test-structures of transistor-open and transistor-short), a comparison is given between: i) the simulated intrinsic structure EM; (ii) the measurement both with the on-wafer TRL and SOLT calibration on alumina substrate followed by pad-open/pad-short de-embedding to set the reference plane just after the pad (see fig. 2) iii) EM simulation including the probe model where the same procedure is applied as for measurements.

First the transistor-open is studied and the capacitances have been extracted considering the well-known π-model $C_{11} = imag(Y_{11} + Y_{12})/(2\pi\, freq)$, $C_{12} = -imag(Y_{12})/(2\pi\, freq)$, $C_{22} = imag(Y_{22} + Y_{12})/(2\pi\, freq)$, see Fig.4 and Fig.5). We obviously observe that if the intrinsic simulation gives constant capacitances over frequency, this is not the case when detailing the measurement data or simulation data including the probes. For the case of the TRL (see Fig.4) with focus on C_{12}, one can clearly see that the crosstalk is not corrected and is different from one probe to the other. At the low frequencies of the band, the coupling within the Infinity probe appears to be lower, but at 40 GHz, a sudden drop of C_{12} indicates that an important coupling is occurring. This is correlated to a drop in the magnitude of S11 which can be explained by a probe to substrate coupling. A second drop arises on C_{12} at 100 GHz but this is not confirmed by simulation and may be due to other reasons. In the case of the Picoprobe (see Fig.4 (C12), right), at the low frequencies of the band, the crosstalk is slightly higher but more homogeneous over the frequency. Around 100 GHz, C_{12} is decreasing showing that a coupling arises around 100 GHz.

C_{11} and C_{22} are more constant in the case of the Picoprobe compared to the Infinity. These conclusions are confirmed by the EM simulations which generally shows the same trend.

In the case of the SOLT calibration (see Fig.5), the C_{12} capacitance measurement is more accurate. This is attributed to the twelve terms error-model which takes partly into account the crosstalk correction. The C_{12} value is nearly constant over frequency and the three data (measurement and simulations) are almost superimposed. Unfortunately, we can notice that C_{11} and C_{22} are underestimated. This is attributed to the de-embedding procedure which performs over-de-embedding due to the fringe capacitance at the end of the pad-open test-structure of the line access (see Fig.8a). The latter does not exist in the case of the transistor-open and thus should not be removed from the transistor-open capacitance. The fringe capacitance of the output of the pad (cross section of the line access) is about 0.8fF. These results are confirmed by the EM simulation.

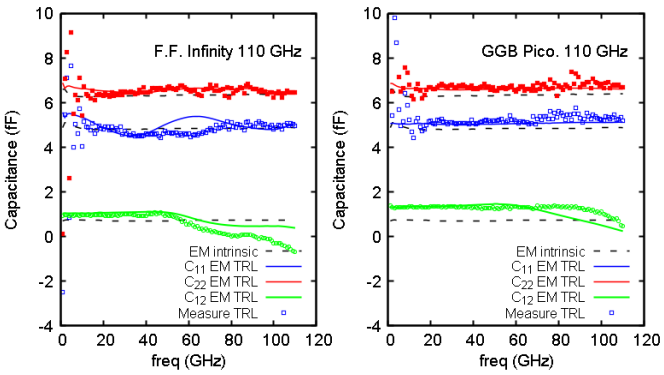

Fig.4. Capacitance of the transistor-open using the π-model versus frequency. Calibration is TRL, comparison between the Infinity and Picoprobe probe.(Dashed line is intrinsic EM simulation, solid line is the result of the simulation of all the test structures where TRL was applied, symbols are measurement data after TRL calibration; applicable to fig.5 to 10)

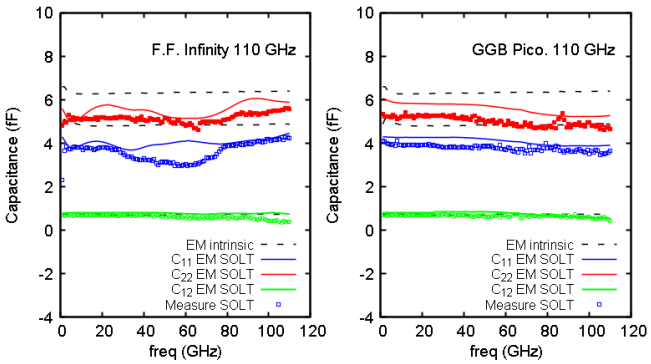

Fig.5. Capacitance of the transistor-open using the π- model versus frequency. Calibration is SOLT followed by open-short de-embedding, comparison between the Infinity and Picoprobe probe.

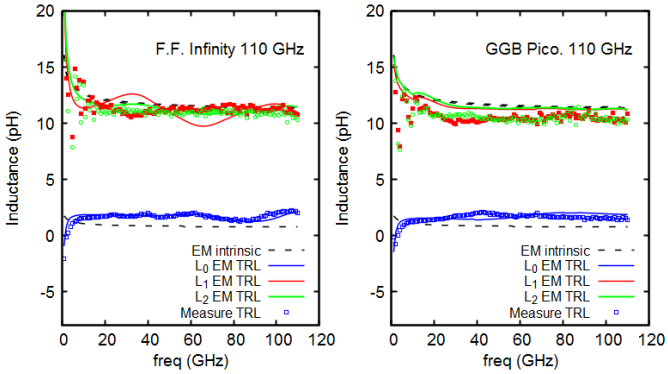

Fig.6. Inductance of the transistor-short using the star model versus frequency. Calibration is TRL, comparison between the Infinity and Picoprobe probe.

Second, a transistor-short is analyzed, see Fig.6, Fig.7, Fig.9 and Fig.10. We have used a T-model to extract the inductors and resistors: $L_1 = imag(Z_{11} - Z_{12})/(2\pi\, freq)$,

$L_0 = imag(Z_{12})/(2\pi\, freq)$, $L_2 = imag(Z_{22} - Z_{21})/(2\pi\, freq)$,

$R_1 = real(Z_{11} - Z_{12})$, $R_0 = real(Z_{12})$, $R_2 = real(Z_{22} - Z_{21})$.

In the case of the TRL (see Fig.6), L_1 and L_2 inductors from measurements are fully superimposed with the intrinsic simulation. The common inductor L0, which is modelling the return path is not perfectly captured for both probes with a difference of about 0.5pH between the reference data and the calibrated data due to the return to ground (see Fig.8b). Again, the EM simulation is well superimposed with the measurements except for L_2 of the Infinity probe which presents an artefact of simulation (not explained at this stage). In the case of the SOLT (see Fig.7), the common inductor L0 is better captured but L_1 and L_2 are shifted. This is once more attributed to over-de-embedding which is due to the small piece of line which connect the signal to grounds (return path from signal to ground (~2pH)).

Looking at the resistor side, after the TRL calibration (see Fig.9) the R_1 and R_2 resistor values are correctly determined but this is not the case for the return path R_0. The behavior is smoother in the case of the Picoprobe while the Infinity probe shows a more "wavy" behavior with frequency. This is confirmed by simulation. In the case of R_0, a first drop of R_0 is observed around 40 GHz and then a second drop around 90 GHz resulting in a negative value, both reproduced by simulation. The Picoprobe gives a more reasonable value over the whole frequency band. The EM model is slightly shifted. We attribute this shift to the fact that the model assumes rigid probe tips whereas the probes are a bit bended when they touch the pads. In the case of the SOLT (see Fig.10), the results after calibration with Infinity probes are closer to the intrinsic ones compared to Picoprobe probes but they start to fail above 80 GHz. In the case of the Picoprobe, a 0.5 Ω shift is observed on the whole frequency band. In both probes, the R0 resistor is well measured and this is due to the SOLT algorithm.

The preliminary conclusion is:

- A more homogeneous behavior over frequency is observed in the case of the Picoprobe compared to the Infinity which shows more artefacts. These are attributed to the topology and geometry of the probe and the coupling within this specific layout.
- The TRL gives more reasonable results for C_{11}, C_{22}, L_1, L_2, R_1 and R_2 but is slightly shifted in the case of the crosstalk capacitance C_{12} and return path of L_0 and R_0. The orders of magnitude of these values are very small since we are talking about hundreds of aF, a few tenths of pH and hundreds of milliohms.
- The SOLT calibration corrects partly the cross-talk and return path but the de-embedding procedure (required because of the off-wafer calibration) introduce a small error which becomes visible for extremely down-scaled devices and which are attributed to fringe capacitances or to the return path to ground which are both usually neglected.

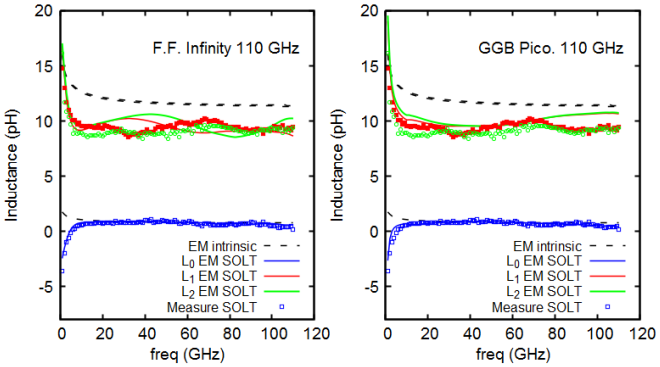

Fig.7. Inductance of the transistor-short using the star model versus frequency. Calibration is SOLT followed by open-short de-embedding, comparison between the Infinity and Picoprobe probe.

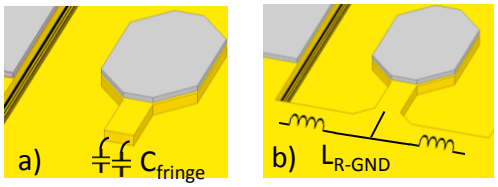

Fig.8. a) Fringe capacitance of the pad-open; b) return to ground inductance: both elements induce over-de-embedding since C_{fringe} and $L_{R\text{-}GND}$ does not exist in transistor open and transistor-short.

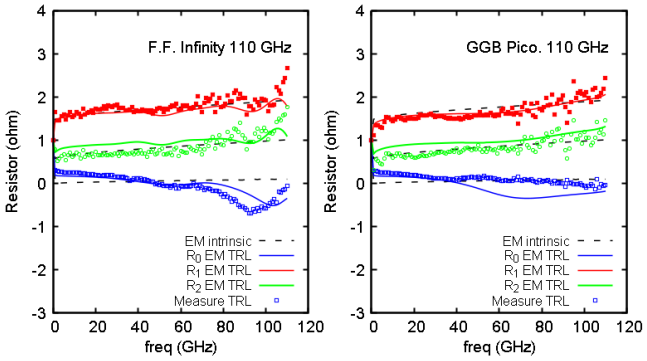

Fig.9. Resistance of the transistor-short using the star model versus frequency. Calibration is TRL, comparison between the Infinity and Picoprobe probe

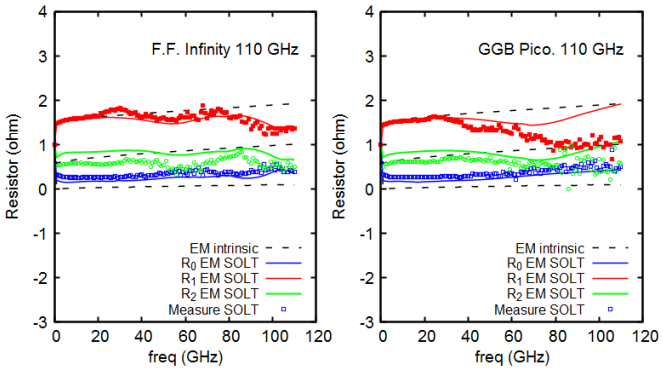

Fig.10. Resistance of the transistor-short using the star model versus frequency. Calibration is SOLT followed by open-short de-embedding, comparison between the Infinity and Picoprobe probe.

B. Transistors

In the case of the transistor, the issues encountered for the passive structures related to de-embedding with respect to the fringe capacitance as discussed above is no more a problem since we do not use the pad-open test structure for de-embedding, but the transistor-open. Also, the pad-short is not used anymore. The de-embedding with the transistor-open and transistor-short allows to set the reference plane at metal 1.

Initially, we are particularly interested in S12 (see Fig.11) because it is strongly impacted by the crosstalk. Up to 40 GHz, the measurement of the magnitude of S12 remains reasonable using a SOLT or a TRL. While the intrinsic simulation grows continuously, a drop-off appears at 40 GHz in the case of Infinity probes for both, the measurement and the EM-SPICE simulation. This is followed by a plateau up to 80 GHz where a second decrease appears. The decrease is less pronounced in the case of Picoprobe probes but the difference between the intrinsic simulation and the co-simulation augments as the frequency increases from 60 GHz. The simulation results with probes are quite good in the case of the Infinity probe up to 80 GHz but less good in the case of the Picoprobe probe. In fact, during measurements, by contacting the pads, a geometrical deformation brings the tip closer to the substrate and leads to a higher coupling. During EM simulation, this deformation is not considered. The Picoprobe probes are much more sensitive to the deformation compared to Infinity probe where not deformation arises in the front-end part.

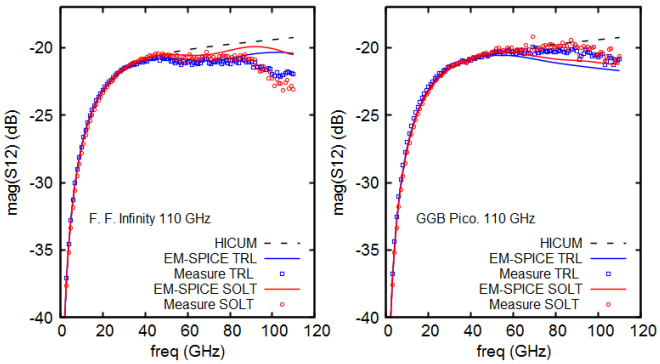

*Fig.11. Measurement of magnitude of S12 versus frequency of the transistor (A_E=0.09*5µm²) at V_{BE}=0.9 BV and V_{CB}=0V. Dashed line is the HICUM compact model. [Applicable for Fig. 11 to 13: Symbols are measurement; solid line are EM simulations with probes for the set of test structures: blue is the TRL + short-open de-embedding and red is SOLT followed by open-short de-embedding.]- De-embedding set reference plane at M1.*

This specific signature that is found on the S_{12} measurement and which was also observed on the measurement of C_{12} of the transistor-open has a strong impact on the FOM such as the f_T and f_{MAX} (see Fig.12 and Fig.13). The Infinity probes show a vale from 40 GHz to 50 GHz followed by an ascent and second dropout appears at 90 GHz. In the case of the Picoprobe, the behavior is smoother but the measurement gradually deviates as the frequency increases. The simulation reproduced the measurement with reasonable accuracy and even the surprising trends that are observed during the measurements for both probe types; even if the accuracy is less good in the case of the Picoprobe.

The intermediate conclusions are:

- f_T / f_{MAX} can be determined from measurements below 40 GHz only; above the unexpected behavior is due to not corrected coupling and crosstalk which are probe and layout specific. Please note that these unexpected "wavy" data are the results of a coupling between the probes and the test structures and that these unexpected data cannot be only attributed to the probes.
- Some of the measurement such as S_{12} or phase of H_{21} (not shown) start to deviate from 40 GHz. This can lead to erroneous SPICE modelling work. If most of the model parameters are static or quasi-static and can be extracted below 40 GHz, some of the parameters such as distributed effects like NQS or like capacitance split in the base or the substrate require accurate measurements far beyond 40 GHz.

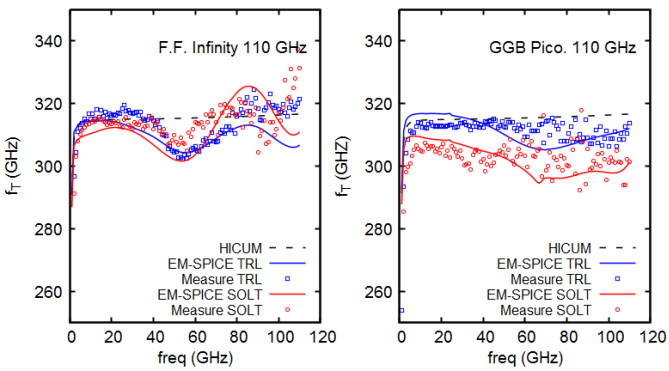

*Fig.12. Determination of f_T (defined as H_{21}*freq) versus frequency for the Formfactor Infinity probe and for the GGB Picoprobe probe. Transistor: (A_E=0.09*5µm²) at V_{BE}=0.9 BV and V_{CB}=0 V. After de-embedding, the data still includes Metal 1 fingers.*

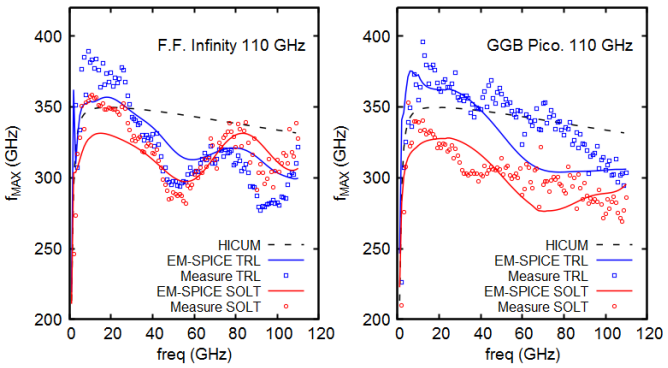

*Fig.13. Determination of f_{MAX} ($U^{1/2}$*freq) versus frequency for the Formfactor Infinity probe and for the GGB Picoprobe: Transistor : (A_E=0.09*5µm²) at V_{BE}=0.9 V and V_{CB}=0 V. After de-embedding, the data still includes Metal 1 fingers.*

V. ANALYSIS

The previous part has clearly shown the incapability of the TRL and SOLT algorithm to fully remove the influence of the measurement environment. It is well known while the TRL has many advantages in terms of accuracy and its ability to work with less input parameters, its limitation is that the crosstalk is not considered. In the case of the SOLT calibration, with the 12 errors terms only isolation is accounted for through EXF and EXR terms. In fact, these errors terms are connecting the error box from the input of port 1 to the output of port 2: it means that it cannot take into account the crosstalk at the probe level. Moreover, this correction is done with a calibration kit having a different EM environment. Hence, the two algorithms do not take into account the distributed nature of coupling and crosstalk all along the probe. Let's detail on each probe how crosstalk and coupling are acting:

-In the case of the Infinity probe, we can first see that the input micro-coaxial cable has a larger diameter than the Picoprobe one, which means that the EM field is less confined at the interface between the coaxial and the micro-strip line. At this interface, i.e. at the solder joint, appears a coupling from the signal to the test structure GND plane (see Fig.14). This solder joint is 176µm above the GND plane and at about 500µm away from the probe tips in the horizontal direction. This coupling induces a path between the signal (at the coax output) towards the GND of the pyramidal tips through the on-wafer ground plane. It is worth to mention that within the microstrip line the coupling to the substrate is completely canceled. The second effect is the crosstalk which is acting from the probe tips (having a pyramid like form) at port 1 to them at port 2. This is illustrated in figure 14 where the E-field is plotted at 60 GHz. As the wave propagates along the probe, a coupling occurs in turn between the welding point and the substrate for a phase of 130°.

- In the case of the Picoprobe probe, the micro-coaxial cable is smaller inducing a better confinement of the EM field at the transition with the CPW line. Nevertheless, a coupling arises at this interface (see Fig.15). While with the Infinity probe the coupling is localized at the soldering point, with the Picoprobe the coupling is distributed all along the CPW down to the contact. In a similar way the crosstalk from port 1 to 2 is also distributed. This is illustrated in figure 15 where the E field is plotted at 60 GHz.

In order to verify these statements and to decouple the crosstalk part from the coupling to the substrate part, we will in turn cut the ground plane below the probe to reduce the substrate coupling and then insert an insulating box between port 1 and port 2. We will focus on the S11 measurement of a transistor-open of both probes for analyzing probe to substrate coupling (see Fig.16) with a reference plane at the tips, i.e. without de-embedding in order to highlight the limitations:

- A localized and strong resonance appears at 60 GHz for the Infinity probe (observed in simulation and measurement) which is explained by a localized coupling to substrate at the solder joint. This can be proven by cutting the GND plane below the solder joint and as shown in Fig.16, this strongly reduces the resonance. A similar resonance effect was also observed in the work from [20] despite the different test structure layout.
- For the Picoprobe, the mag of S_{11} decrease progressively as the frequency increases and it is due to a more distributed nature of coupling. Cutting the GND plane below the probe shows less difference with the original layout than in the case of the Infinity Probe. We can conclude that this probe is less prone to substrate coupling.

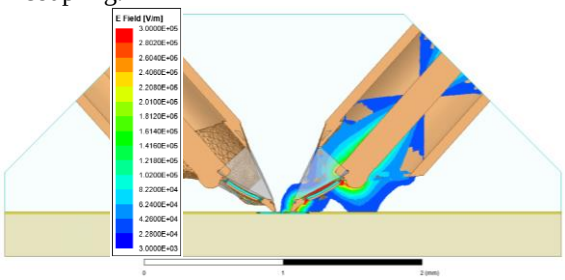

Fig.14. Simulation of a pad-load with Infinity probes with E-field contour (a) frequency = 60 GHz, EM wave phase= 130° highlighting coupling to substrate

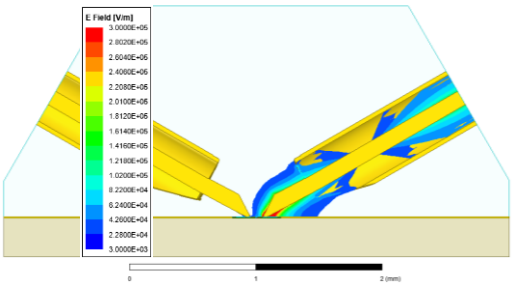

Fig.15. Simulation of a pad-load with Picoprobe probes with E-field contour highlighting crosstalk and coupling ; frequency = 60 GHz, angle= 130°

Now we will focus on S_{12} of the transistor-open, to analyze the crosstalk effect: contrary to popular belief, crosstalk starts already at low frequencies. This is demonstrated by the introduction of an "insulating" box (epsr=0.001) between the probe tips which aims to suppress the crosstalk in air. Figure 17 shows indeed a not negligible difference between the simulations without "insulating" box and this for both probe types.

-In the Infinity probe case, we observe a first effect from 10 GHz to 60 GHz which is due to crosstalk and starting from 60 GHz the S12 behavior is more affected by the probe to substrate coupling than by the crosstalk.

- A similar effect is observed with the Picoprobe probe. But in this case, the crosstalk is stronger. Adding the insulating box allows to strongly reduce this effect over the whole frequency band.

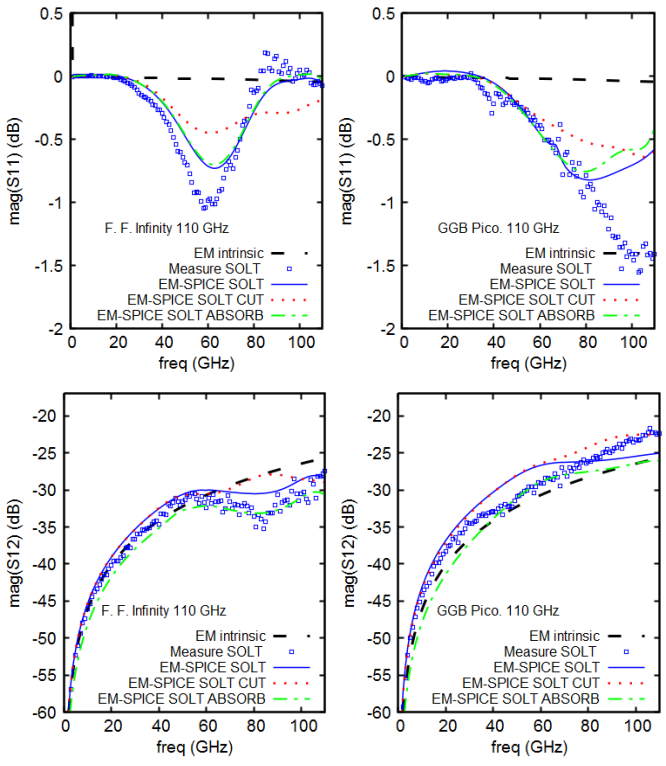

Fig.16. Evaluation of importance of probe to substrate coupling on the measurement of a transistor-open: Measurement of magnitude of S11 & S12 versus frequency for both probe topologies. Dashed line is the intrinsic EM simulation. Symbols are measurement; solid line is EM simulations with probes + SOLT; red dot is EM simulations simulation + SOLT but the GND plane is cut below the probe, green dashed is EM simulations simulation + SOLT but having an absorber/insulator between the 2 probes to reduce crosstalk in air. (reference plane is at probe tips in this case)

VI. Conclusion

The unexpected pattern observed in the f_{MAX} determination of SiGe HBTs was studied with two commercial RF probe topologies. Indeed, this study demonstrates that the observed f_{MAX} pattern is strongly correlated to the association of the on-wafer test-structure layout with the probe topologies but not to the intrinsic device itself or to any measurement inaccuracies. The analysis has been performed applying two different standard calibration procedures (SOLT and TRL) highlighting the advantages and defaults of each one to correct for the different coupling schemes. Summarizing the different couplings that can be observed on each probe, we can say that even if crosstalk and probe to substrate coupling are present in both probe technologies, the Infinity probes are more affected by probe to substrate coupling in the case of a continuous ground plane between DUTs while the Picoprobe probes are more prone to crosstalk. The distributed nature of crosstalk together with coupling in the probes used in the study is the red brick wall which prevents to assess the intrinsic device performances of high frequency SiGe HBTs. These results are applicable to advanced RF SOI MOSFETs or to high frequency small inductors. Possible solutions to solve this problem are either the use of a more complex calibration method (e.g. 16-term error model) [21]–[23] or of better EM-field confining probes to limit coupling between probes or with the substrate.

Acknowledgment

We would like to thank Patrick Scheer, and Benjamin Dormieu, Cédric Durand from STMicroelectronics for fruitful discussion.

References

[1] P. Chevalier *et al.*, « Si/SiGe:C and InP/GaAsSb Heterojunction Bipolar Transistors for THz Applications », *Proc. IEEE*, vol. 105, nº 6, p. 1035-1050, juin 2017, doi: 10.1109/JPROC.2017.2669087.

[2] B. Saha *et al.*, « Reliable Technology Evaluation of SiGe HBTs and MOSFETs: fMAX Estimation From Measured Data », *IEEE Electron Device Lett.*, vol. 42, nº 1, p. 14-17, janv. 2021, doi: 10.1109/LED.2020.3040891.

[3] B. Heinemann *et al.*, « SiGe HBT with fx/fmax of 505 GHz/720 GHz », in *2016 IEEE International Electron Devices Meeting (IEDM)*, déc. 2016, p. 3.1.1-3.1.4. doi: 10.1109/IEDM.2016.7838335.

[4] F. Arnaud *et al.*, « Truly Innovative 28nm FDSOI Technology for Automotive Micro-Controller Applications embedding 16MB Phase Change Memory », in *2018 IEEE International Electron Devices Meeting (IEDM)*, déc. 2018, p. 18.4.1-18.4.4. doi: 10.1109/IEDM.2018.8614595.

[5] J. Rimmelspacher, A. Werthof, R. Weigel, A. Geiselbrechtinger, et V. Issakov, « Experimental Considerations on Accurate fT and fmax Extraction for MOS Transistors Measured Up to 110 GHz », in *2019 92nd ARFTG Microwave Measurement Conference (ARFTG)*, janv. 2019, p. 1-4. doi: 10.1109/ARFTG.2019.8637249.

[6] C. B. Sia, « Minimizing discontinuities in wafer-level sub-THz measurements up to 750 GHz for device modelling applications », in *2017 89th ARFTG Microwave Measurement Conference (ARFTG)*, juin 2017, p. 1-4. doi: 10.1109/ARFTG.2017.8000843.

[7] S. Fregonese *et al.*, « On-Wafer Characterization of Silicon Transistors Up To 500 GHz and Analysis of Measurement Discontinuities Between the Frequency Bands », *IEEE Trans. Microw. Theory Tech.*, p. 1-10, 2018, doi: 10.1109/TMTT.2018.2832067.

[8] K. Yau, E. Dacquay, I. Sarkas, et S. P. Voinigescu, « Device and IC Characterization Above 100 GHz », *IEEE Microw. Mag.*, vol. 13, nº 1, p. 30-54, janv. 2012, doi: 10.1109/MMM.2011.2173869.

[9] S. Fregonese *et al.*, « Comparison of On-Wafer TRL Calibration to ISS SOLT Calibration With Open-Short De-Embedding up to 500 GHz », *IEEE Trans. Terahertz Sci. Technol.*, vol. 9, nº 1, p. 89-97, janv. 2019, doi: 10.1109/TTHZ.2018.2884612.

[10] C. Andrei, D. Gloria, F. Danneville, P. Scheer, et G. Dambrine, « Coupling on-wafer measurement errors and their impact on calibration and de-embedding up to 110 GHz for CMOS millimeter wave characterizations », in *2007 IEEE International Conference on Microelectronic Test Structures*, mars 2007, p. 253-256. doi: 10.1109/ICMTS.2007.374494.

[11] M. Seelmann-Eggebert *et al.*, « On the Accurate Measurement and Calibration of S-Parameters for Millimeter Wavelengths and Beyond », *IEEE Trans. Microw. Theory Tech.*, vol. 63, nº 7, p. 2335-2342, juill. 2015, doi: 10.1109/TMTT.2015.2436919.

[12] M. Cabbia, C. Yadav, M. Deng, S. Fregonese, M. De Matos, et T. Zimmer, « Silicon Test Structures Design for Sub-THz and THz Measurements », *IEEE Trans. Electron Devices*, vol. 67, nº 12, p. 5639-5645, déc. 2020, doi: 10.1109/TED.2020.3031575.

[13] C. Yadav, « Importance and Requirement of frequency band specific RF probes EM Models in sub-THz and THz Measurements up to 500 GHz », *IEEE Trans. Terahertz Sci. Technol.*.

[14] G. N. Phung, F. J. Schmückle, et W. Heinrich, « Parasitic effects and measurement uncertainties in multi-layer thin-film structures », in *2013 European Microwave Conference*, oct. 2013, p. 318-321. doi: 10.23919/EuMC.2013.6686655.

[15] F. J. Schmückle, T. Probst, U. Arz, G. N. Phung, R. Doerner, et W. Heinrich, « Mutual interference in calibration line configurations », in *2017 89th ARFTG Microwave Measurement Conference (ARFTG)*, juin 2017, p. 1-4. doi: 10.1109/ARFTG.2017.8000823.

[16] S. Fregonese, M. Deng, M. Cabbia, C. Yadav, M. D. Matos, et T. Zimmer, « THz Characterization and Modeling of SiGe HBTs: Review (Invited) », *IEEE J. Electron Devices Soc.*, vol. 8, p. 1363-1372, 2020, doi: 10.1109/JEDS.2020.3036135.

[17] D. F. Williams et R. B. Marks, « Transmission line capacitance measurement », *IEEE Microw. Guid. Wave Lett.*, vol. 1, nº 9, p. 243-245, sept. 1991, doi: 10.1109/75.84601.

[18] « infinity-probe-mechanical-layout-rules.pdf ». Consulté le: mars 22, 2021. [En ligne]. Disponible sur: https://www.formfactor.com/download/infinity-probe-mechanical-layout-rules/?wpdmdl=1850&refresh=60580ade1255e1616382686

[19] S. Fregonese *et al.*, « Analysis of High-Frequency Measurement of Transistors Along With Electromagnetic and SPICE Cosimulation », *IEEE Trans. Electron Devices*, vol. 67, nº 11, p. 4770-4776, nov. 2020, doi: 10.1109/TED.2020.3022603.

[20] N. Waldhoff, C. Andrei, D. Gloria, S. Lepilliet, F. Danneville, et G. Dambrine, « Improved Characterization Methology for MOSFETs up to 220 GHz », *IEEE Trans. Microw. Theory Tech.*, vol. 57, nº 5, p. 1237-1243, mai 2009, doi: 10.1109/TMTT.2009.2017359.

[21] X. Wei, G. Niu, S. L. Sweeney, Q. Liang, X. Wang, et S. S. Taylor, « A General 4-Port Solution for 110 GHz On-Wafer Transistor Measurements With or Without Impedance Standard Substrate (ISS) Calibration », *IEEE Trans. Electron Devices*, vol. 54, nº 10, p. 2706-2714, oct. 2007, doi: 10.1109/TED.2007.904362.

[22] J. V. Butler, D. K. Rytting, M. F. Iskander, R. D. Pollard, et M. Vanden Bossche, « 16-term error model and calibration procedure for on-wafer network analysis measurements », *IEEE Trans. Microw. Theory Tech.*, vol. 39, nº 12, p. 2211-2217, déc. 1991, doi: 10.1109/22.106567.

[23] X. Wei, G. Niu, S. L. Sweeney, et S. S. Taylor, « Singular-Value-Decomposition Based Four Port De-embedding and Single-step Error Calibration for On-chip Measurement », in *2007 IEEE/MTT-S International Microwave Symposium*, juin 2007, p. 1497-1500. doi: 10.1109/MWSYM.2007.380537.