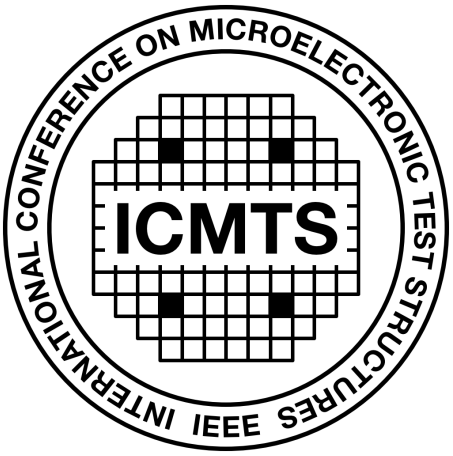
SESSION 1
Layout Dependent Effects



16th April 2024, 09:10–10:10

Session Co-Chairs: Yoshio Mita, *University of Tokyo, Japan*
Yuzo Fukuzaki, *Rapidus US, USA*

Layout Dependent Effects Of Passive Devices And Their Impact On Analog Integrated Circuits

A. Jayakumar, M. van Dort, M. Vertregt, G.D.J. Smit, R. Lander, I. Liu and P. Volf
NXP Semiconductors B.V, Gerstweg 2, 6534 AE Nijmegen, Netherlands
Email: aarthi.jayakumar@nxp.com

***Abstract*—Analog blocks on products built using advanced CMOS technologies were seen to have deviating behavior on silicon than estimated by post layout simulations, especially in circuits used for biasing and reference generation. Circuit investigations pointed towards trimming resistor banks used for bias and reference current derivation. To detect possible unmodelled silicon effects experienced by these trimming resistor banks which have different configurations in the various analog blocks on products, a scribe-line test structure was implemented, aimed to capture major product design use-cases and study their effects. The outcome of this study is presented in this paper.**

Keywords—bias current generation, resistor banks, device models, layout dependent effects, model to silicon gaps

I. INTRODUCTION

Analog integrated circuits designed and implemented using advanced CMOS technology nodes are often seen to encounter unmodelled silicon effects of individual circuit components, causing the circuits to operate outside the expected electrical specification, thereby affecting their overall performance. Layout dependent effects (LDE) are often seen to be one among those commonly unmodelled effects. While extensive studies have been undertaken to study their influences on active devices [1]-[2], there are still gaps observed in capturing these effects for passive devices. One of the most fundamental passive devices from a CMOS process used in analog circuits such as RC filters, reference voltage and current generation circuits, data converters etc. are resistors.

For obtaining voltage references, a commonly used circuit is a bandgap voltage reference circuit as shown in Fig.1(a). Upon establishing a voltage reference, it is used in combination with resistors to generate reference and bias currents, as seen in Fig.1(b). Resistor banks are employed in trimming circuits that form a part of such bias and reference circuits. In all these cases, any deviation in the value of the integrated resistors used in these circuits will affect the quality and precision of the generated bias and reference current and voltage outputs [3]. In our case, the circuit deviations were a result of deviations in the narrow-width resistor banks used for trimming purposes. The deviation was confirmed using nano-probing technique.

Based on the observed circuit deviations, we investigated the different configurations of trimming resistor banks built out of unit resistors, employed in various analog blocks and systematically examined the layout environments in their realization for capturing their associated effects on silicon, as relevant for our product designs. Using this we implemented a layout DOE by means of engineering test structures to identify and understand the root cause behind the observed deviation. Similar engineering test structures comprising of analog sub-circuits of interest for root cause investigations and silicon debugging have been demonstrated in [4].

II. TEST STRUCTURE DESIGN

A. Motivation For A Screening Design

Prior to the creation of a screening design aimed to effectively capture and understand the local layout factors inducing silicon offsets, physical inspections such as nano probing were performed to confirm the observed deviations. In addition, cross-section TEM images were studied for line-width deviations and the associated impact due to geometry effects on the observed offset. While these could be compensated for by applying a optimal proximity correction (OPC) [5], this alone was inadequate to quantify and explain the total silicon offset. Also potential process optimization and retargeting the intrinsic offset contributed by process/technology was not only unable to explain the total deviation on silicon, but was also infeasible, as the device did not have its own dedicated process layer despite a dedicated photomask. So it became imperative to understand the layout dependent effects and its interaction with the process, in order to be able to completely quantify and understand the silicon deviation. And in order to make judicious use of limited silicon space, a screening layout DOE was proposed covering the identified the layout differences and to study their impact.

B. Engineering Test Structure Implementation

The goal was to have a discrete parametric test structure in the limited space available on product silicon, and still have a statistically significant determination of the impact of layout variants of interest. Since the existing device model was derived from only a single finger structure for a given layout scenario, this approach was critical to ensure the silicon results seen in products can be explained. Fig.2 illustrates the structure that was designed to have trimming resistor banks for various combinations of poly width (PO_W), poly finger spacing (PO_S), number of poly fingers (NF), in isolated and nested resistor bank environments and placed in different well configurations depending on the reference domain of the circuit for a given poly length, as relevant for our various product design use cases.

For all these DUTs only a single poly finger was electrically connected for measurement, to be able to compare the layout effects against the structure used for device model build. This experiment was aimed to serve as a screening DOE that used 2^k factorial design approach to study the interaction between the different layout variants (especially, PO_W, PO_S and a known process variable) and investigate the significance of each these on the measured on-silicon resistance and to identify the most dominant layout factors affecting the measurements [6]. This screening experiment was designed predominantly to detect silicon effects relevant

for our products and not aimed to capture and cover all possible silicon effects that could potentially be incorporated into the device model.

The structure was measured by applying a voltage across the resistor unit as seen in the circuits of products and the corresponding sheet resistance (Rs) was calculated for the various duts. The measurement was performed on matrix lots containing process corner splits for two different CMOS processes.

III. SILICON RESULTS

With the measured data, the designed experiment was analyzed with factorial design analysis approach using Minitab software. The impact of poly finger spacing and its interaction with poly finger width, the effects of well dopants, the impact of the number of poly fingers, the effects on resistor banks placed as an isolated DUT versus placed among identical dummies are presented here.

Fig.3 shows a pareto chart of the standardized effects in decreasing order of magnitude that helps to compare the relative magnitude and the statistical significance of main layout factors influencing the measured resistance and the effect of their interactions. The reference line on the pareto displays which effects are significant with a significance level of 0.05 being used to determine the reference line. While the process variable impact (C) is seen to be the most dominant factor as expected, it is followed by poly finger spacing (B) as a statistically significant dominant factor affecting the measured resistance, for a given poly length. The statistically significant two way interaction term of poly width and poly finger spacing (AB) also confirms the relationship between these two factors and their impact on the measured resistance value, for a given poly length. This demonstrates that only accounting for poly width and length in the device model would render it in adequate for actual product design evaluations.

The analysis of this 2^k factorial design essentially helps to derive a mathematical model that establishes a relationship between the experiment outcome (sheet resistance in this case) and the statistically significant factors and their interaction impacting the outcome (layout experiments in this case). Such a derived model is shown as a regression equation in Fig.3. To assess how well this derived model fits the data, goodness of fit statistics, R^2, was used, and it was seen to be at 96.99%. In addition to this, a residual plot approach as shown in Fig.3. has also been used to complement the goodness of fit assessment.

In Fig.4, the effect of each factor (poly width, poly finger spacing, process variable) in the regression equation, the effect of their interactions and their influence on the model outcome (measured resistance) has been illustrated in single plots. This helps to visualize the significance of poly finger spacing and it's interaction with poly width on the measured resistance.

With the dominant layout factors being established, a contour plot as shown in Fig.5 helps to visually gauge the different resistance zones for various poly width and spacing combinations, for design decision making. This analysis has been performed on the test structure data measured from two different CMOS processes and the observations are seen to be consistent.

In Fig.6, the impact of well dopants (NW, PW) on the resistor banks has been studied. Since the earlier analysis indicated the non-normality of the dataset and the presence of outliers, mood's median non parametric hypothesis test has been used to judge the NW and PW data groups and it is seen that the type of well used for the placement of the resistor banks plays a role on the measured resistance (p-value < 0.05 and a very high χ^2 from the above statistical test demonstrates that the data groups are statistically different) [7]. The associated silicon effect that is attributed to this is believed to be the impact of anneal process following well doping [8].

In Fig.7, a statistically significant impact of the number of poly fingers (NF) on the measured output has been demonstrated. Device models with only single finger considerations renders the model inadequate for circuit evaluations.

Fig.8. shows the experimental result where multi-finger resistor banks are nested among identical resistor bank dummies and it is seen that that the nested design shows no impact for all NF>1 designs.

But, in Fig.9, we see a statistically significant difference between isolated and nested designs for NF=1 design cases. The associated silicon effect is believed to be potentially linked to poly etching effects.

IV. CONCLUSION AND FUTURE WORK

There are different ways to address the design vulnerability to layout effects. If the process/technology is not frozen then it can be compensated at the patterning stage using either OPC or etch. For a mature process node, such dependencies identified from silicon can be incorporated in the associated device model or the designers can be made aware of the offsets caused per factor by means of design rules to enable robust designs.

From this design experiment, the significance of accounting for poly finger spacing (PO_S), number of poly fingers (NF) and the type of well doping in the resistor device model has been demonstrated, for robust circuit evaluations. This work employs a screening DOE approach as a first attempt to capture dominant layout modulators using limited silicon area available on product silicon. As a result, the dataset obtained from this is limited and is used to derive only a linear statistical model with respect to all the interactions among the factors. A more detailed DOE design has been implemented on the next silicon to capture additional higher order factors in the regression equation to improve the accuracy of the statistical model.

ACKNOWLEDGMENTS

The authors would like to thank Andreas Drejfert, Didier Depreeuw, Geert Spierings and Han Lokker for their technical support. Also, many thanks to Christophe Landez and Arjan Mels for supporting this work.

REFERENCES

[1] S. Saxena et al., "Test structures and analysis techniques for estimation of the impact of layout on MOSFET performance and variability," Proceedings of the 2004 International Conference on Microelectronic Test Structures (IEEE Cat. No.04CH37516), Awaji, Japan, 2004, pp. 263-266, doi: 10.1109/ICMTS.2004.1309492.

[2] Ndiaye, C. & Huard, Vincent & Bertholon, R. & Rafik, M. & Federspiel, X. & Bravaix, Alain. (2016). Layout Dependent Effect: Impact on device performance and reliability in recent CMOS nodes. 24-28. 10.1109/IIRW.2016.7904894. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] Tony, Chan, Carusone., David, A., Johns., Kenneth, W., Martin. (2012). Analog integrated circuit design. 2nd ed.

[4] D. McQuirk, C. Baker and B. Smith, "Test structures to evaluate the impact of parasitic edge FET on circuits operating in weak inversion," 2018 IEEE International Conference on Microelectronic Test Structures (ICMTS), Austin, TX, USA, 2018, pp. 16-20, doi: 10.1109/ICMTS.2018.8383754.

[5] Weste, Neil H. E.CMOS VLSI design : A circuits and systems perspective - 4th ed.

[6] Antony, J. (2014). Design of Experiments for Engineers and Scientists. (2nd ed.) Elsevier. https://doi.org/10.1016/C2012-0-03558-2.

[7] Page, E.B., & Marcotte, D.R. (1966). Chapter III: Nonparametric Statistics. Review of Educational Research, 36, 517 - 528.

[8] Optical Properties and Microstructure of TiNxOy and TiN Thin Films before and after Annealing at Different Conditions.Coatings.2019;9(1):22. https://doi.org/10.3390/coatings9010022

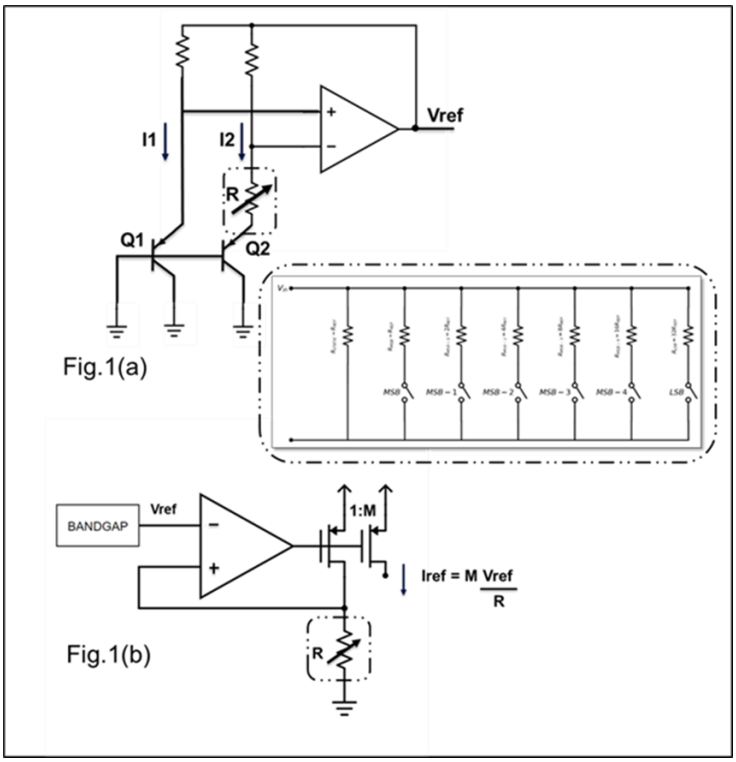


Fig.1(a): A simplified schematic of a bandgap voltage reference circuit. Fig.1(b): A simplified schematic of a reference current generation circuit using a voltage reference. In both circuits, use case of trimming resistor banks has been illustrated.

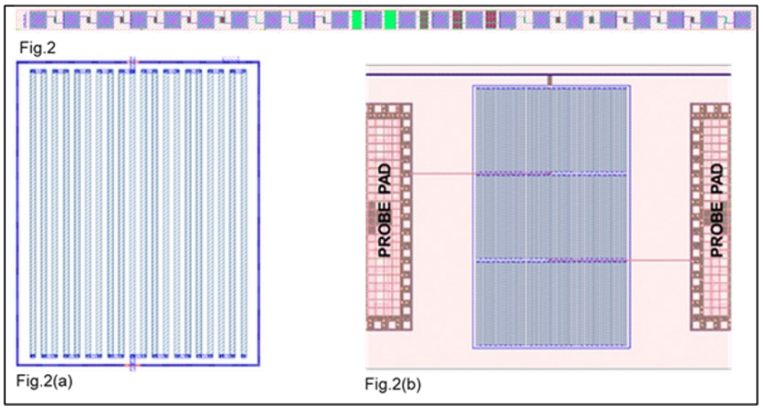


Fig.2: Single-scribe discrete DOE test structure comprising of 21 different layout variants used for layout dependent effects evaluation. Fig.2(a): A multi-poly finger resistor bank DUT with only the center finger electrically connected to evaluate against a single poly finger layout case. Fig.2(b): A nested resistor bank DUT, with the central active resistor bank surrounded by identical dummies as seen between two probe pads.

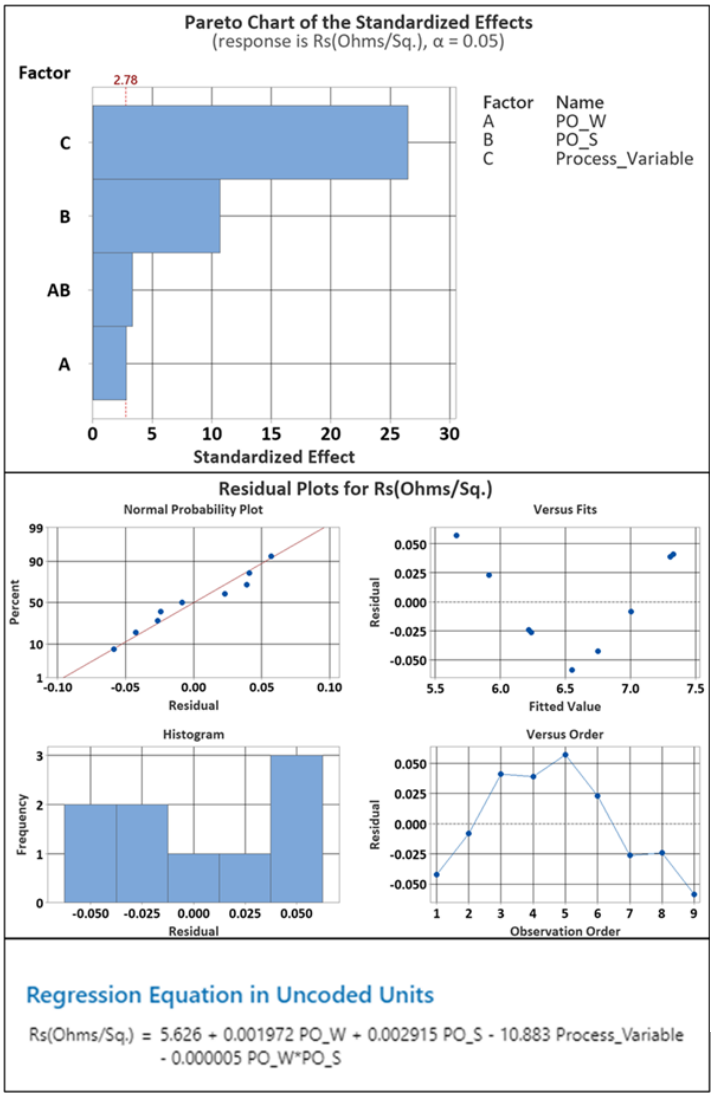


Fig.3: Pareto plot showing the statistically significant dominant layout modulators and the associated regression equation that represents the statistical model along with goodness of fit assessment residual plots, for a given poly length.

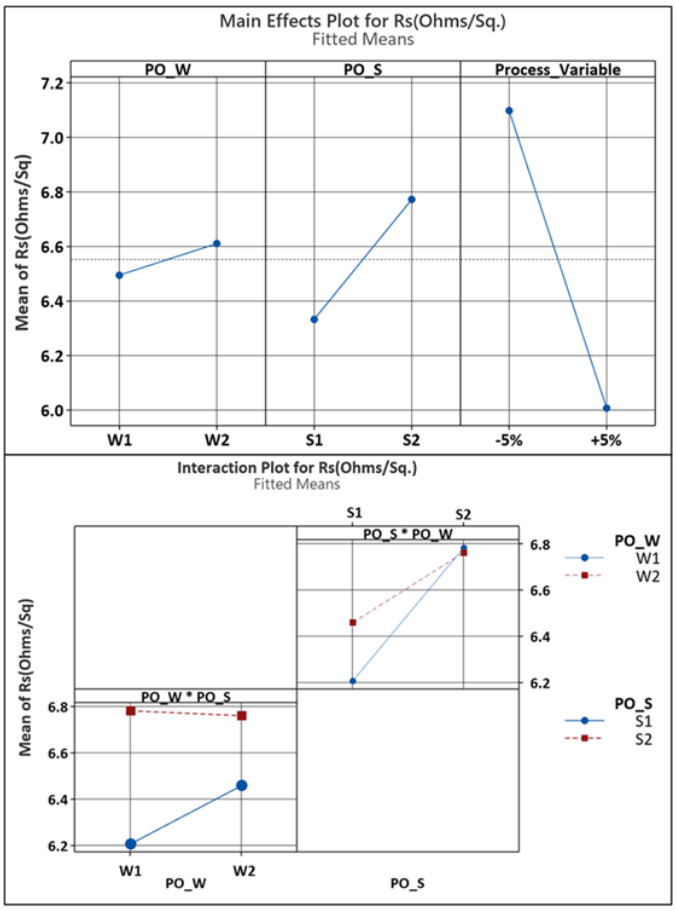


Fig.4: Effects and Interaction plot from 2^k factorial design analysis, for a given poly length.

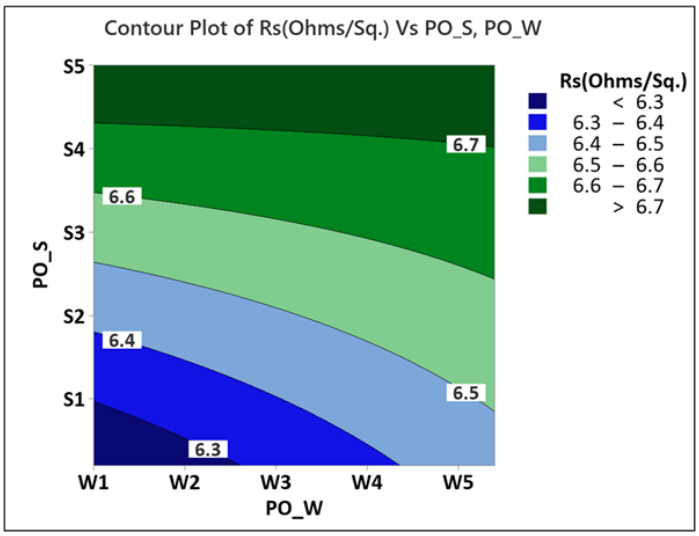


Fig.5: Contour plot illustrating the various resistance zones for varying poly widths and poly finger spacing derived from the above regression equation for design decision making.

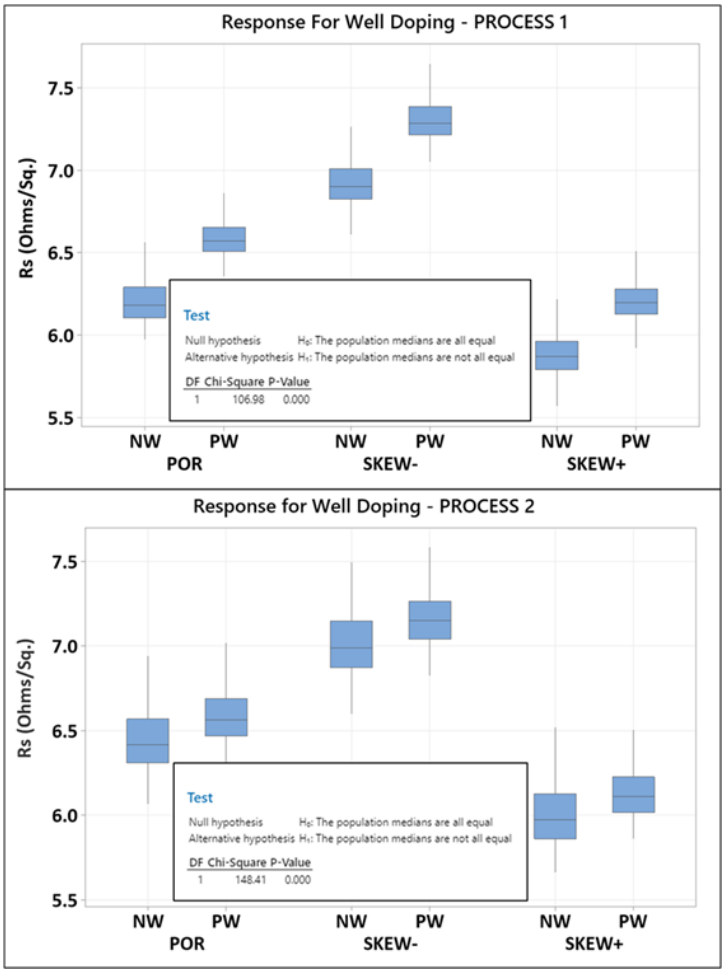


Fig.6: Impact of well-dopants on the measured resistance values.

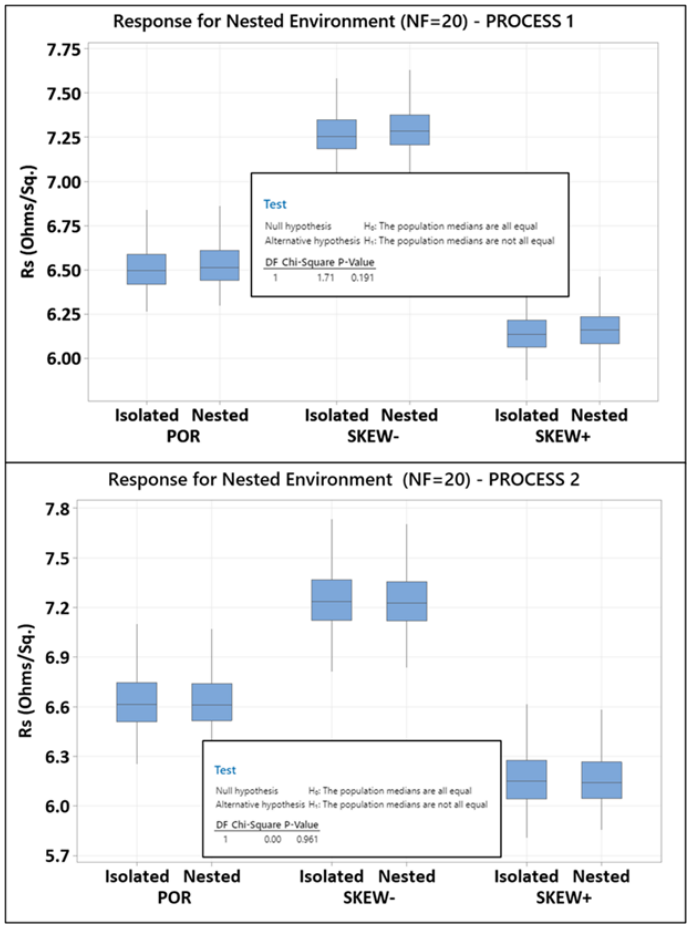


Fig.8: Impact of isolated resistor banks and resistor banks nested among identical dummies, for multi poly finger case, on the measured resistance values.

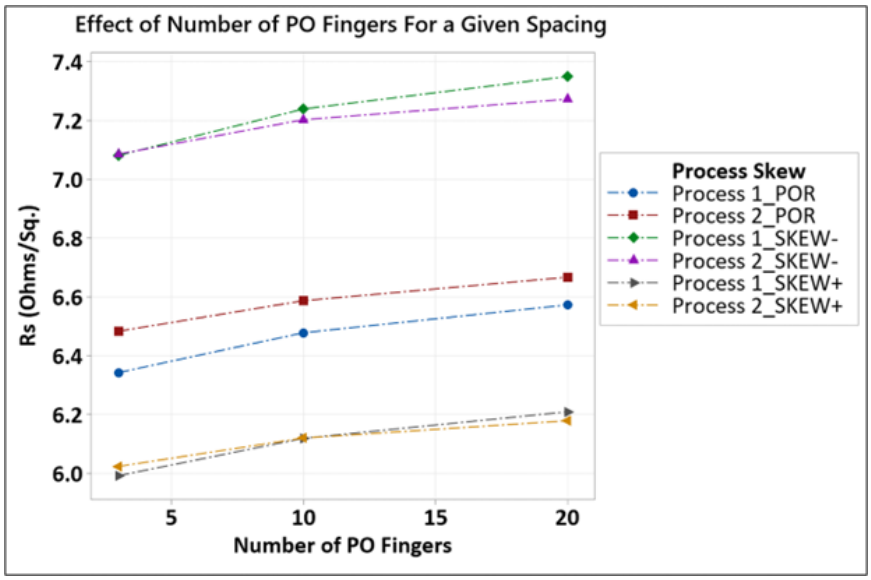


Fig.7: Effect of the number of poly fingers on the measured resistance values.

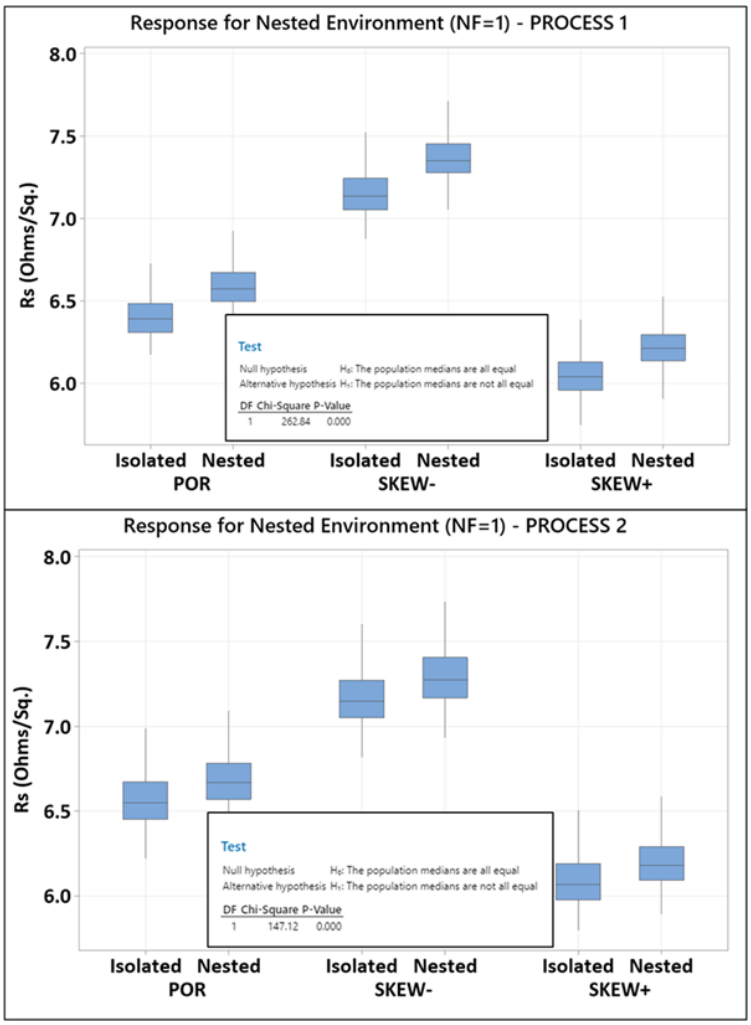


Fig.9: Impact of isolated resistor banks and resistor banks nested among identical dummies, for single poly finger case, on the measured resistance values.

Test Structures for studying the impact of the backend contact metallization on the performance and stress sensitivity of SiGe HBTs

Oliver Dieball[1], Hans Tuinhout[1] and Jeroen Zaal[2]
[1]NXP Semiconductors, Modeling, Eindhoven, The Netherlands
[2]NXP Semiconductors, Package Innovation, Nijmegen, The Netherlands
oliver.dieball@nxp.com, hans.tuinhout@nxp.com, jeroen.zaal@nxp.com

Abstract— **Based on a collection of layout realizations of a simplified SiGe HBT test structure, this paper addresses the impact of mechanical stress induced by the backend metallization placed above the base, emitter and collector regions on the electrical performance of these HBTs. Subsequently, the temperature dependencies of the collector currents and their sensitivity to externally applied mechanical stress is investigated. This study provides valuable insights into test structure layout-related performance shifts relevant for layout and design of modelling test structures and high-precision circuit design.**

Keywords—test structures, backend, contact metallization, stress sensitivity

I. INTRODUCTION

The performance of a silicon-germanium (SiGe) heterojunction bipolar transistor (HBT) is determined by many technological and layout factors. One of these is the metal construction used for the transistor's electrical connections. The intrinsic mechanical stress arising from the device architecture is caused by the different thermal expansion coefficients of metals, dielectrics, and silicon. Already in 2003 it was demonstrated at ICMTS that metal connections can have a significant effect on BJT current mirrors [1]. More recently, it has been demonstrated that strategically placing large metal dummy structures above an HBT can lead to a notable improvement in performance, particularly in the ideal region of operation [2,3]. However, the influence of the transistor architecture itself, typically based on a layout utilizing multiple metal layers and vias with potentially similar effects, remains less explored for the more advanced high-performance RF SiGe technologies.

In the broader context of packaging-induced stress effects, it becomes relevant to investigate whether altering the metallization layout of the transistor can affect mechanical stress sensitivities. The metallization, by changing the stiffness of the backend, can potentially alter how stress propagates from the die surface to the device [4]. This paper addresses these questions through investigating a selection of HBT test structures fabricated in an in-house BiCMOS technology.

To assess the impact of metallization variations more systematically, this study concentrates on the individual effects of additional metal layer stacks above the base, emitter, and collector regions, as well as combinations of these constructions. The exploration begins with an analysis of the differences amongst the electrical characteristics of the different layout options, and this is followed by an investigation of the effects of temperature changes on the observed mechanical-stress-induced changes.

Finally, the sensitivity of two structures to additional external mechanical stress is investigated at room temperature. The pressing probe needle technique that was introduced at ICMTS in 2023 [5] is employed for this, imposing additional external mechanical stress on the die surface, primarily generating a significant additional out-of-plane stress in the Device Under Test (DUT). This part of the study demonstrates that metallization layers do play a discernible, albeit relatively small, role in the device's sensitivity to additional external mechanical stress. This observation is not only relevant for packaging-induced stress effects in IC products, but also for extracting piezo electric coefficients for compact models. It will be shown that the backend metallization must be considered in the utilization of the Pressing Probe Needle (PPN) technique for extracting modeling coefficients, as it can affect the stress propagation from the die surface to the DUT.

The findings as discussed in this paper are essential considerations in the design and layout of process control and modeling test structures, as well as for the layout of high-precision analog circuits.

II. THE TEST STRUCTURES

A collection of layouts was fabricated to investigate the impact of metallization/vias and the additional mechanical stress they impose over certain regions of an in-house SiGe BiCMOS technology with 300 GHz f_T/f_{MAX}. The layout variants are founded on a standardized test structure frame, featuring a device with a length of 10µm, a 0.24µm wide emitter window, and incorporating a CBEBC symmetric layout to allow a systematic assessment of the influence of supplementary metallization and additional vias on the SiGe HBT's performance.

Within this standardized frame, the designated reference structure was meticulously laid out, utilizing only metal 1 connections to the HBT. A microscope photograph and layout illustration are shown in Figs. 1 and 2, respectively, where the metal 1 region is represented by a distinctive light blue color. Importantly, the surrounding metallization layers, positioned more than 5µm away from the active emitter area, ensure a controlled environment for the investigation, enabling a focused examination of the effects of additional metallization and vias on the SiGe HBT's performance. The connections to the probe pads are routed in metal 2, metal 3, and metal 4,

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

maintaining enough distance to solely study the effects of the additional metal and via stacks above the transistor's contact regions. This routing strategy ensures that the exploration is concentrated on the impact of additional metal stacks on the transistor terminals.

This layout approach by itself, with metal 1 connections only, does not necessarily result in optimal (higher current density) transistor performance as would normally be used in a high-performance RF circuit layout, as the access and terminal contact resistances are obviously not minimized for this test structure device.

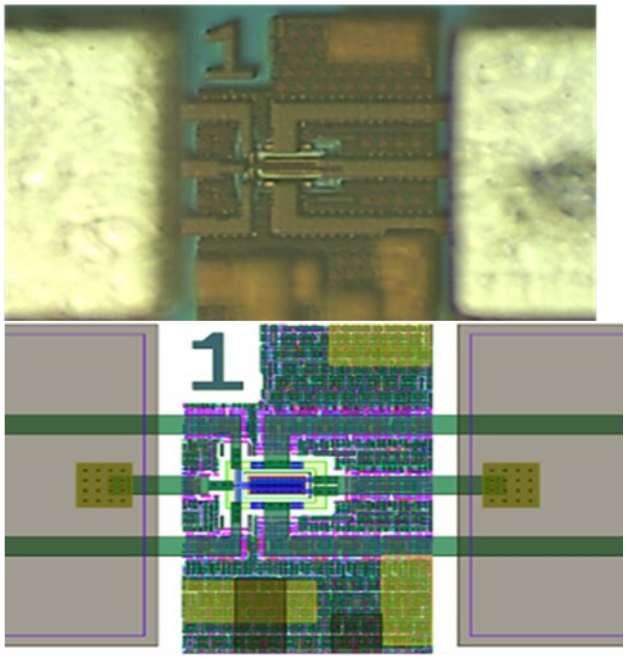


Fig. 1. Microscope photograph (top) of the SiGe HBT and the corresponding layout top view (bottom) of the reference structure. Note that litho dummy tiles are blocked above the device.

In Fig. 2, an illustration of the reference structure provides a more detailed look at the metal 1 connections and the transition to the pad connections. This highlights the structural intricacies of metal 1 connections, offering the visual context crucial for subsequent analyses. Additionally, Fig. 3 presents a schematic cross-section of the HBT structure, indicating the four (aluminum) metal layers and related (tungsten) vias used in this study, each designated with their respective names and coloring. This schematic aids in comprehending the layered composition and facilitates a clearer understanding of the collection of metal stacks on which this study is based.

To study the effects of additional metal stacks above the emitter, collector, and base regions, as well as combinations of these, a collection of 15 derived layouts was designed and fabricated. The schematic cross-sections depicting these structures are provided in Figs. 4, 5 and 6. Some combinations of layout variants are shown in Fig. 7, mimicking typical layout approaches used in RF circuits to minimize the transistor's access resistances. Please note that the "last mile" of the electrical connections to the HBT are always routed in metal 1 only. An example (Type 16) is given in Fig. 8, which shows a layout option with additional metal stacks placed above base, emitter and collector regions. The extra metal stacks are therefore placed such that they do not change the electrical resistance directly because there is no shunting of lower and higher metal layers to the probe pads.

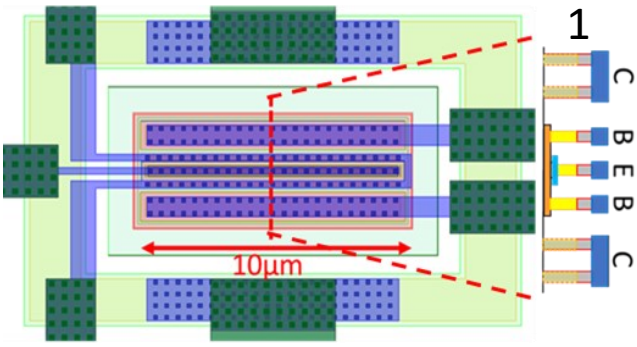


Fig. 2. The reference layout shown on the left uses only metal 1 (light blue) in the "last mile" to the active area and on the terminals (cross section, right). The connections to the pads are routed in metal 3 and 4 (dark green areas). The dashed red line indicates where the cross-section is taken.

III. Electrical Characteristics

For each fabricated layout, the base and collector currents were measured on 33 samples from a single wafer. A typical example of the measurements from one of the layouts is shown in Fig. 9, which shows the Gummel plots that are utilized to calculate the median currents used in this study. To mitigate the impact of random variability, the median values of the (33) observed data points are used for assessing the mechanical stress effects.

Figure 10 shows the relative changes (of the medians) of the collector currents across different layout options for four different bias conditions, indicated in Fig. 9. These measurements are conducted with V_{CB}=0.0V and at room temperature (25°C).

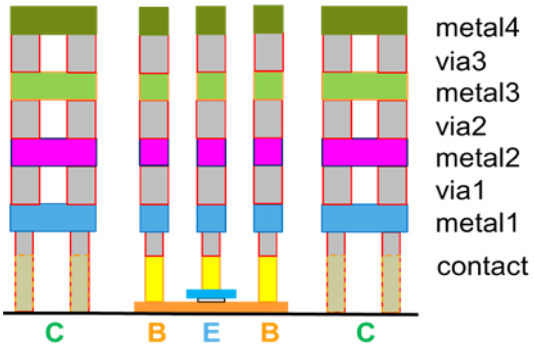


Fig. 3. Schematic cross-sectional view that illustrates the individual metallization layers as well as the vias with their respective names and colors.

Notably, at low base-emitter voltages (V_{EB}=-0.55V and -0.65 V), individual metal stacks exhibit distinct effects on the collector current: The emitter metal stacks (types 2 to 7) give up to a 5% increase, the base metal stacks (types 8, 9, 10) induce up to a 13% increase, and the collector metal stacks (types 11, 12, 13) give an increase of up to 7%. These changes are attributed to a change in the bandgap due to mechanical stress.

Figure 10 also includes combined layout options (types 14, 15, 16), showcasing the cumulative effects of multiple metal stacks. Notably, combinations, such as those with the densest base and densest collector metal stacks (type 16), exhibit an increase of up to 20% in collector current, which corresponds remarkably well to the accumulated result from individual shifts for the base and the collector metal stack types (10 and 13). These layout options represent scenarios in which as much metal as possible was strategically placed above the

structures, comparable to what is done in actual high-performance RF power amplifier circuits.

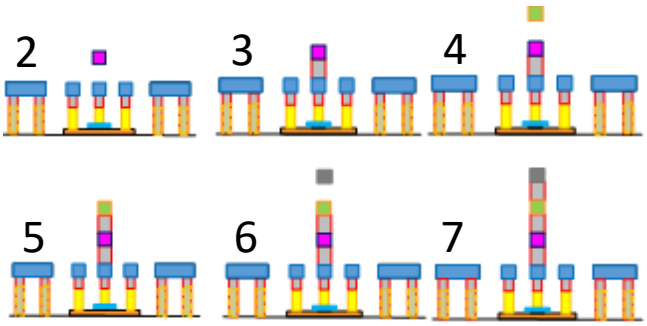


Fig. 4. Illustration of emitter layout options in which the metal layers 2-4 and the respective vias have been included.

Fig. 5. Illustration of base layout options where up to two large metal bars were placed above the base and emitter region.

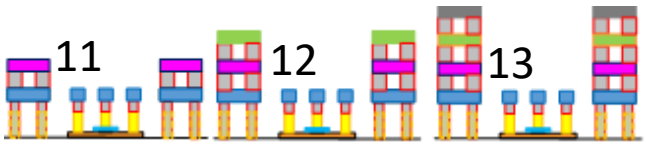


Fig. 6. Illustration of collector layout options where up to three additional layers of metallization and their vias are placed above the collector region.

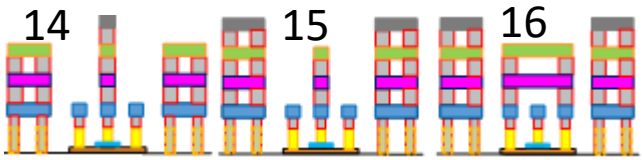


Fig. 7. Illustration of the combined layout options that aim to create the largest possible parametric shifts of the HBT performance.

Note that the observations discussed above focus on the "ideal" (exponential) part of the Gummel plots. These are the regions where the process control measurements are typically collected and which also form the basis of any compact model parameter extraction. The observations from Fig. 10 also show the behavior under higher bias conditions, which is significantly different. No discernible increase in performance is found for the case of collector and combined layout options,

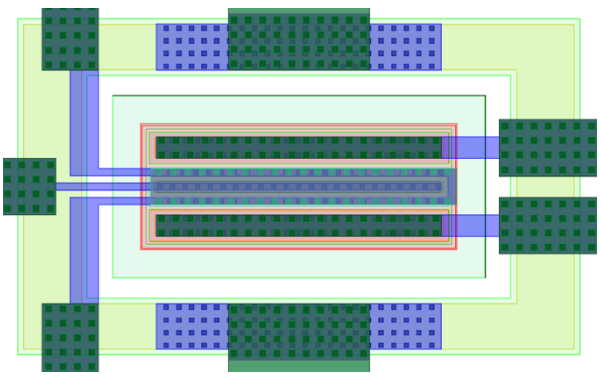


Fig. 8. Layout type 16, which is a structure with an additional metal stack above the base, emitter and collector regions. Please note that the electrical connection is only routed in metal 1 (light blue). Compare to Fig.2 for the layout differences.

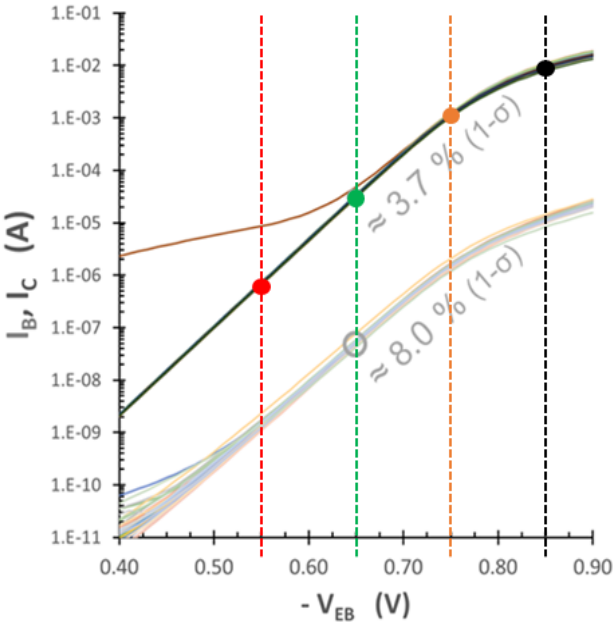


Fig. 9. Example of the Gummel plots for V_{CB}=0.0V at T=25°C obtained from 33 samples measured across the wafer. An uncertainty of +/-1.4% on the median estimators is found due to wafer spread and population size.

particularly at the high base-emitter voltages, which, for these devices, are considered to be affected by the internal emitter resistances, and probably also by the added resistance from the metal 1 (last mile). At the higher bias conditions, a counterintuitive decrease in collector current is noticeable, particularly prominent for collector metal stacks. This peculiar phenomenon could be attributed to an increase of the resistivity in the collector region due to mechanical stress in these regions [6]. In such scenarios, the only discernible benefit in terms of performance is attributed to the presence of emitter/base metal stacks.

IV. Temperature Dependency

Additionally, the impact of temperature on mechanical stress induced by different thermal expansion coefficients in the backend is explored. This section focuses on the effects of dense metallization stacks, hypothesized to yield the largest stress magnitudes. Three specific metallization stacks for analysis, including the emitter, collector, and base metal stacks are selected (types 7, 10 and 13). Additionally, two combination structures are analyzed: emitter+collector and base+collector (types 14 and 16).

To isolate the impact of metal stacks, the analysis involves plotting changes in collector currents relative to the corresponding (bias and temperature) measurements on the reference structure (type 1). This approach again allows for a focused examination of the effects of metal stacks while holding other variables (biasing and temperature effects) constant.

This study spans a temperature range from -50 to +130 degrees Celsius in 10-degree steps, a range that covers common circuit simulation specifications.

The obtained results (Fig. 11) reveal clear trends in stress-induced changes at different temperatures. Note that at low temperatures, the differences between the reference device and those with extra metal on the contact regions can rise to several tens of percents! For instance, while the high emitter metal stack (type 7) yields a relative increase of 7%, at -50°C,

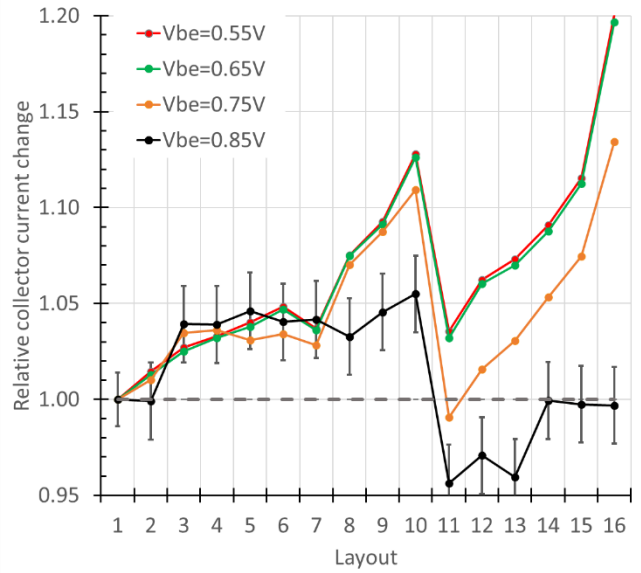


Fig. 10. Change of the collector current at T=25°C of the different layouts relative to the reference structure evaluated at different base-emitter bias conditions with V_{CB}=0.0V. At low voltages, almost all layout styles improve the performance, whereas for higher biases certain layouts lead to a reduced current. The three-sigma uncertainties of the median estimators of the collector current differences are indicated for V_{BE}=0.85V.

maximum metallization (type 16) results in up to a 35% increase. A clear mitigating trend emerges with increasing temperature, indicating that the relative differences in collector currents between the augmented metal stack devices and the reference layout decrease. This behavior (as to be expected) indicates that the mechanical stress in the structures (partially) decreases. At higher temperatures, structures with less metal almost reach the level of the reference (type 1) structure at 130°C, while the structure with densest metallization patterns (type 16) still exhibits a notable difference, converging to a 10% increase. Possibly, this is attributable to the stress contribution of the tungsten via's, which are typically deposited at a higher temperature than aluminum.

These findings not only provide valuable insights into the temperature-dependent behavior of mechanical stress in SiGe HBT structures with varying metallization patterns but are also crucial for the derivation of compact models using these test structures. The good thing is that the metal-induced mechanical stress decreases at higher temperatures, such as those at which RF and analog circuits usually operate. However, at low temperatures, for instance when systems must start up in cold winter conditions, metal-induced performance asymmetries will deteriorate system performance. Moreover, such discrepancies can result in significant circuit simulation deficiencies, as the modeling test structures used to derive the compact models, have, almost by definition, different metal connections compared to those used in actual circuits.

V. Stress Sensitivity

Another potentially relevant consideration for high-precision analog circuits lies in their sensitivity to externally applied mechanical stress, particularly in the context of recent packaging technologies like solder bumps, under bump metallization, interposers, stacked dies, heterogeneous integration, etc.. Such sensitivities become increasingly

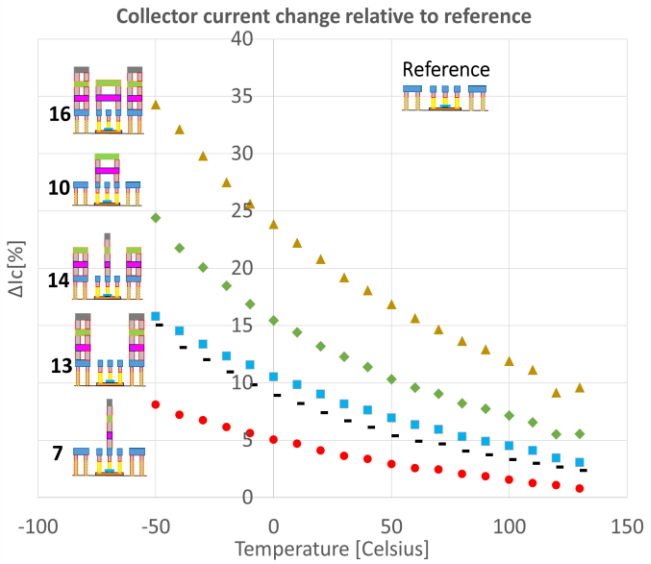


Fig. 11. Change of the collector current evaluated at V_{BE}=0.65V relative to the reference structure as a function of temperature. The largest impact at low temperatures is introduced by the metallization above the base and the collector regions. As the temperature is increased additional mechanical stress from the backend partially relaxes and the relative differences become smaller.

important as new advances in packaging introduce additional mechanical stress to the integrated circuits and the used metal layout could potentially affect how the stress propagates, e.g. from the die surface to the device [4]. In particular, there is the notion amongst analog circuit designers that placing a metal (aluminum) plate over a transistor will mitigate stress effects, as the supposedly "soft" aluminum will absorb mechanical stress.

The pressing probe needle technique [5], which uses a 9 µm round tip tungsten probe needle to press on the die area above the HBT (between the contact pads; see Fig. 12) is utilized to investigate these effects. By pressing the needle at multiple depths (representing multiple forces), each controlled through a motorized positioner, stress sensitivities can be extracted [5]. An example of an extraction result is shown in Fig. 13, which shows the change of the collector current as a function of needle position. Here, the minimum of the collector current coincides with the position of the needle pressing exactly on top of the center of the HBT.

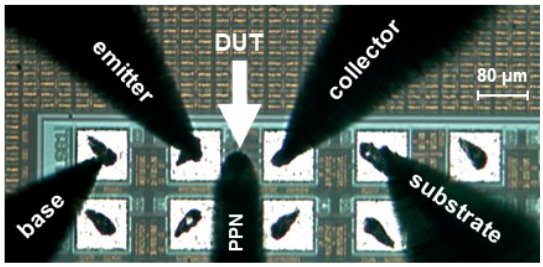


Fig. 12. Example of an HBT modeling test structure with four measurement probe needles and the pressing probe needle placed on (or near) the DUT.

Fig. 14 summarizes results obtained through the pressing probe needle technique, in this case displaying the stress sensitivity of two different structures. Notably, the reference structure exhibits a stress sensitivity of -0.16 %/µm, while the structure with a denser metallization pattern demonstrates a

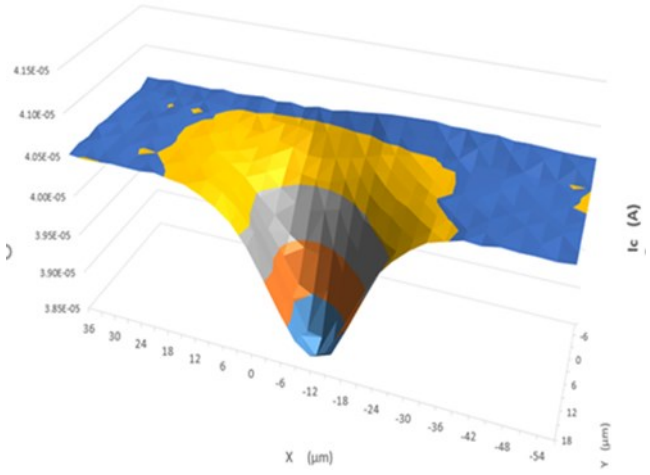


Fig. 13. An example of a full 2D sensitivity scan of the collector current on the layout obtained with the PPN [5] technique for a fixed press depth. A maximum decrease in current of almost 7% is obtained by pressing the needle onto the center of the HBT.

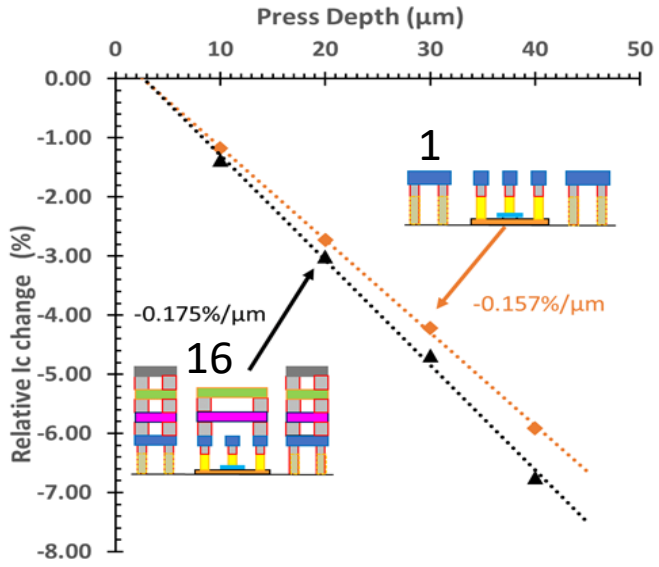


Fig. 14. Relative change of the collector current extracted in the ideal region of operation as a function of press depth obtained from the pressing probe needle experiment. If the needle presses deeper into the surface, the stress magnitude at the DUT increases, resulting in a parametric shift.

slightly higher sensitivity at -0.18 %/µm or approximately 10% difference. Note that the increase of press depth corresponds to a larger force (up to 1 gf at approximately 30 µm) and therefore to a larger stress magnitude. The press depth / force calibration procedure was explained in detail in [5].

To corroborate that the sensitivity is indeed slightly higher in structure 16, we performed a finite-elements mechanical stress simulation experiment using Marc/Mentat [7]. This experiment accounts for the actual backend with the different metallization layers and the various materials with their mechanical properties. Shown in Fig. 15 are illustrations of the mechanical simulations with a round-tip needle (tungsten, red) pressing onto the die surface. The simulation domain in this experiment spans 50 µm in the x-direction, 25 µm silicon under the backend stack, and mimics the full backend stack on the test transistors with a total height of approximately 18.5 µm including a nitride top layer, TEOS, silicon, aluminum and poly silicon with a simulation grid of 1/6 µm.

The simulation results of an axi-symmetric finite-elements simulation for a fixed load of 10 mN (approximately 1 gf) are

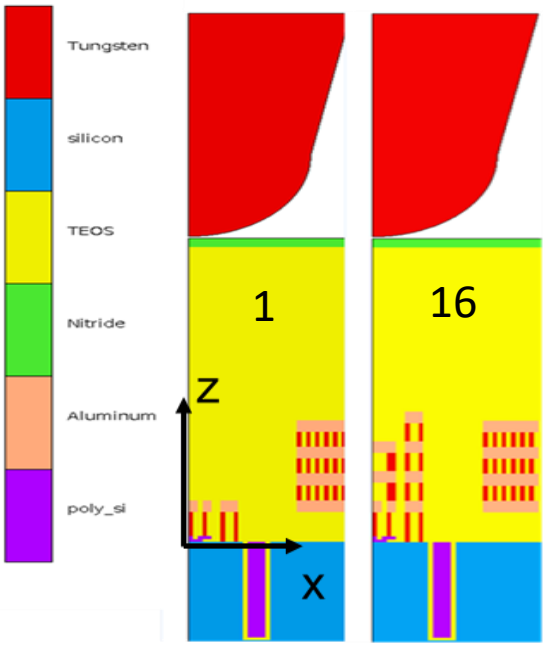


Fig. 15. Illustration of the mechanical stress simulation setup for the structure with sparsest (reference 1, left) and densest metallization (16, right) pattern. The pressing probe needle (red, tungsten) applies a load of 10mN on the die surface. The HBT is located at the center of the coordinate system. The cross-sections only show part of the simulated regions.

displayed in Fig. 16, indicating the out-of-plane stress on the y-axis and the distance from the center of the DUT on the x-axis. No intrinsic stress, e.g., due to the manufacturing and different temperature steps of the backend, was included in this analysis. The approach is based on the assumption that the external mechanical stress can be superimposed upon the intrinsic stress. Our experience is that this methodology yields a good agreement between simulations and experiments if the temperature and therefore the intrinsic stress is kept constant.

The 10 mN needle pressure results in an increase in the out-of-plane stress magnitude of 5.3 MPa (19%) in the case of the structure with the densest metallization. This alteration in the propagation of stress from the die surface to the DUT in fact indicates that the backend has become less soft, or rather, that the stress propagates downwards more effectively.

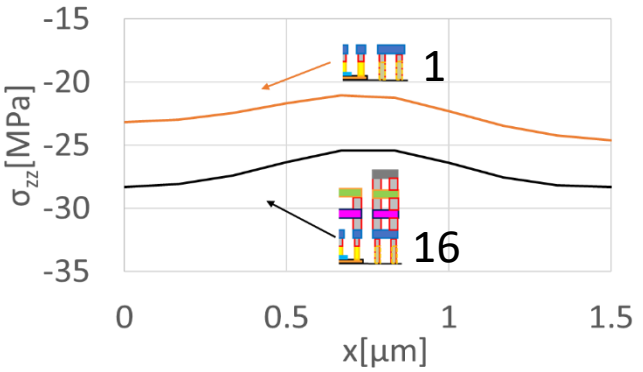


Fig. 16. Simulated out-of-plane stress σ_{zz} as a function of distance to the center of the device at z=-0.5 µm. The needle is assumed to press with a load of 10mN onto the die surface. The denser metallization pattern (16) leads to an increase of the stress magnitude by 19%.

Such observations underscore the importance of considering the layout of test structures when extracting stress sensitivities or piezo electric coefficients, as there is a small, but significant difference visible.

Moreover, as mentioned above, many circuit designers will still assume that placing an aluminum plate over a transistor will reduce the mechanical stress sensitivity of the transistor below. Our measurement (Fig. 14) shows the (small) opposite, and this is indeed substantiated through the mechanical simulations, indicating that the stress at the DUT level indeed increases.

These findings obviously emphasize the necessity of using both physical (real) and simulation approaches to comprehensively understand the mechanical behavior of test structures.

CONCLUSIONS

This paper demonstrates that using different realizations of wired parameterized cells used for process control modules (often utilizing only metal 1) and compact modeling modules (generally utilizing multiple metals to mitigate access resistances) can lead to large differences in their electrical characteristics. Similarly, transistor implementations, as laid out by circuit designers, aiming to improve specific properties of a circuit by optimizing the layout, may again perform differently. This can create errors and circuit fails as the interpretation of the measurement data and their respective compact modeling coefficients, and hence circuit simulation results, can be different by up to tens of percents.

The second important message is that at high temperatures the differences between the collector currents of different layout realizations become smaller, implying that with increasing temperature the mechanical stress generated in the backend relaxes. A corresponding finding is that at lower temperatures a simple layout can show 35% less collector current than a layout with a very dense metallization. This indicates that the mechanical stress in the backend increases with decreasing temperature and the amount of stress depends on the number of metals and vias placed above the device.

Finally, using the PPN method we demonstrated that it is important to carefully select layouts for mechanical stress analysis. By simulation and experiment we demonstrated for instance that metal stacks covering devices do not absorb stress but in fact increase the sensitivity of a device to externally applied out-of-plane mechanical stress.

ACKNOWLEDGMENT

The authors would like to thank Peter Magnée, Johan Donkers, Thanh Viet Dinh and Stephen John Sque for their help and support during this investigation.

REFERENCES

[1] Tuinhout, H. P., et al. "Current mirror test structures for studying adjacent layout effects on systematic transistor mismatch." *International Conference on Microelectronic Test Structures, 2003*. IEEE, 2003.

[2] Rücker, H., et al. "Impact of back end stress on SiGe bipolar transistors." *2011 International Electron Devices Meeting*. IEEE, 2011.

[3] Canderle, E., et al. "Impact of BEOL stress on BiCMOS9MW HBTs." *2013 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*. IEEE, 2013

[4] Tuinhout, H., et al. "Characterization of parametric mismatch attributable to plastic chip encapsulation." *2020 IEEE 33rd International Conference on Microelectronic Test Structures (ICMTS)*. IEEE, 2020.

[5] Tuinhout, Hans, and Oliver Dieball. "The Pressing Probe Needle Technique for Characterizing Mechanical Stress Sensitivity of Semiconductor Devices." *2023 35th International Conference on Microelectronic Test Structure (ICMTS)*. IEEE, 2023.

[6] Dieball, Oliver, et al. "Microscopic Simulation of the RF Performance of SiGe HBTs With Additional Uniaxial Mechanical Stress." *IEEE Transactions on Electron Devices* 69.9 (2022): 4803-4809.

[7] https://en.wikipedia.org/wiki/MSC_Marc

A Step-by-Step Layout Transformation Approach to Differentiate How Multiple Layout Dependent Effects Modify Device and Circuit Performance

Luke Lu[1], Kejun Xia[2], Ronald van Langevelde[1], Colin C. McAndrew[1], and Wuxia Li[1]
[1]NXP Semiconductors, Front End Innovation
Email: wuxia.li@nxp.com
[2]TSMC, Power Management Business Development

***Abstract*—Ring oscillators (ROs) are important components of electronic circuits. Their performance, however, is susceptible to parasitic effects when device sizes scale down. We present a methodology, that uses a series of test structures which step-by-step transform an RO layout into single-device modeling layouts, to differentiate and quantify the impact of individual layout dependent effects on device DC performance. A model that is tuned to fit the DC characteristics of the devices in an RO, which are subject to multiple LDEs, gives correct timing behavior.**

Keywords—Ring oscillator, device modeling, layout dependent effects, MOSFET, design technology co-optimization

I. Introduction

With the continuous scaling down of CMOS feature sizes, devices with same channel width and length but with different surrounding layout patterns show significant variations in their electrical performances [1-2]. There are numerous such layout dependent effects (LDEs) as summarized in [3], including the shallow trench isolation effect (STI or LOD), the well proximity effect (WPE), the deep trench effect (DTE), and the active spacing effect (OSE). These effects impact circuit performance and should be characterized and modeled. However, it is very challenging to mimic the real application environment due to a number of facts, namely: (i) the individual effects are typically small and need many well-designed test structures to single out the real magnitude of the effect; (ii) some effects interact and interfere with each other, thus cannot be decoupled [4]; (iii) LDE effects are challenging to model accurately due to the intricate irregular layout patterns encountered in real applications; and (iv) disparities that exist between the structures used for modeling and those present in actual circuits.

For a ring oscillator (RO) fabricated in a 0.13 µm SOI BCD process [5], we observed a more than 10% gap between simulated RO frequency and the mean value of the measured frequency. The SPICE models of the pmos and nmos transistors, extracted using Si data from DC modeling structures, included STI, WPE, and DTE effects. Even after addressing inaccuracies in the capacitance model, the gap persisted. This prompted us to investigate the accuracy of, and potential oversights in, our LDE models.

One possibility is that the layout 2D effects are oversimplified. In [6] the 2D STI stress distribution of a PLL was simulated with a finite element analysis, based on circuit layout, and then the extracted stress tensors were fed into a 2D-aware model to modulate the bandgap, mobility, and device threshold. This method can provide good results, after calibration, when one LDE dominates. However, there are several LDEs that contribute at about the same level in our case, so we could not use this approach. Furthermore, 2D stress analysis is not applicable to the WPE.

In this work, we created a series of layout structures that, step-by-step, transform the RO into the standalone devices in our DC modeling structures. To the best of our knowledge, this is the first time such a method is used to investigate LDEs.

In each of these structures, both nmos and pmos transistors were connected to pads, to enable measurement of their individual DC parameters threshold voltage VTH and drain saturation current IDSAT. Consequently, we successfully identified the contributions of different LDEs to the discrepancies between Si and simulation results. We found that models fine-tuned to the DC performance of the devices in the RO could give correct timing behavior and close the gap between models and Si. This validates our approach of analyzing LDEs in a real circuit and improving LDE modeling. Moreover, our methodology offers valuable insights for design technology co-optimization (DTCO).

II. Test Structures

The RO consists of 31 stages of NAND inverters, with its output connected to a frequency divider and a buffer. Each NAND-cell consists of four nmos transistors and four pmos transistors, see Fig 1. The nmos and pmos transistors used in this study are 1.5V devices with regular threshold voltages.

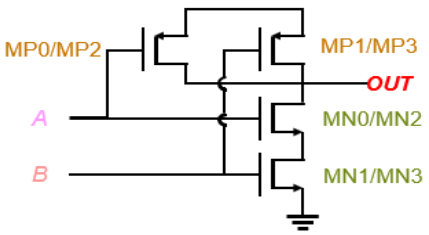


Fig. 1: NAND2 schematic of the 31-stage RO circuit. Each transistor is split into 2 sections and the sections of each transistor type are laid out in a common centroid configuration.

In an RO, the oscillation frequency is determined by the number of stages N and the gate propagation delay τ_d per stage, which is approximately linearly proportional to the supply current I_{DD} and inversely proportional to the load capacitance of each stage. We created a series of layout structures that step-by-step transformed the RO into a single

979-8-3503-2987-2/24/$31.00 ©

NAND cell and then further transformed that cell into a standalone pmos transistor and a standalone series connection of two nmos transistors (to match the pull-down devices in the NAND cell of Fig. 1), respectively. Both the pmos and series-nmos in each of these structures were connected to pads so that we could measure them separately. Gate protection diodes were added to protect the devices from plasma induced damage.

Structure Type	Layout No.	Layout Changes
RO	1	Ring oscillator (31-stage NAND inverters, with a frequency divider and a buffer)
	2	Remove neighboring poly & metal
	3	Remove neighboring actives
	4	Remove neighboring NAND
	5	Shrink NWELL & PPLUS
	6	Standalone NAND (keep allTiBlk[a] & DTI[b] as RO)
	7	Standalone NAND (shrink allTiBlk, keep DTI as RO)
NAND	8	**Standalone NAND (shrink DTI)**
	9	Add Active tilling in length direction (>20um in source and drain ends)
	10	Remove 4 nmos
	11	Remove neighboring pmos
	12	Change SB[c] & contact as those in modeling DC test structures
	13	Change SA[d] as those in modeling DC test structures
	14	Change SA & contact as those in modeling DC test structures
	15	Remove active leg (drain side)
	16	Stretch DTI
	17	Stretch DTI + Stretch NWELL
	18	Stretch DTI + Stretch NWELL + Add dummy gates
Standalone Device	19	**Single pmos**

Fig. 2: The step-by-step transformation from ring oscillator into NAND cell and further into standalone device for the pmos transistor. Note: a) for allTiBlk- all tiling block layer; b) for DTI- deep trench isolation ring; c) for SB- the distance from poly gate to active edge at drain side; d) for SA- the distance from poly gate to active edge at source side.

Fig. 2 shows the detailed layout changes for the pmos transistors. From layout #1 to #7, the process includes removing the neighboring poly/metal, actives and NANDs, shrinking the NWELL and p^+ doping to get the standalone NAND, and then further shrinking the all-tiling-block layer while keeping the deep trench isolation (DTI) the same as in the RO. Layout #8 results in a standalone NAND with a shrunk DTI. From layout #9 to #11, active tiling with a size larger than 20 μm was added at both source and drain sides along the channel length direction, followed by removal of the 4 nmos transistors, then the neighboring pmos transistors, until a single pmos was fully isolated. From layout #12 to #19, we further transform the single pmos used in RO to those used in SPICE modeling: First, the distance from poly gate to active edge at the drain side (SB) and the distance from poly gate to active edge at the source side (SA) were changed, as well as the contacts. Then the active leg at the drain site was removed, the DTI was stretched, followed by further stretching the DTI and NWELL, and adding dummy gate to a real standalone pmos transistor used for SPICE modeling.

Fig. 3 shows the flow of layout transformation from RO to NAND and then to two nmos in series for the nmos transistor. We characterize two nmos in series as a whole because there is no contact between the two poly gates in the NAND cell and we do not want to insert contacts, which would inevitably change device performance.

Structure type	Layout No.	Layout Changes
RO	1	Ring oscillator (31-stage NAND inverters with a frequency divider and a buffer)
	2	Remove neighboring poly & metal
	3	Remove neighboring actives
	4	Remove neighboring NAND
	5	Standalone NAND (keep allTiBlk[e] & DTI[f] as RO)
	6	Standalone NAND (shrink allTiBlk, keep DTI as RO)
NAND	7	Standalone NAND (shrink DTI)
	8	Add active tilling in length direction (>20um in source and drain ends)
	9	Remove 4 pmos
	10	Remove neighboring nmos pair
	11	Change SB[g] as DC as those in modeling DC test structures
Standalone Device	12	**2-nmos in series**

Fig. 3: The step-by-step transformation from ring oscillator into NAND cell and further into standalone 2-nmos in series for the nmos. Note: e) for allTiBlk- all tiling block layer; f) for SB- the distance from poly gate to active edge at drain side; g) for SA- the distance from poly gate to active edge at source side.

Fig. 4 (a) shows the layout of the 31-stage RO and a single NAND that consist of 4 pmos and 4 nmos transistors, which are each laid out in a common centroid configuration. Fig. 4 (b) shows the PCell layouts for individual transistors; these are the layouts used for the SPICE modeling structures. The n-type buried layer (NBL) is placed under all digital circuits and single devices to prevent latch-up but it is not placed in the SPICE modeling test structures. However, we verified that the

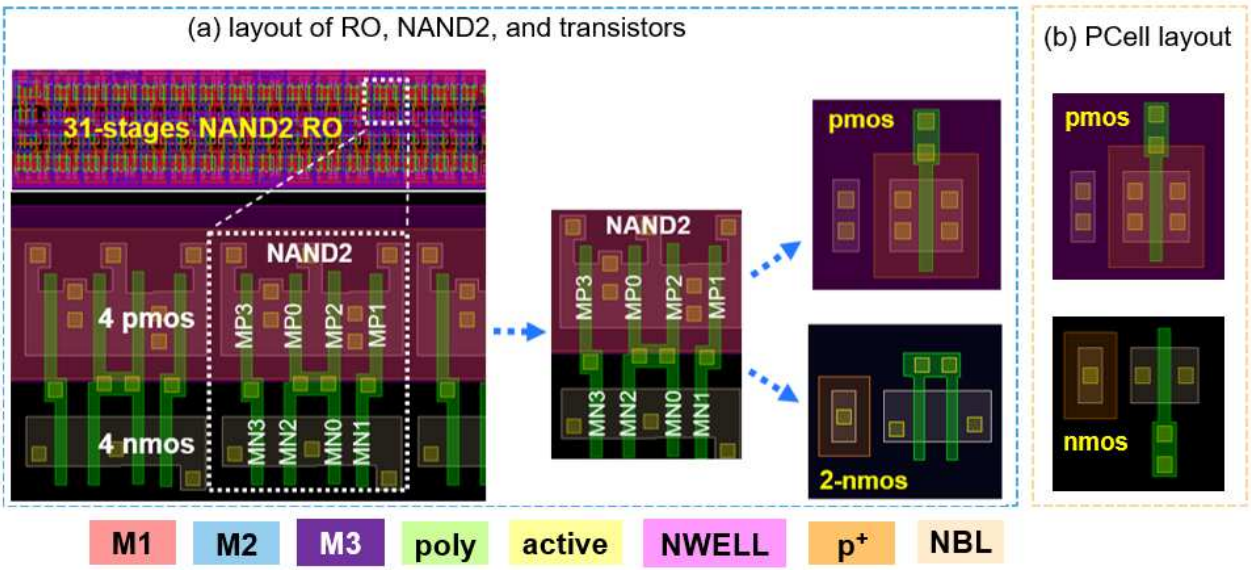


Fig. 4: Layout transformation from ring oscillator to NAND and then to single device for pmos transistor and 2-nmos in series for nmos transistor. The n-type buried layer (NBL) is placed under all digital circuitry and single devices to prevent latch-up; it is not included in the PCells, which are the basis for the SPICE modeling test structures.

presence or absence of NBL has a negligible influence on DC behavior.

Fig. 5 shows the layout for one of the key steps by stretching the NWELL, which caused the most significant change in the drain saturation current for pmos, from the preceding layout step.

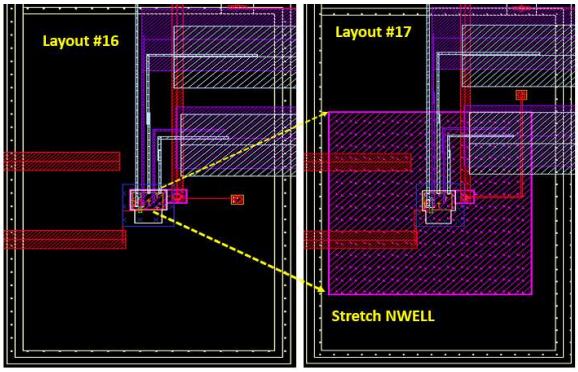


Fig. 5: Layout change with stretching the NWELL which resulted in dramatic change in DC performance. In layout #16, the spacing between active and well edges are 0.415µm, 0.33µm, 0.735µm and 1.095µm, respectively in the four directions; for Layout #17, the spacing is 10µm in all directions. In both structures, NBL is aligned with DTI and all other conditions are identical.

III. Result and Discussion

A. Impact on the threshold voltage (V_{TH})

In Fig. 6(a), the relative changes in V_{TH} are illustrated, with the pmos in the original RO (layout #1) serving as the reference for all the layouts. Removing the active tiling of poly and metal resulted in a subtle 1 mV decrement in V_{TH} (absolute value increases), while removing the active tiling led to an increment of 26 mV in V_{TH} which is mainly caused by stress relaxation [1]. We also found that shrinking the NWELL caused an additional increment of 22 mV due to WPE effect. From the NAND structure further adding active tiling led to a decrement of 18 mV in V_{TH} due to stress.

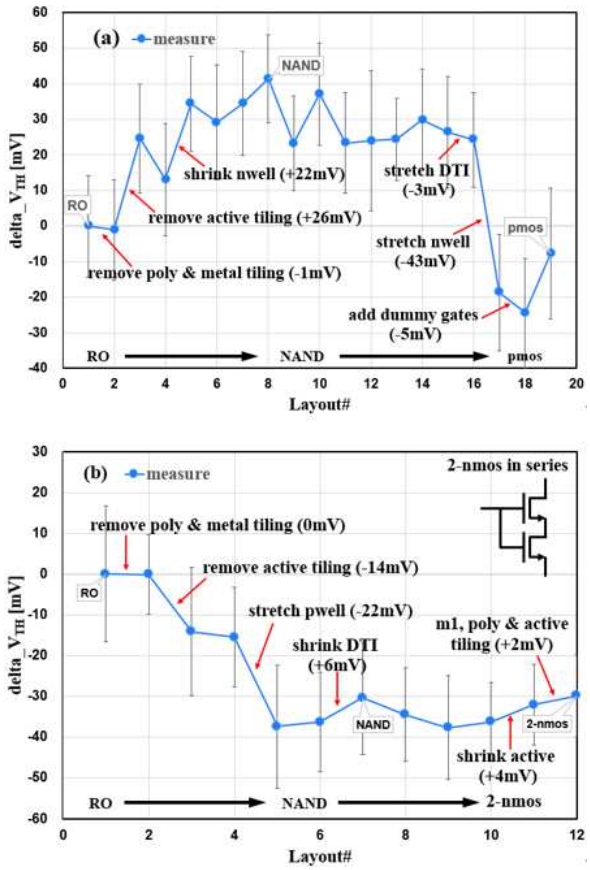


Fig. 6: Delta in threshold voltage (V_{TH}) in step-by-step transformation from ring oscillator to standalone devices: (a) single pmos; and (b) 2-nmos in series. The measured delta_V_{TH} are the median of wafer mapping data of 38 samples; the error bars are ± 1 σ variation of the samples.

Removing the neighboring 4 nmos transistors led to an increment of 14 mV while further removing the neighboring 3 pmos transistors resulted in a 14 mV drop which can be explained by the stress relaxation effect from different directions. Thereafter, the layout changes exerted minor changes V_{TH} till an abrupt drop of 43 mV was observed when stretching the NWELL due to WPE and a mild decrement of 5 mV by adding dummy gates to the device. An additional increment of 14 mV was observed when compared to the identical PCell device used in SPICE modeling extraction. Thus, the most significant LDE on V_{TH} for pmos here is the WPE. A similar observation has also been reported and the impact on the digital circuit timing has been discussed [7]. It could be suppressed by increasing the separation between gate poly and the enclosing well edge, as we observed here. Change of the active tiling, by removing or shrinking, has only a minor impact. The V_{TH} variation vs layout change for the 2-nmos in series is shown in Fig. 6(b). The changes of each step are smaller than those observed for pmos, the most pronounced impact is from removing the active tiling and stretching the p-well, which caused decrements of 14 mV and 22 mV in V_{TH}, respectively. Modifying the active tiling, either by removal or shrinking the enclosing well edge [8], has only a minor impact on nmos.

B. Impact on the on-current I_{DSAT}

Fig. 7(a) and 7(b) show the changes in the drain current I_{DSAT} per width during the step-by-step transformation from the RO to the standalone devices for pmos and 2-nmos in series, respectively. For pmos, by comparing the I_{DSAT} values at step #1 and #16, the totally variation is about 9%. The most significant change of 14% is observed when stretching NWELL due to WPE. For nmos, the main LDEs the device in RO suffered are the stress from the active tiling and WPE, which result in about 7% and 3% decrement, respectively. What's interesting, albeit unfortunate, is that our LDE models created based on standalone device structures are not able to accurately model the LDEs we observed in these layouts.

C. RO timing improvement with updated model

Our model has captured STI, WPE, DTE effects, but does not include other effects such as active tiling and its direction dependence. A general model to capture these missed LDEs is difficult. To close the gap between simulations and Si for RO frequency, we tuned both nmos and pmos models to fit the measured I_{DSAT} of the MOSFETs in the RO: The simulated oscillation frequencies then matched the measurements. This indicates that the LDEs, primarily influenced by the well proximity effect and active/poly/metal tiling, play a significant role. Decoupling these effects is feasible but convoluted. Design optimization requires special care and attention to navigate these complexities effectively.

IV. Conclusion

In this paper, we used a series of test structures which step-by-step transforms the RO layout into a single-device modeling layout, to differentiate and quantify the impact of individual layout dependent effects on the device level DC performance for both pmos and nmos transistors. Our findings revealed that the active tiling and well proximity effects exerted predominant impacts on DC performance, accounting for the timing deviation gap observed in RO simulation compared to Si performance. Our experimental results confirmed that for digital circuit design, the impact of LDEs on timing prediction can be decouplable but complicated and should be carefully considered in design-technology co-optimization.

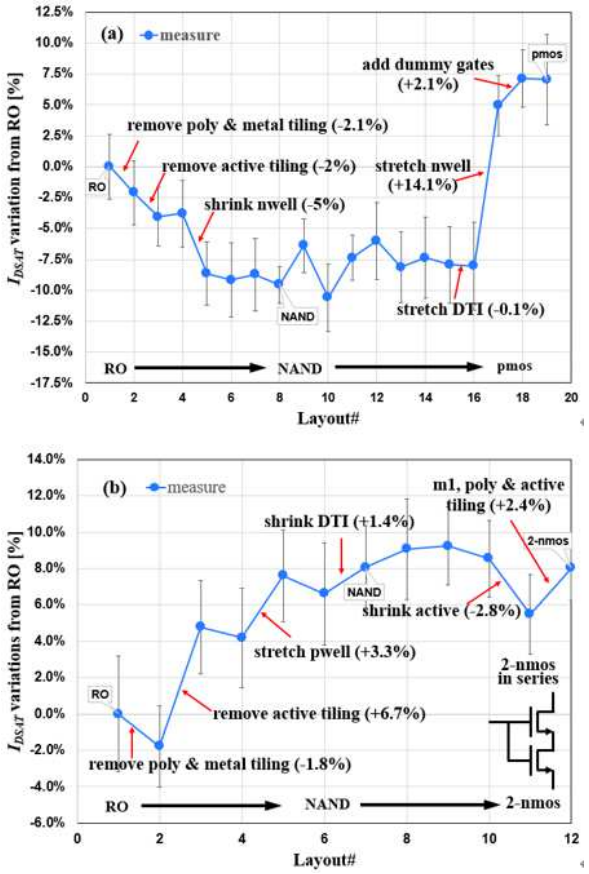


Fig. 7: I_{DSAT} per width and its changes in step-by-step transformation from ring oscillator to standalone devices: (a) single pmos; and (b) 2-nmos in series. The measured I_{DSAT} are the median of wafer mapping data of 38 samples; the error bars are ± 1 σ variation of the samples.

References

[1] J. V. Faricelli, "Layout-dependent proximity effects in deep nanoscale CMOS," *Proc. IEEE Custom Integrated Circuits Conference*, pp. 1-8 (2010).

[2] C.-W. Wu, K.-T. Chen, R. Lee, W.-S. Kao, H.-J. Liao, J. Chang and S. Natarajan, "A 28nm ROM with Two-Step Decoding scheme and OD-space-effect minimization to achieve 30% speed and 190mV Vmin improvement," *Symposium on VLSI Circuits*, pp. 1-2 (2013).

[3] C. C. McAndrew, "Compact models for MOS transistors: successes and challenges", *IEEE Trans. Electron Devices*, vol. 66, no. 12, pp. 12-18 (2019).

[4] Q. Lin, H. Pan and J. Chang, "Test structures for debugging variation of critical devices caused by layout-dependent effects in FinFETs," in *Proc. Int. Conf. Microelectron. Test Struct.*, pp. 3–6 (2018).

[5] R. Zhu, V. Khemka, T. Khan, W. Huang, X. Cheng, P Hui, M. Ger, and P. Rodriquez, "A high voltage Super-Junction NLDMOS device implemented in 0.13μm SOI based Smart Power IC technology," *ISPSD*, pp. 79-82 (2010).

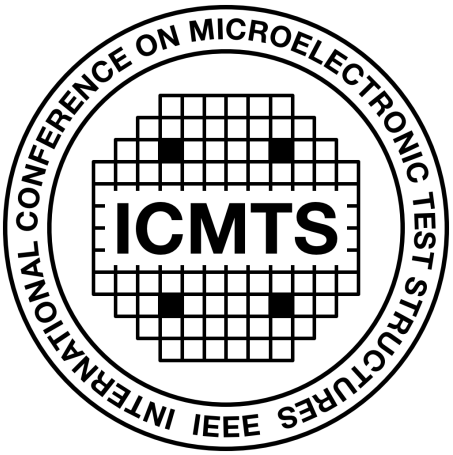
[6] J. Xue, Z. Ye, Y. Deng, H. Wang, L. Yang, Z. Yu, “Layout-dependent STI stress analysis and stress-aware RF/analog circuit design optimization”, *IEEE/ACM ICCAD*, pp. 521-528, (2009).

[7] T. Kanamoto, Y. Ogasahara, K. Natsume, K. Yamaguchi, H. Amishiro, T. Watanabe and M. Hashimoto, "Impact of well edge proximity effect on timing," *ESSCIRC 2007 - 33rd European Solid-State Circuits Conference*, pp. 115-118 (2007).

[8] T. B. Hook, J. Brown, P. Cottrell, E. Adler, D. Hoyniak, J. Johnson and R. Mann, “Lateral ion implant straggle and mask proximity effect”, *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp.1946–1951, (2003).

SESSION 2

Reliability



16th April 2024, 11:20–12:20

Session Co-Chairs: Francesco Driussi, *Università di Udine, Italy*
Brad Smith, *NXP Semiconductor, USA*

A novel test structure with two active areas for eNVM reliability studies

Khaled Alkema[1,2], Franck Melul[1], Vincenzo Della Marca[2], Marc Bocquet[2], Madjid Akbal[1], Arnaud Regnier[1], Stephan Niel[3], Francesco La-Rosa[1]

[1]STMicroelectronics Rousset, France

[2] Aix-Marseille University, CNRS, IM2NP UMR 7334, Marseille, France

[3]STMicroelectronics Crolles, France

Abstract— **This paper presents a test structure with a poly floating gate shared on two actives areas. Programming and erase can be split toward these two regions with a specific arsenic implantation. The aim is to study the tunnel oxide degradation and the injection efficiency of embedded charge storage memory cells.**

I. Introduction

Embedded non-volatile memory using floating gate technology is still largely used in automotive and IOT products [1] despite the development of new memory technologies. The embedded Select in Trench Memory (eSTM™) is a 40 nm split-gate technology developed by STMicroelectronics [2] to optimize size, cost, and performance for microcontrollers applications. Compared to 1T and 2T cells, eSTM™ can manage faster programming/erasing and reading operations with a lower energy consumption [3]. This article proposes a novel Test Structure with Two Actives areas (TSTA) to study eSTM™ cell cycling degradation by dissociating programming and erasing operations. Firstly, the TSTA will be detailed as well as the electrical activations used in this work. Finally, endurance tests will be used to determine the impact of Source Side Injection (SSI) programming versus the Fowler-Nordheim (FN) erase on the oxide degradation.

II. TSTA description

A. Device architecture

Our test structure process begins with NISO and P active well implants, representing the source and substrate terminals respectively. This step is followed by the select in trench transistor fabrication and the tunnel oxide growth. Finally, the memory stack polysilicon floating gate/ONO/polysilicon control gate is deposited. This is fabricated on two active areas, that enables the common poly floating gate sharing as shown on Fig. 1a. Two operating regions called Prog-Zone and Erase-Zone are thus defined and separated by a Shallow Trench Isolation (STI). A phosphorus implant named CAPA is diffused in the Erase-Zone with high annealing time to optimize p-n junction for FN erasing. On the other hand, there is an arsenic buried implant in the Prog-Zone, that makes the cell conductive when it's in the virgin state (normally on) [2]. Finally, the n-type drain region is realized, this implies the floating implant (FI) implementation across the select and sense transistors. Fig. 1b and Fig. 1c show the associated Transmission Electron Microscopy (TEM) pictures of length (L) and wide (W) directions respectively. These validated the TSTA morphology with the poly-floating gate sharing on both sides and the eSTM™ like cut in length direction.

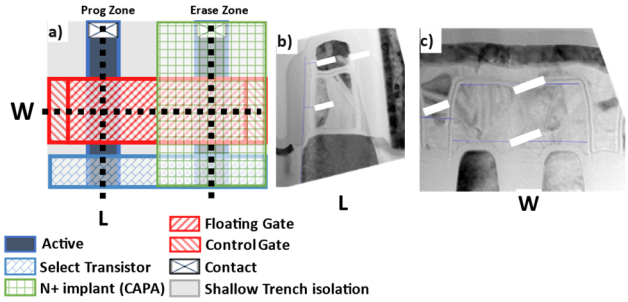


Fig. 1: (a) Layout and (b, c) TEM cuts of the TSTA cell.

B. TSTA coupling factor

TSTA has been designed for reliability understanding of floating gate based non-volatile memory cells (NVM). To study TSTA programming and erasing schemes, we consider the poly floating gate capacitive model. This allows to calculate the coupling factors using planar and fringing capacitances [4].

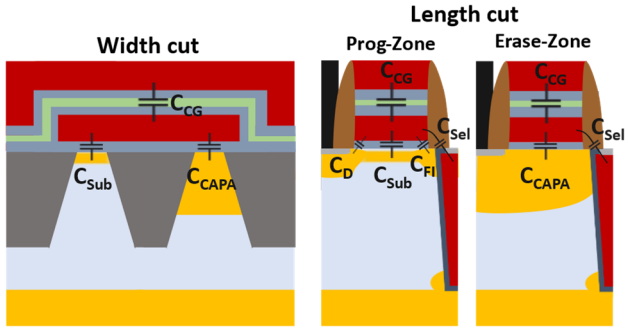


Fig. 2: Capacitive model for TSTA architecture. Wide direction cut with the poly floating gate sharing, and length cut with both programming and erasing zones.

Each α-factor is calculated considering the capacitance between the floating gate and all nodes defined in Fig. 2 for both eSTM™ and TSTA device. In Tab.1 the coupling factor results are reported for: the control gate (α_{CG}), the substrate in the Prog-Zone (α_{Sub}), the CAPA implant (α_{CAPA}), the trench selector (α_{Sel}), the floating gate neighbors (α_{FG}), the floating implant (α_{FI}) and the drain implant (α_D).

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

	eSTM	TSTA
α_{CG}	0.610	0.544
α_{CAPA}	N. A.	0.181
α_{D}	0.109	0.057
α_{Sub}	0.181	0.099
α_{FI}	0.046	0.024
α_{Sel}	0.029	0.058
α_{FG}	0.025	0.036

Tab. 1: Coupling factors for eSTM™ based architectures.

Comparing the results in Tab. 1, one can notice that $\alpha_{CG,eSTM}>\alpha_{CG,TSTA}$, then it is important to take this parameter into account for the reliability study, adapting the programming and erase electrical scheme. To study the degradation of eSTM™ tunnel oxide using the TSTA, it is crucial to have the same electrical field applied in the tunnel oxide, i.e. equal floating gate potential, then ones observed in eSTM™:

$$V_{FG,TSTA} = V_{FG,eSTM} \quad (1)$$

Thus, using this condition, the threshold voltage of TSTA ($V_{T,TSTA}$), in programmed state, using SSI scheme, can be calculated using the eSTM™ experimental threshold voltage ($V_{T,eSTM}$).

$$V_{T,TSTA} = \frac{\alpha_{CG,eSTM}}{\alpha_{CG,TSTA}}\left(V_{T,eSTM} + \frac{\sum_i \alpha_{i,TSTA} V_{i,TSTA} - \sum_i \alpha_{i,eSTM} V_{i,eSTM}}{\alpha_{CG,eSTM}}\right) \quad (2)$$

With: $V_{FG} = \sum_i \alpha_i V_i$ [5]

Considering the erased state, in order to compare different FN tunneling methods, for the TSTA device, it is necessary to calculate the threshold voltage variation as a function of applied biases to any x nodes (V_{CG} or V_{CAPA}) with the following formula:

$$\frac{\partial V_T}{\partial V_X} = \frac{1}{\alpha_{CG}}\left(\sum_i \alpha_i \frac{\partial V_i}{\partial V_X} - \frac{\partial V_{FG}}{\partial V_X}\right) \quad (3)$$

The floating gate potential at the end of erase ($V_{FG,Erase}$) operation can be calculated by:

$$V_{FG,Erase} = -\xi_{min} * t_{ox} + V_{sub} \quad (4)$$

Where ξ_{min} is the minimum electrical field in the tunnel oxide needed to start the Fowler-Nordheim tunneling: $\xi_{min}=10^7 Vcm^{-1}$ [5], t_{ox} is the tunnel oxide thickness and V_{sub} is the substrate potential.

III. ELECTRICAL ACTIVATION

This part will describe the behaviors of eSTM™ and TSTA for all electrical operations. The programming operation is performed using SSI, while erasing is done using FN tunneling mechanism. Then the threshold voltages are extracted at a fixed current of 3µA.

A. *Electrical SSI scheme*

In Fig. 3a and 3b, the SSI programming kinetics are presented for TSTA and eSTM™ respectively. For both structures, the voltage pulses are applied on the select gate, drain and control gate terminals. The select transistor works in saturated regime to generate hot electrons, while the memory transistor is in linear regime ($V_{CG}>>0V$) to transfer the drain potential to the floating implant ($V_D \approx V_{FI}$). At the same time, the control gate is biased in order to create a vertical electrical field to inject electrons in the floating gate [6]. Once can notice that with the same electrical activation conditions, in general the $V_{T,TSTA}$ is 2V lower than $V_{T,eSTM}$. This difference is related to the reading operation made in the Prog-zone while the stored charges are distributed in the entire floating gate whole the two active areas in the case of TSTA. Threshold saturation is observed for the three different V_{CG} values meaning that the electric field at oxide terminals is not enough important to inject more electrons into the floating gate. Indeed, concerning the drain current, the Fig. 3c and Fig. 3d show a weak dependence on V_{CG} variation at the beginning of the programming kinetic, while the consumption remains constant when the kinetics tend to saturate, and it is independent on the control gate bias. This is due to the main role of in-trench select transistor able to control the current flowing in the cell. To establish that the condition in equation (1) is respected, we calculate the $V_{T,TSTA}$ using (2). This implies to use the $V_{T,eSTM}$ experimentally extracted from the Fig. 3b at the end of the programming kinetic. In Tab. 2 we compare for all programming conditions, the model results and the experimental $V_{T,TSTA}$ obtained with the SSI programming scheme. The model results fit well with the experimental data, meaning that the condition (1) is verified, thus it is possible to adapt V_{CG} to reach the same programming window (PW) for both structures, that is suitable for the reliability tests.

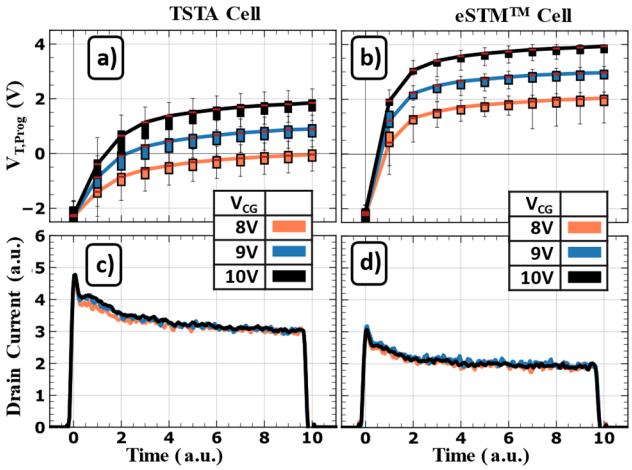


Fig. 3: (a, b) Programming kinetics, (c, d) drain current of eSTM™ and TSTA depending on V_{CG}, V_D=4.5V, V_{Sel} = 1.1V (10 samples).

			eSTM	TSTA	TSTA
V_{CG}	V_D	V_{Sel}	$V_{T,Prog}$	Calculated $V_{T,Prog}$	Measured $V_{T,Prog}$
8 V	4.5 V	1.1 V	1.93 V	-0.26 V	-0.11 V
9 V	4.5V	1.1 V	2.86 V	0.77 V	0.84 V
10 V	4.5 V	1.1 V	3.78 V	1.80 V	1.78 V

Tab. 2: Programmed threshold voltages of eSTM™ and TSTA.

B. Impact of SplitVoltage method for the erase operation

Concerning erasing methods for TSTA cell, two activations schemes will be compared to highlight the interest of this test architecture for reliability study of eSTM™. The typical scheme to erase this memory cell enables the voltage bias applied on the control gate only, as shown in Fig. 4a. We define this method as the full control gate (FullCG) FN tunnelling. By this way, electrons are ejected from the floating gate in both Prog and Erase zones, because of the constant tunnel oxide thickness. In Fig. 4b, we present the erase threshold voltage evolutions. These characteristics show the same slope $\left|\frac{\partial V_{T,Erase}}{\partial V_{CG}}\right|$ for eSTM™ and TSTA, while an offset of 2V is still obtained. This difference is coherent with the results obtained on the programming kinetics presented before.

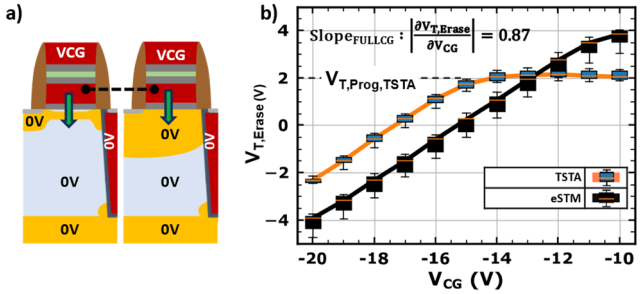


Fig. 4: (a) Schematic of FullCG erase operation. (b) Erase threshold voltage evolution (10 samples).

Thanks to the new TSTA architecture, a SplitVoltage erase method can be implemented. Fig. 5a shows a constant voltage applied to the control gate while the CAPA implant biasing is varied up to 8V to avoid the p-n junction breakdown.
Fig. 5b shows the $V_{T,Erase}$ variation for different SplitVoltage conditions. The highest electric field is applied to the Erase-Zone tunnel oxide, causing electrons to flow mainly through this specific region. Hence, the Prog-Zone oxide remains free from FN charge trapping contribution. The main goal of TSTA is thus demonstrated, using the SplitVoltage scheme it is possible to separate the injection/ejection mechanisms toward two different active areas.

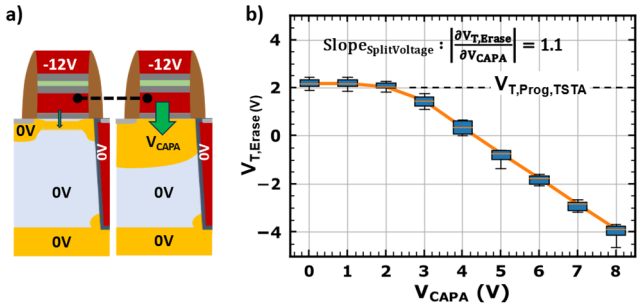


Fig. 5: (a) Schematic of erase operation (b) Evolution of erasing threshold using SplitVoltage FN tunneling at fixed V_{CG} (10 samples).

Moreover, the erase efficiency is enhanced when the characteristics in Fig. 4b and Fig. 5b are compared, showing that: $Slope_{SplitVoltage} > Slope_{FullCG}$. These two parameters are lower than those expected theoretically, respectively 1.5 and 1, using (3) and (4). The remaining difference between theoretical and measured slopes refers firstly to deep depletion of interface region over the CAPA implant [5]. The presence of the space charge region reduces the effective coupling capacitance with floating gate [7], and V_{CAPA} is different from V_{sub} in (4) and thus in (3). To conclude, this part shows how TSTA can be activated using different methods to investigate the tunnel oxide degradation and the effect on the reading operation during endurance tests.

IV. Reliability Study on TSTA

To study the degradation of both tunnel and select transistor oxides fabricated in the memory process flow we performed endurance tests of TSTA with different erasing scheme. eSTM™ and TSTA cells have been thus cycled with SSI programming and FullCG versus SplitVoltage erasing (Fig. 6a). Cycling is pushed to 10M to exacerbate cycling-related defects.

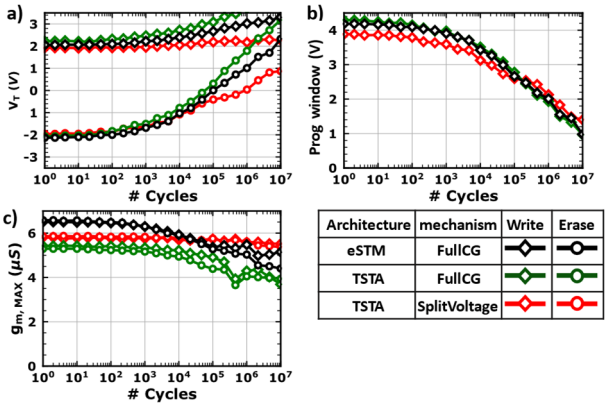


Fig. 6: Cycling of eSTM™ and TSTA using different erase mechanisms. (a) Evolution of treshold voltage (b) Programming windows and (c) maximum transconductance evolutions.

Notice that all reliability trials are carried out at different electrical conditions chosen from Fig. 3a,b, Fig. 4b and Fig. 5b, to start with a symmetrical 4V programming window (Fig. 6b) between $V_{T,Prog}$=2V and $V_{T,Erase}$=-2V for all devices. The results remain consistent regardless of the architecture or mechanism schemes employed. This approach is particularly pertinent considering the model deviation from experimental results, as demonstrated in the previous section. In the case of FullCG erasing, the PW_{TSTA} centering is degraded due to more charge trapping in the oxide volume as well as the PW_{eSTM}. Even if the TSTA has two active regions to erase the cell, an higher control gate voltage is necessary to achieve the same $V_{T,Erase}$ of eSTM™ device. In the case of SplitVoltage, once can appreciate the fact that the PW_{TSTA} centering is less impacted due to a lower read-side oxide degradation caused by FN tunneling. In Fig. 6c, the maximum of transconductance ($g_{m,MAX}$) is extracted at each reading and compared in order to study the tunnel oxide interface degradation in Prog-Zone. Comparing eSTM™ and TSTA devices in FullCG mode, it is highlighted the best performance of the new test structure where the FN mechanism is shared on two active regions.

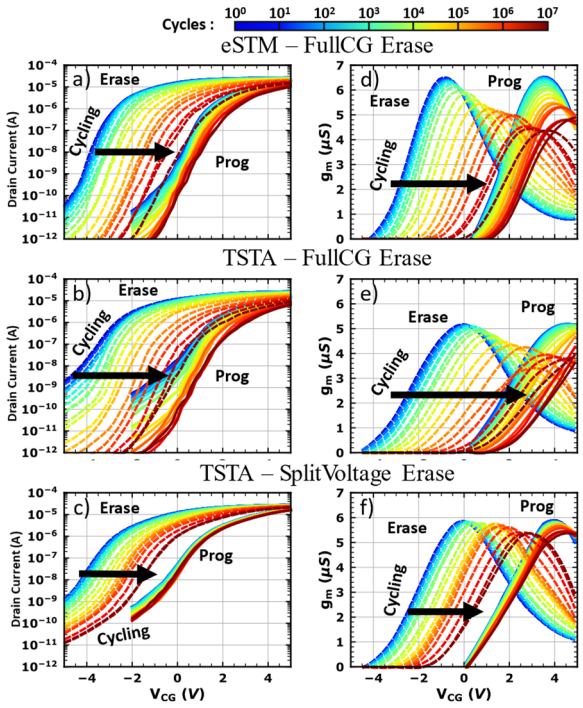


Fig. 7: I_D-V_{CG} and g_m-V_{CG} characteristics during cycling. (a), (d) of eSTM™, (b), (e) for TSTA with FullCG erase and (c), (f) for TSTA using SplitVoltage method. Select Transistor in 'ON' state: V_{Sel}=3V.

In Fig. 7a,b,d,e, I_D and g_m versus V_{CG} measured during cycling for eSTM™ and TSTA, are reported. Using FullCG scheme, both structures exhibit the same behavior in terms of programming and erase characteristics, where I_D-V_{CG} are strongly affected by charge trapping, resulting in a rigid shift of the curves [8]. Moreover, the reduction of g_m in both cases, shows a degradation of the silicon/oxide interface [9]. These degradations imply two respective impacts: a reduction of subthreshold slopes and a shift of the PW. The erase efficiency loss pushes the characteristics toward positive V_{CG}, combined with the SSI efficiency loss. To avoid the interface trap generation, the SplitVoltage scheme is used for TSTA. In this case the electrons are mainly ejected through the Erase-Zone, this implies that the electron trapping in the tunnel oxide during the SSI operation cannot be compensated, a drift of erase characteristics is thus present (Fig. 7c). On the other hand, a quasi-constant $g_{m,max}$ (Fig. 7f) demonstrates the interface defects due to the FN tunneling are localized in the Erase-Zone only, and they are not due to the SSI mechanism. The TSTA device enables to separate the interface oxide/CAPA degradation from the volume electrons trapping due to SSI mechanism that impacts the reading of memory cell as a static aging phenomenon. In conclusion, from [10], the tunnel oxide can be stressed by each operation: static aging from read operation, SSI degradation [11] at floating implant side and a non-localized FN degradation [12]. In our case we demonstrate the static aging comes mainly from programming, and the TSTA enables the dissociation of the memory cell writing operations. Moreover, to deeply understand the degradation mechanisms, we decided to characterize also the vertical select transistor during the endurance tests. In Fig. 8a,b,c it is shown the selector oxide is still intact after 10Mcycles. The current, at high V_{SEL} values, decreases during the cycling and consequently, the effect of g_m degradation (Fig. 8d,e,f) is due to the sense charge trapping and the memory threshold voltage increasing. The $g_{m,max}$ quickly decreases using FullCG for both structures (Fig. 8d,e), the interface in Prog-Zone is continuously altered by interface states leading. However, the state transistor channel should be free of any state. By using SplitVoltage, the channel of sense transistor presents a lower density of interface states in Prog-Zone (Fig. 8f).

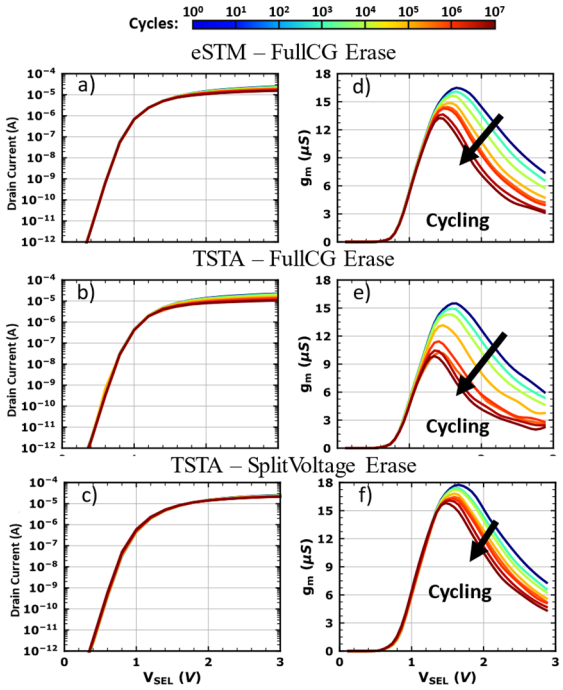


Fig. 8: I_D-V_{SEL} and g_m-V_{CG} characteristics during cycling. (a), (d) of eSTM™, (b), (e) for TSTA with FullCG erase and (c), (f) for TSTA using SplitVoltage method. Sense Transistor in 'ON' state: V_{CG} = 10V.

V. Conclusion

In this paper, a novel test structure named TSTA with two distinct active areas for SSI and FN operations is presented. The goal of this structure was to demonstrate the possibility to study oxides degradation of any floating gate based non-volatile memory cells with classical reliability characterization. To achieve it, a CAPA n-type implant has been added in the Erase-Zone allowing SplitVoltage erasing and thus to enable oxide degradation decorrelation between SSI and FN mechanisms. TSTA cell has been built following eSTM™ process. By using the capacitive model, we ensure that this test structure could be used for reliability studies. After that, endurance studies are carried out to highlight the oxides bulk trapping and interface states generation, for SSI programming and FN erase operations. Both memory cell and select transistor have been characterized during the cycling, showing the vertical selector is not impacted by the degradation. An improvement of this structure could be to add another active region to separate the reading operation, from the SSI, in order to really estimate the static cell aging.

VI. Acknowledgment

The authors gratefully acknowledge all manufacturing teams of STMicroelectronics for supporting this program.

VII. References

[1] R. Strenz, IEEE IMW 2020, pp. 1 – 4.
[2] S. Niel et al., IEEE IEDM 2018, pp. 7.4.1 – 7.4.4.
[3] F. L. Rosa et al., IEEE 11th IMW 2019, pp. 1 – 4.
[4] J. Lacord et al., *IEEE* TED 2012 pp. 1332 – 1344.
[5] A. Kolodny et al., IEEE TED 1986, pp. 835 – 844.
[6] F. Melul et al., IEEE IMW 2021, pp. 1 – 4.
[7] A. Spessot et al., IEEE EDL 2010, pp. 647 – 649.
[8] C. Papadas et al., Solid-State Electronics 1991, pp. 1375 – 1379.
[9] F. Matteo et al., DTIS 2021, pp. 1 – 5.
[10] G. Torrente et al., Microelectronics Reliability 2017, pp. 647 – 649.
[11] A. Bravaix et al., IEEE IRPS 2009, pp. 531 – 548.
[12] Y-B Park and D.K. Schroder, IEEE TED 1998, pp. 1361 –1368.

Test Structures of Cross-Domain Interface Circuits with Deep N-Well Layout to Improve CDM ESD Robustness

Huai-Min Huang and Ming-Dou Ker*
Institute of Electronics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
hmhuang.ee10@nycu.edu.tw, *mdker@ieee.org

***Abstract*— Charged-device model (CDM) electrostatic discharge (ESD) event is a complex reliability issue for integrated circuits (ICs) in advanced CMOS technology. With the development of ICs toward system-on-chip (SoC) applications, various circuit blocks have been integrated into a single chip. In order to avoid noise coupling between circuit blocks or even to reduce power consumption, the SoC chip was often equipped with separated power domains for different circuit blocks. However, the cross-domain interface circuits between different power domains are particularly susceptible to gate-oxide rupture caused by CDM ESD during cross-domain ESD events. In this study, CDM ESD robustness of cross-domain interface circuits with deep N-well (DNW) was investigated through test structures fabricated in a 0.18-µm CMOS technology.**

Keywords—Charged-device model (CDM), electrostatic discharge (ESD), cross-domain interface circuits, separated power domains, gate oxide rupture.

I. Introduction

While scaling down CMOS technologies can bring faster operation speed and lower power consumption for IC products, it also makes modern IC products more sensitive to ESD threats, especially the CDM ESD events. In contrast to the human body model (HBM) and machine model (MM) ESD events [1], where charges are transferred from the external environment into the IC, CDM involves the IC in a floating condition with accumulated charges inside its body. When some pins of the IC make contact to the ground, the accumulated charges inside the IC begin to discharge out. This phenomenon represents the ESD event where the charges are discharged from the internal part of IC to the external ground. IC with a larger chip dimension or a larger package size can store more CDM charges in itself, leading to a larger discharging current during CDM ESD events. Compared to HBM and MM ESD events, the transient current in CDM ESD is quite faster and larger [1], which increases the risk of gate-oxide damage on MOS devices. Recently, CDM ESD issues have got more attentions from many companies in the IC industry [2]. The thinner gate oxide and lower gate-oxide breakdown voltage can easily result in internal circuits damage caused by CDM ESD events [3]. In some previous studies, various CDM ESD protection designs were developed to prevent internal circuits damage during CDM ESD events. For example, the vulnerable parts of internal circuits were surrounded by deep N-well (DNW) [4], [5]. CDM charges could be blocked by DNW from the common P-substrate to avoid the ESD current flowing through the internal circuits. Another advantage of DNW to prevent internal circuits damage is due to the parasitic junctions between the common P-substrate and DNW in forward-bias mode to discharge CDM ESD current [6].

II. Cross-Domain Interface CDM ESD Damage

The cross-domain interface circuits can easily be damaged during cross-domain ESD events [7]. Due to CDM charges distributed throughout the entire chip, grounding either the VDD1 or VDD2 pin can lead to cross-domain ESD events. The schematic diagram of CDM charges damage to the interface circuits is shown in Fig. 1. In order to offer an ESD current path between the different power domains, the bi-directional diodes had been often used to connect the separated power lines [8]. The effectiveness of reducing noise coupling can be enhanced by increasing the number of series diodes on bi-directional diodes. However, a higher number of series diodes on bi-directional diodes leads to a higher voltage drop across them during cross-domain ESD events. In addition to circuit damage caused by the discharging path through the interface circuits during cross-domain ESD events, the high voltage drop generated when the discharge current flows through the power-rail ESD clamp circuit and bi-directional diodes could also lead to gate oxide breakdown in the interface circuits. The increase in the number of additional discharge paths reduces the voltage drop generated during cross-domain ESD events, thereby preventing voltage overstress on transistors in the interface circuits. In addition to the bi-directional diodes, other approaches for ESD protection in the cross-domain interface circuits were also developed. Those include the addition of local ESD protection devices [9] or even modifications of the internal circuits architecture on the receiver side of interface circuits [10].

In this work, the investigation focuses on CDM ESD protection for cross-domain interface circuits through the incorporation of deep N-well (DNW) in the interface circuits to achieve better CDM ESD robustness.

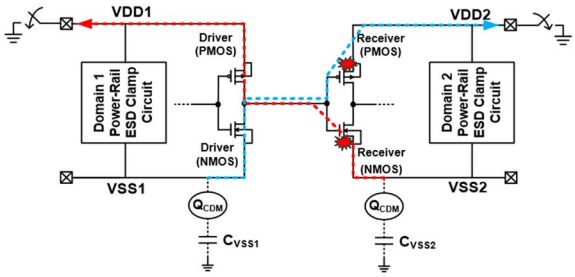


Fig. 1. The schematic diagram of simplified cross-domain interface circuits. When either the VDD1 or VDD2 pin is grounded, CDM ESD current flows from one power domain to another.

III. Cross-Domain Interface Circuits with Deep N-Well

The test circuits developed to investigate CDM ESD robustness of cross-domain interface circuits with or without deep N-well (DNW) are illustrated in Fig. 2. The baseline

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

(without DNW), as shown in Fig. 2(a), incorporates an RC-based power-rail ESD clamp circuit for each power domain, I/O ESD protection circuits, bi-directional diodes between separated VSS power lines, a pair of inverters serving as a driver and receiver for interface circuits, and input/output buffers to ensure signal driving capability. The interface circuits drawn with DNW layout are shown in Fig. 2(b), and listed as DNW21, DNW12, DNW11, and DNW22. The DNW is used to surround the P-well (PW) of NMOS in the interface circuits from the common p-type substrate. The DNW may be connected to VDD1 or VDD2 in the layout. The description of such test circuits with the DNW biased at different power supplies is listed in Table I.

The gate-grounded NMOS (Mn1) in the input ESD protection serves as ESD device for protecting against input-to-VSS ESD stress, while the gate-VDD PMOS (Mp1) serves as ESD device against input-to-VDD ESD stress. In the output ESD protection, Mn2 and Mp2 can be connected through different numbers of fingers of the output buffer to ensure adequate driving capability. The other fingers of output buffer are connected as Mn3 and Mp3, serving as ESD protection devices against output-to-VSS or output-to-VDD ESD stresses. The RC-based power-rail ESD clamp circuit [11] for each power domain, as illustrated in Fig. 3, can detect ESD events by selecting the appropriate RC time constant and subsequently activating the main ESD clamp device (Mclamp) to discharge ESD current across the power rails (VDD and VSS). Under normal power-on conditions, the ESD clamp device is kept off to prevent leakage current across the power rails. The device parameters of the test circuits are listed in Table II. The sizes of MpD and MnD in one inverter are larger than those of MpR and MnR in the other inverter, facilitating the easier detection of damage at the cross-domain interface circuits.

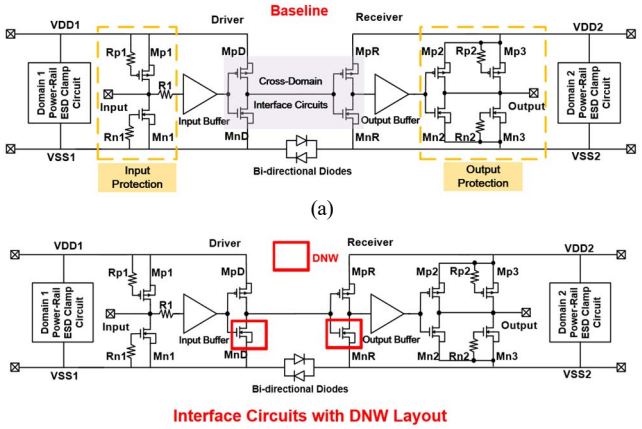


Fig. 2. The schematic diagram of test circuits: (a) interface circuits without DNW, labeled as Baseline, and (b) MnD and MnR of interface circuits with DNW biased at VDD1 or VDD2, labeled as DNW21, DNW12, DNW11, and DNW22.

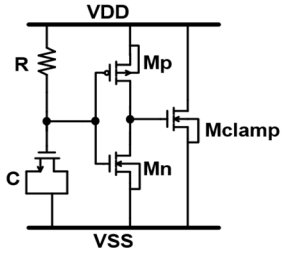


Fig. 3. The schematic diagram of the RC-based power-rail ESD clamp circuit [11].

TABLE I. COMBINATIONS OF TEST CIRCUITS

Description	Symbol	DNW Bias	
		MnD	MnR
Interface Circuits w/o DNW	Baseline	N/A	N/A
Interface Circuits w/ DNW	DNW21	VDD2	VDD1
	DNW12	VDD1	VDD2
	DNW11	VDD1	VDD1
	DNW22	VDD2	VDD2

TABLE II. DEVICE PARAMETERS OF TEST CIRCUITS

Subcircuit	Device Parameter	
Input Protection	Resistor (Ω)	Rp1 = Rn1 = 927, R1 = 311
	MOS W/L (μm/μm)	Mp1 = 360/0.25, Mn1 = 360/0.5
Cross-Domain Interface Circuits	MOS W/L (μm/μm)	MpD = 6.85/0.18, MnD = 5/0.18, MpR = 0.685/0.18, MnR = 0.5/0.18
Output Protection	Resistor (Ω)	Rp2 = Rn2 = 927
	MOS W/L (μm/μm)	Mp2 = 120/0.25, Mn2 = 60/0.5, Mp3 = 240/0.25, Mn3 = 300/0.5
RC-Based Power-Rail ESD Clamp Circuit	Resistor (Ω)	R = 94.8k
	MOS W/L (μm/μm)	C = 96.24/1.5, Mp = 120/0.25, Mn = 24/0.25, Mclamp = 700/0.18
Bi-directional Diodes	W × L (μm ×μm)	4.62 × 17.4

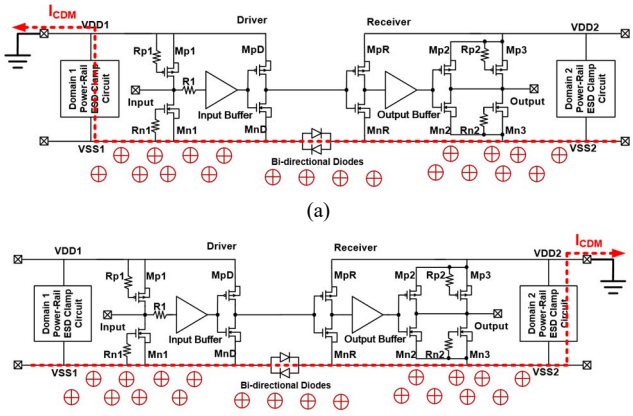


Fig. 4. The main CDM discharge path of the baseline when positive CDM charges are accumulated in the common P-substrate, when (a) VDD1 is grounded and (b) VDD2 is grounded.

The main CDM discharge current in the baseline flows through the bi-directional diodes and the power-rail ESD clamp circuit of domain 1 (or domain 2) during positive CDM discharge when the VDD1 (or VDD2) pin is grounded, as the dashed red lines depicted in Figs. 4(a) and 4(b).

In addition to the main CDM discharge path of the baseline, the DNW test circuits exhibit additional discharge paths depending on the bias connection to the DNW. For example, in the scenario where positive CDM charges are accumulated in the common P-substrate, if DNW is biased at VDD1 and the VDD1 pin is grounded, additional discharge paths are formed through parasitic junctions in forward-bias mode. Therefore, DNW21 and DNW12 each incorporate one DNW and P-substrate to create additional discharge paths in forward-bias mode. DNW11 involves two DNWs and the P-substrate, forming additional discharge paths in forward-bias mode. DNW22 does not exhibit any additional discharge paths under this condition. Those are illustrated by the dashed red lines in the cross-sectional views of the interface circuits with DNW shown in Figs. 5(a)-5(d).

Under another condition, when positive charges are accumulated in the common P-substrate, and the VDD2 pin is grounded, additional discharge paths are also formed. DNW21 and DNW12 still incorporate one DNW and P-substrate to establish additional discharge paths in the

forward-bias mode. In contrast to the scenario where VDD1 is grounded, DNW11 does not exhibit any additional discharge paths under this condition. Meanwhile, DNW22, with two DNWs and the P-substrate, creates additional discharge paths in the forward-bias mode. Those are depicted by the dashed red lines in the cross-sectional views of the interface circuits with DNW shown in Figs. 6(a)-6(d).

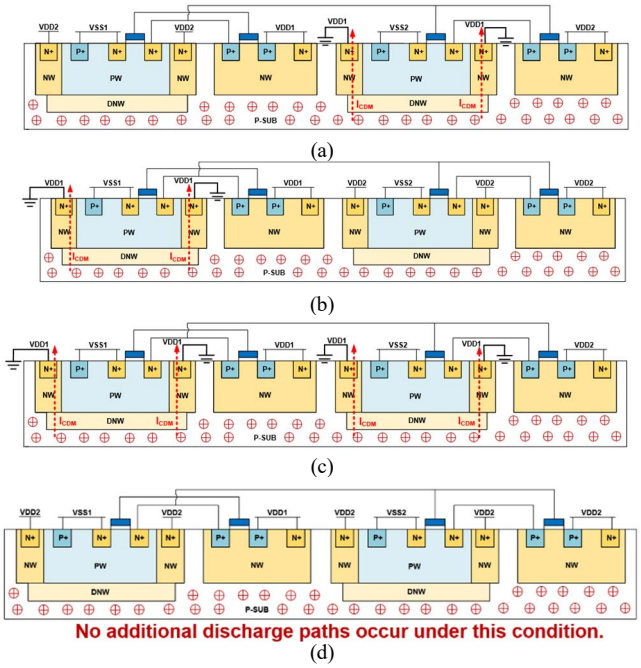


Fig. 5. When positive charges are accumulated in the common P-substrate, and the VDD1 pin is grounded, additional discharge paths are illustrated in the cross-sectional views of the interface circuits with DNW: (a) DNW21, (b) DNW12, (c) DNW11, and (d) DNW22.

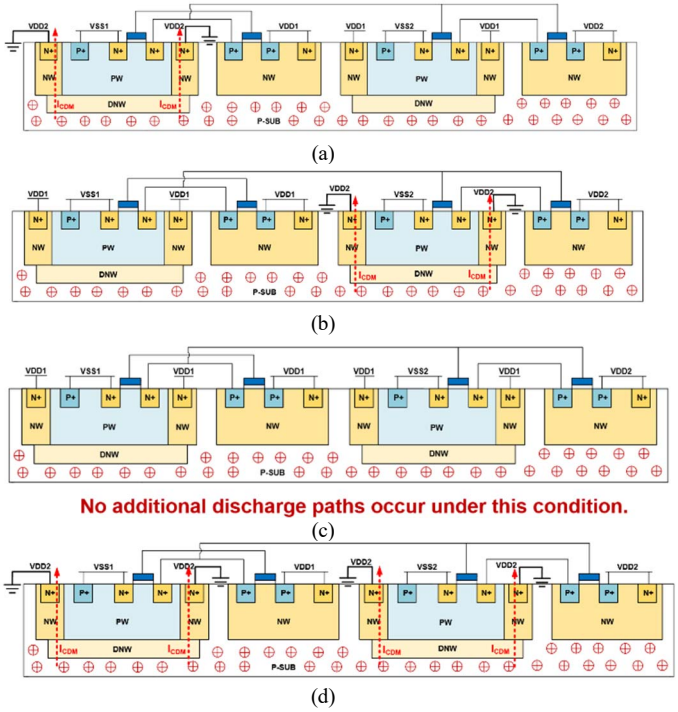


Fig. 6. When positive charges are accumulated in the common P-substrate, and the VDD2 pin is grounded, additional discharge paths are illustrated in the cross-sectional views of the interface circuits with DNW: (a) DNW21, (b) DNW12, (c) DNW11, and (d) DNW22.

During negative CDM discharge, some unexpected discharge paths may occur to cause ESD damage along the interface circuits, as the dashed red lines shown in Figs. 7(a) and 7(b), and their corresponding cross-sectional views are shown in Figs. 8(a) and 8(b), respectively. Such unexpected discharge paths can be prevented by incorporating DNW into the layout of interface circuits. Due to the high breakdown voltage of the parasitic junction between the common P-substrate and DNW, it will be difficult for the negative CDM charges in the P-substrate to flow into interface circuits. The unexpected CDM discharge paths prevented by DNW, when either VDD1 or VDD2 is grounded, are illustrated in Fig. 9(a). Its corresponding cross-sectional view is shown in Fig. 9(b).

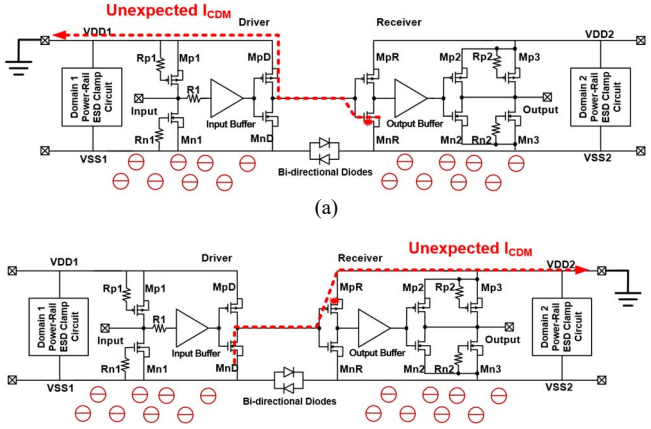


Fig. 7. The unexpected CDM discharge path in the baseline when negative CDM charges are accumulated in the P-substrate, when (a) VDD1 is grounded and (b) VDD2 is grounded.

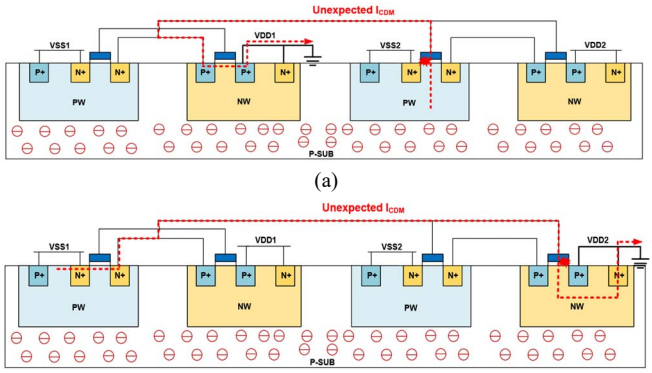


Fig. 8. The cross-sectional view shows the unexpected CDM discharge path in the baseline under the same condition of Fig. 7, when (a) VDD1 is grounded, and (b) VDD2 is grounded.

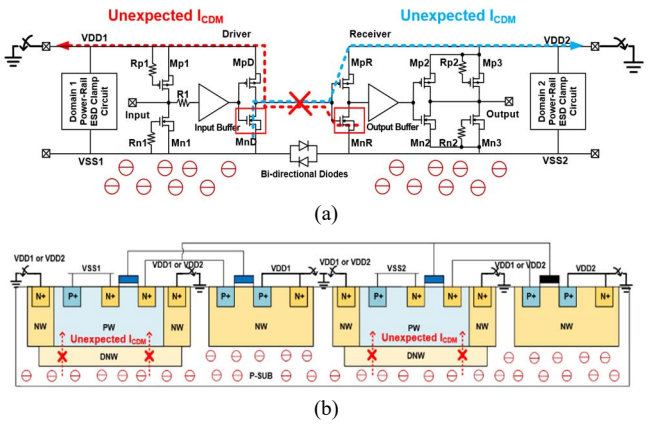


Fig. 9. (a) The schematic diagram of negative CDM discharge paths prevented by DNW when either VDD1 or VDD2 is grounded, and (b) its corresponding cross-sectional view.

IV. Experimental Results and Failure Analysis

The optical microscope (OM) micrograph of the test chip fabricated in a 0.18-μm CMOS technology is shown in Fig. 10. Each test circuit was equipped with its own pads for measurement.

The CDM ESD test is performed by grounding the VDD1 or VDD2 pin with either positive or negative CDM charges accumulated in the p-substrate, respectively. The pin combinations of CDM ESD test are listed in Table III. The CDM ESD test is conducted by using a field-induced CDM tester [12], as depicted in Fig. 11. The field plate is pre-charged to a specified voltage using the high voltage (HV) supply. Subsequently, the device under test (DUT) is fully charged through the induction process. Once the DUT is fully charged, the pogo pin is lowered to establish contact with the pin of the DUT, creating a low-impedance discharge path to ground. The accumulated charges in the DUT are then discharged to the ground. The equivalent circuit of the CDM tester with all parasitic resistors, inductors, and capacitors is also drawn in Fig. 11. C_{DUT} represents the parasitic capacitance between the DUT and the field plate, C_{DG} represents the parasitic capacitance between the DUT and the ground plate, and C_{FG} represents the parasitic capacitance between the field plate and the ground plate. The pogo pin is represented by L_{pogo} and a 1Ω resistor, and R_{spark} is employed to represent the air discharge resistance. All ESD measured results among the test circuits under positive or negative CDM ESD stresses are summarized in Table IV. In some tests, the CDM ESD test was performed multiple times, because variations in the air discharge resistance and charge distribution on the test chip might cause deviations in each individual test. Therefore, re-testing was conducted to verify the CDM ESD level for the test circuits.

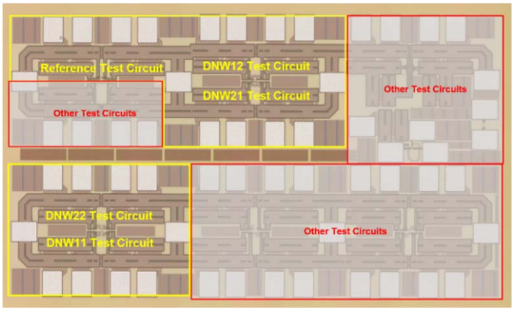


Fig. 10. The micrograph of CDM ESD test circuits on a silicon chip fabricated in a 0.18-μm CMOS technology.

TABLE III. Pin Combinations of CDM ESD Test on a Chip with Separated Power Domains

	Positive charges are accumulated in the P-sub	Negative charges are accumulated in the P-sub
VDD1 pin is grounded	VDD1 (+)	VDD1 (-)
VDD2 pin is grounded	VDD2 (+)	VDD2 (-)

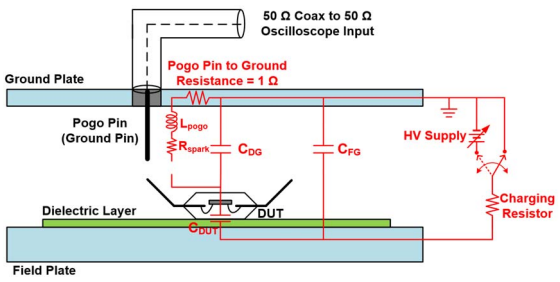


Fig. 11. A simplified diagram with the equivalent circuit of the field-induced CDM tester [12].

Before and after the CDM tests, electrical verification and functional verification are conducted. Electrical verification is to observe the DC I-V curves of the test circuits. If the leakage current under DC exceeds 10 μA, the CDM test is considered not to pass. The I-V curves of the baseline from VDD1 to VSS2 before and after the positive CDM tests on VDD1 are depicted in Fig. 12. Before the CDM test, the leakage current was minimal (in the range of nA). After the +375 V CDM test, the I-V curves began to shift. With increasing CDM test levels, the shift of I-V curves becomes more significant. Functional verification is also used to observe the function of the test circuits after the CDM test. The input signal is a periodic square wave of 1 MHz with an amplitude of 1.8 V, where the operating voltage of the test circuits is 1.8 V. The normal output signal and input signal are inverted since the signal is propagated through an odd number of inverters. The voltage waveforms before and after the positive CDM tests on VDD1 are illustrated in Fig. 13. The function of the baseline remains normal before and after the CDM tests till to 750V. It is speculated that potential damage to the gate oxide, despite causing an increase in leakage current, but does not affect the normal function. With both of electrical verification and functional verification, the CDM ESD levels among all test circuits were listed in Table IV.

TABLE IV. CDM ESD Robustness (Pass Level) of All Test Circuits

	VDD1 (+)	VDD1 (-)	VDD2 (+)	VDD2 (-)
Baseline	+250 V, +375 V, +250 V	-250 V, -125 V	+625 V	-500 V, -375 V
DNW21	+625 V, +500 V	-250 V, -250 V, -375 V	+625 V, +750 V	-500 V, -500 V
DNW12	+375 V, +500 V	-250 V, -250 V	+750 V	-500 V, -500 V
DNW11	+375 V, +375 V, +500 V	-125 V, -250 V, -250 V	+625 V	-500 V, -500 V
DNW22	+375 V, +375 V	-375 V, -375 V	+875 V, +1000 V	-750 V, -625 V

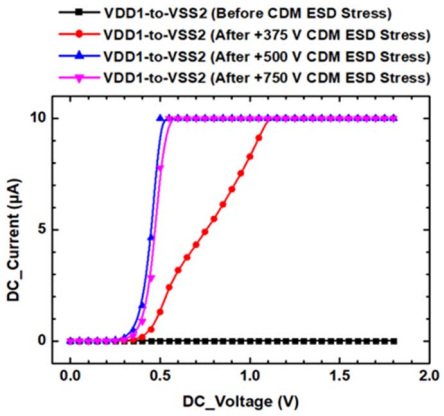


Fig. 12. Traced DC I-V characteristics from VDD1 to VSS2 of the baseline before and after CDM ESD stresses.

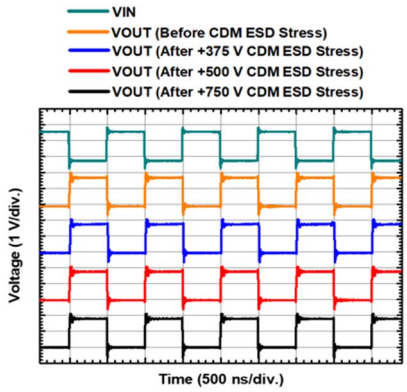


Fig. 13. Measured voltage waveforms of the baseline before and after CDM ESD stresses.

The electrical failure analysis (FA) is performed on the test circuits (after CDM ESD test) with InfraRed Optical Beam Induced Resistance Change (IR-OBIRCH) to locate the failure point. The OBIRCH mechanism is the scanning of a laser beam on the IC surface, inducing temperature fluctuations in the scanned region and leading to changes in resistance. Applying a constant voltage during the scanning process can detect the current variations. The intensity of the current change is highlighted by the hot spot, with the red hot spot representing an increase in leakage current. This methodology is valuable for investigating leakage paths within the IC.

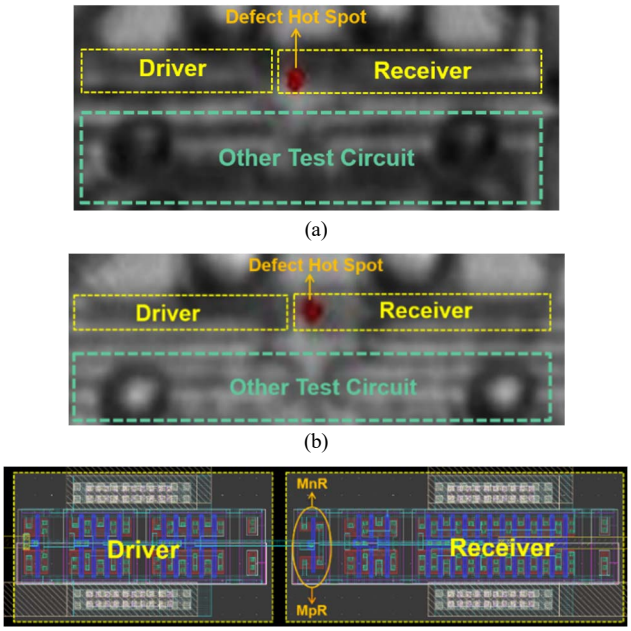


Fig. 14. IR-OBIRCH failure spots in the baseline are located on the receiver side of the interface circuits, (a) after +375 V CDM ESD test on VDD1, (b) after +750 V CDM ESD test on VDD2, and (c) the corresponding layout top view of the interface circuits.

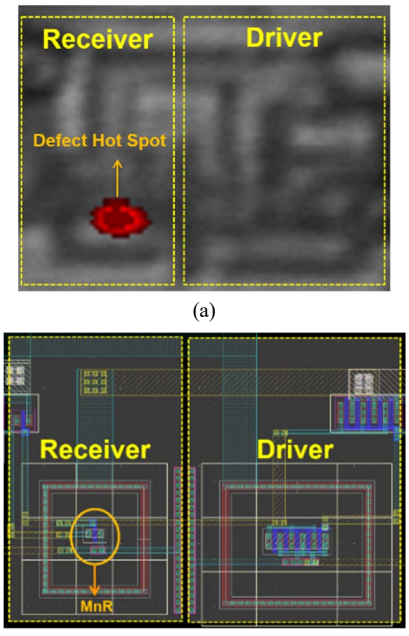


Fig. 15. IR-OBIRCH failure spots in the DNW21 are located on the MnR transistor of the interface circuits, (a) after +875 V CDM ESD test on VDD2, and (b) its corresponding layout top view of the interface circuits.

The failure points of the baseline, after +375 V CDM ESD test on VDD1 and +750 V CDM ESD test on VDD2, are located on the MpR and MnR transistors of the interface circuits, respectively. The corresponding FA pictures and layout are shown in Figs. 14(a), 14(b), and 14(c), respectively. The failure point of the DNW21, after the +875 V CDM ESD tests on VDD2, is located on the MnR transistor of the interface circuits. The corresponding FA picture and location in the layout of interface circuits are illustrated in Figs. 15(a) and 15(b), respectively. The failure point of the DNW12, after the +625 V CDM ESD tests on VDD1, is located on the MpR transistor of the interface circuits. The corresponding FA picture and location in the layout of interface circuits are illustrated in Figs. 16(a) and 16(b), respectively.

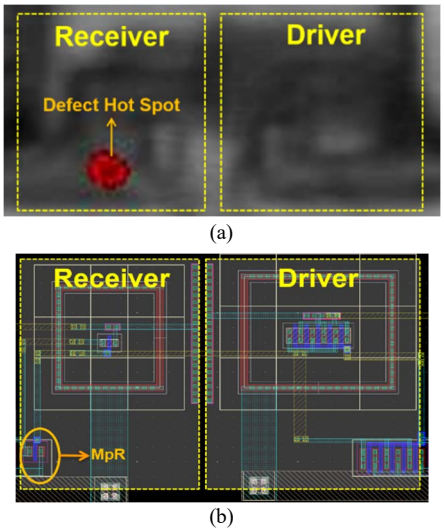


Fig. 16. IR-OBIRCH failure spots in the DNW12 are located on the MpR transistor of the interface circuits, (a) after +625 V CDM ESD test on VDD1, and (b) its corresponding layout top view of the interface circuits.

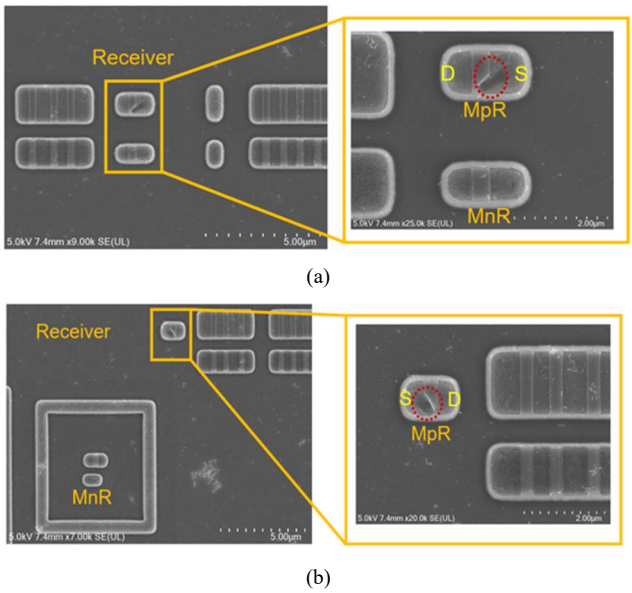


Fig. 17. The SEM images of (a) the baseline after +375 V, and (b) the DNW12 after +625 V, CDM tests on VDD1.

Based on the IR-OBIRCH failure locations, physical FA can further identify defects on the test circuits through total de-layering and scanning electron microscopy (SEM). The total de-layering process involves using a chemical solution to etch the IC layer down to the substrate layer. Subsequently,

the poly profile is observed through high acceleration voltage SEM. This allows for the observation of gate oxide breakdown in MOS devices. Similar results are observed for different test circuits after VDD1 (+) mode stress. The SEM images of failure points on the baseline and DNW12 after +375 V and +625 V CDM tests on VDD1 are presented in Figs. 17(a) and 17(b), respectively. The failure points on the gate-to-source of the MpR transistor in the receiver of the interface circuits are clearly observed in Figs. 17(a) and 17(b) through physical FA on the CDM test chip.

V. Discussion and Conclusion

The baseline of test structure demonstrates lower CDM ESD robustness, as compared to that of the test circuits with DNW structure. In positive discharge mode, the DNW test circuits exhibit enhanced CDM performance due to the presence of additional discharge paths. For instance, the DNW22 demonstrates a higher CDM level during VDD2 (+) test mode due to the presence of parasitic junctions between the common P-substrate and two DNWs in forward-bias mode, but its performance in VDD1 (+) test mode was not significantly better than that of the other DNW test circuits. This is attributed to the absence of additional parasitic paths under VDD1 (+) test mode. DNW11 shows a similar trend but in the opposite direction, as compared to that of DNW22. Therefore, DNW22 and DNW11 only improve CDM performance under certain CDM ESD test conditions.

In negative discharge mode, the DNW test circuits exhibit better CDM performance than that of the baseline. This is attributed to the ability of DNW to prevent CDM charges from flowing into the interface circuits, thereby preventing damage to the interface circuits.

While both DNW21 and DNW12 exhibit enhanced CDM performance under all pin combinations of CDM ESD tests, DNW21 introduces layout complexities due to the need for cross-connecting DNW biases to different power domains. In contrast, DNW12 offers a less complex DNW bias in the layout connection, making it a more favorable choice for DNW implementation in cross-domain interface circuits to improve CDM robustness.

Acknowledgment

This work was supported in part by the National Science and Technology Council (NSTC), Taiwan, under Contract of NSTC 112-2221-E-A49-147-MY3. The authors would like to thank *Taiwan Semiconductor Research Institute* (*TSRI*) for the support of EDA Tools for test chip layout and silicon fabrication via TSMC. The authors also express their appreciation to MA-tek (*Materials Analysis Technology Inc.*) for providing assistance with CDM ESD test and failure analysis.

References

[1] M.-D. Ker, "ESD protection design in nano CMOS," *Advanced Signal Processing, Circuits, and System Design Techniques for Communications*, *IEEE Press*, 2006, Chapter 7, pp. 217-279. (ISBN: 1-4244-0460-6)

[2] *White Paper 2: A Case for Lowering Component-level CDM ESD Specifications and Requirements Part II: Die-to-Die Interfaces*, Industry Council on ESD Target Levels, Aug. 2023.

[3] C.-Y. Lin and M.-D. Ker, "CDM ESD protection design with initial-on concept in nanoscale CMOS process," in *Proc. Int. Symp. Phys. Failure Anal. Integr. Circuits*, 2010, pp. 1-4.

[4] M.-D. Ker, C.-Y. Lin, and T.-L. Chang, "Layout styles to improve CDM ESD robustness of integrated circuits in 65-nm CMOS process," in *Proc. Int. Symp. VLSI Design, Autom. Test*, 2011, pp. 374-377.

[5] J.-H. Lee *et al.*, "The study of sensitive circuit and layout for CDM improvement," in *Proc. Int. Symp. Phys. Failure Anal. Integr. Circuits*, 2009, pp. 228-232.

[6] Y.-C. Huang and M.-D. Ker, "Study on CDM ESD robustness among on-chip decoupling capacitors in CMOS integrated circuits," *IEEE J. Electron Devices Soc.*, vol. 9, pp. 881-890, 2021.

[7] M. Okushima, "ESD protection design for mixed-power domains in 90 nm CMOS with new efficient power clamp and GND current trigger (GCT) technique," in *Proc. EOS/ESD Symp.*, 2006, pp. 205-213.

[8] M.-D. Ker, "ESD protection circuit for mixed mode integrated circuits with separated power pins," U.S. Patent 6075686 A, Jun. 13, 2000.

[9] S.-H. Chen, M.-D. Ker, and H.-P. Hung, "Active ESD protection design for interface circuits between separated power domains against cross-power-domain ESD stresses," *IEEE Trans. Device Mater. Rel.*, vol. 8, no. 3, pp. 549-560, Sep. 2008.

[10] C.-Y. Hsueh and M.-D. Ker, "Stacking-MOS protection design for interface circuits against cross-domain CDM ESD stresses," *IEEE Trans. Electron Devices*, vol. 68, no. 4, pp. 1461-1470, Apr. 2021.

[11] M.-D. Ker, "Whole-chip ESD protection design with efficient VDD-to-VSS ESD clamp circuits for submicron CMOS VLSI," *IEEE Trans. Electron Devices*, vol. 46, no. 1, pp. 173-183, Jan. 1999.

[12] *Electrostatic Discharge (ESD) Sensitivity Testing: Charged Device Model (CDM)-Device Level*, ESD Association, Standard ANSI/ESDA/JEDEC JS-002-2018, 2018.

A 4H-SiC Trench MOS Capacitor Structure for Sidewall Oxide Characteristics Measurement

Huai-Lin Huang, Li-Tien Hsuesh, Yen-Cheng Tu, and Bing-Yue Tsui
Institute of Electronics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan, R.O.C.
lynnhuang3026.ee11@nycu.edu.tw

***Abstract*—Test structure for evaluating gate oxide properties on the trench sidewall in 4H-SiC is proposed. Using the thick bottom oxide and poly-Silicon spacer structure, we are able to measure the capacitance characteristics directly and extract the interface state density. It is observed that typical NO annealing process cannot passivate the trench etching induced defects effectively.**

I. Introduction

Silicon carbide has attracted a lot of attention due to its superior properties on the high-power applications. With high critical electric field, wide band gap, and high thermal conductivity, these features allow SiC MOSFETs to have better performance than the silicon devices with the same breakdown voltage as switching power devices [1]. Among several different types of power MOSFETs, trench gate MOSFET (UMOSFET) can achieve smaller cell pitch and thus lower on-resistance compare to the vertical double-implantation MOSFET (VDMOSFET). Furthermore, the $(11\bar{2}0)$ and $(1\bar{1}00)$ faces on trench sidewall have higher bulk mobility than the (0001) face on the planar surface and should provide better channel mobility.

However, there are still many problems while fabricating SiC UMOSFETs, and one of the most concerned issues is the gate oxide integrity on the trench sidewall, which are rather hard to be identified and measured. Complicated evaluation or analyzation is often needed to obtain the characteristics. J. Berens et al. presented the determination of trench MOSFET oxide interface states by thermal dielectric relaxation current [2]. The extraction procedure is more complex than the conventional capacitance-voltage methods. Z. Guo et al. proposed using several trench MOS capacitors with variable bottom, sidewall, and mesa length to extract the sidewall capacitance [3]. This requires multiple samples and undesirable indirect evaluation.

In this work, we propose a trench MOS capacitor test structure that can help obtain the gate oxide characteristics on trench sidewall directly. Using thick bottom oxide and poly-Si spacer structure, the measurement results should have rather small parasitic capacitance influence. The capacitance characteristics and interface states can hence be observed without complicated extraction procedure.

II. Test Structure Design

The cross-sectional schematic of the trench MOS capacitor is shown in Fig. 1. The lateral and vertical dimensions are listed in Table 1. For more accurate measurement results of the sidewall capacitance, the parasitic capacitance must be minimized. Thus, thick bottom oxide (TBOX) and thick interlayer dielectric (ILD) are required.

Table 1 Lateral and vertical dimensions of the trench MOS capacitor used in TCAD simulation.

Item	value	unit
Trench depth	1.8	µm
Trench width	2	µm
GOX thickness	50	nm
TBOX thickness	300	nm
FOX thickness	0.8	µm
N_{drift}	1×10^{16}	cm^{-3}

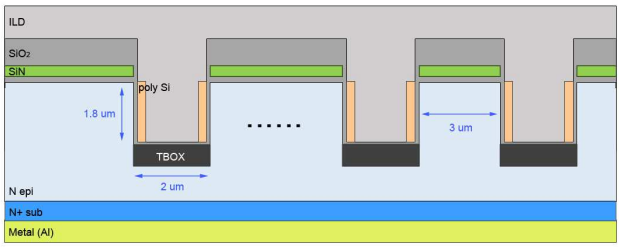


Fig. 1. Schematic cross-sectional structure of the trench MOS capacitor.

To eliminate the parasitic capacitance of probing pad, the pad is located right above the trench array as shown in Fig. 2. Since metal pad and poly-silicon are connected, there is no voltage difference between the two. The parasitic capacitance of the pad is therefore negligible.

The parasitic capacitance can be evaluated by TCAD simulation [4]. Fig. 3 shows the schematic of a unit cell used in TCAD simulation. The trenches are 500 µm in length and 1.8 µm in depth. The total gate oxide area of the sidewall MOS capacitor is 90000 μm^2 as the trench number is 50. The sidewall capacitance and parasitic capacitance are summarized in Table 2. The parasitic capacitance consists of fringing capacitance at top corner and bottom corner, capacitance at trench bottom, and capacitance of field oxide. The B'-B part structure, which has less parasitic capacitance, is the major part of the trench layout so the data measured can focus on the sidewalls capacitance wanted. Accuracy close to 90% can be achieved by direct measurement. The accuracy can be further improved by

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

Table 2 Capacitance components of the trench MOS capacitor from TCAD Simulation.

Structure Region	Capacitance (pF)	Proportion in each section
A' – A region structure (length 28 µm)		
2 sidewalls	3.319	71.9%
parasitic	1.298	28.1%
B' – B region structure (length 522 µm)		
2 sidewalls	62.11	89.3%
parasitic	7.423	10.7%
Overall layout (length 550 µm)		
sidewalls	65.43	**88.2%**
parasitic	8.721	11.8%
Total	74.15	100%

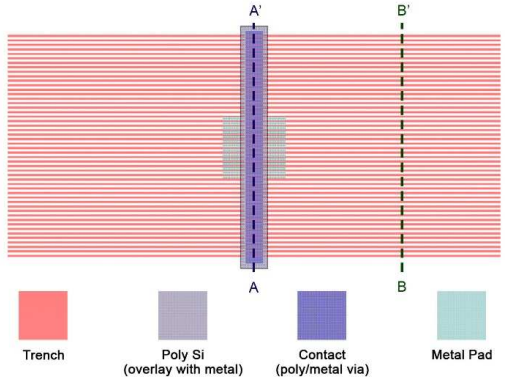


Fig. 2. Layout of the trench MOS capacitor. In the design, the gate electrode pad is located on the top of trench array to minimize parasitic capacitance.

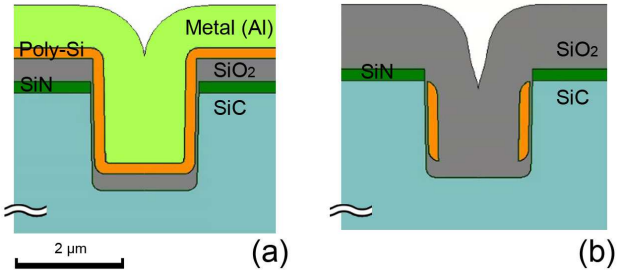


Fig. 3. Schematic cross-sectional structure of the trench MOS capacitor in TCAD simulation. (a) A'-A cross section and (b) B'-B cross section.

increasing the trench depth to create larger sidewall gate oxide area and increasing the TBOX thickness to suppress the capacitance at trench bottom.

III. Evaluation Method

To evaluate the interface density of states (D_{it}), typical high-low capacitance-voltage (C-V) method is often used [5, 6]. Due to the interface states, a difference between high frequency capacitance and low frequency capacitance is often observed. It should be noted that linearly ramping voltage method often acts as low frequency measurement of capacitance.

Interface density of states can be determined by the equation:

$$D_{it} = \frac{1}{Aq}\left[\left(\frac{1}{C_{LF}} - \frac{1}{C_{ox}}\right)^{-1} - \left(\frac{1}{C_{HF}} - \frac{1}{C_{ox}}\right)^{-1}\right] \quad (1)$$

where C_{ox} is oxide capacitance, C_{LF} and C_{HF} are low and high frequency capacitance measured, A is capacitor area, and q is the electron charge.

In this method, capacitance in the accumulation region is usually used as C_{ox}. However, this means all C_{ox}, C_{LF}, and C_{HF} will all be under the influence of parasitic capacitance. Assume parasitic capacitance C_p does not change with the voltage and frequency applied, the D_{it} obtained from measurement is:

$$D_{it,meas.} = \frac{1}{Aq}\left[\left(\frac{1}{C_{LF}+C_p} - \frac{1}{C_{ox}+C_p}\right)^{-1} - \left(\frac{1}{C_{HF}+C_p} - \frac{1}{C_{ox}+C_p}\right)^{-1}\right], \quad (2)$$

where C_p can be estimated from simulation.

By eqs. (1) and (2), the influence of C_p can be calculated as:

$$\frac{D_{it,meas}}{D_{it}} = \left(\frac{C_{ox}+C_p}{C_{ox}}\right)^2 \quad (3)$$

It is clear that overestimation of D_{it} will occur due to the existence of parasitic capacitance. A 10% parasitic capacitance would be results in a 21% overestimation of D_{it}. With C_p obtained from the simulation, some correction can be made while evaluating the interface states density to obtain results closer to reality.

IV. Test Structure Fabrication

The main process flow is shown in Fig. 4 while the schematics of several key process steps are shown in Fig. 5. The test structure was fabricated on a Si-face 4H-SiC wafer with a 5-µm-thick epi-layer of 1×10^{16} cm^{-3} doping concentration on heavily doped n-type 4H-SiC substrate. Thermal oxidation followed by LPCVD Si_3N_4 and SiO_2 deposition were used to form an etching hardmask (Fig. 5(a)). The Si_3N_4 layer is utilized to prevent oxidation of the SiC top surface during the TBOX

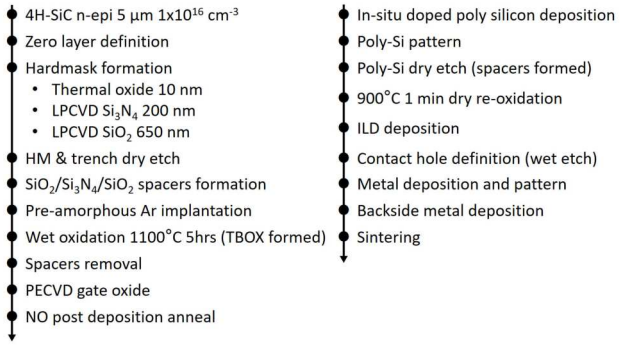


Fig. 4. Process flow of the trench MOS capacitor.

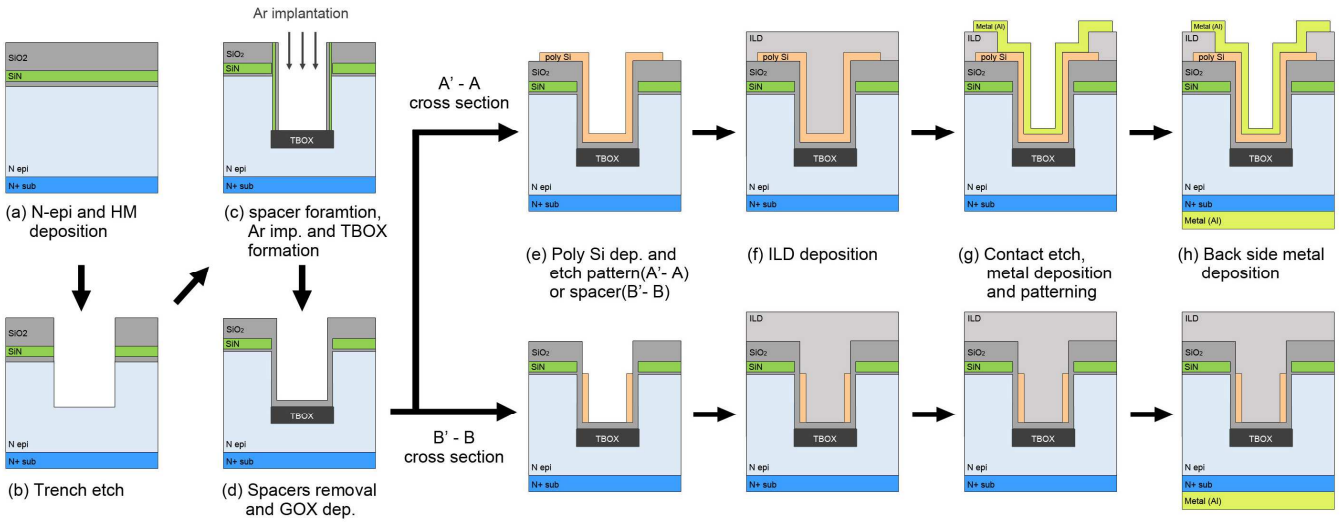


Fig. 5. Schematic cross-sectional structures of the trench MOS capacitor at main process steps.

oxidation step. Plasma etching were used to etch hardmask and SiC trench to a depth of 2 μm (Fig.5(b)).

Thermal SiO_2/LPCVD Si_3N_4/LPCVD SiO_2 spacer on the trench sidewall were formed by thermal oxidation and LPVCD followed by dry etching to prevent the sidewall from being damaged by the subsequent argon ion implantation process. The pre-amorphous Ar implantation with several different energy and dose were performed to obtain an implant damage profile that covered the trench bottom and corners. Wet oxidation of the pre-amorphous layer was carried out under 1100°C for 5 hours, created a thick bottom oxide TBOX (Fig. 5(c)). This process is similar to the local oxidation of SiC (LOCOSiC) process [7-9]. The trench depth after TBOX formation was about 1.8 μm. A 50-nm-thick gate oxide on the sidewall was deposited by PECVD after the removal of the nitride and oxide spacer (Fig. 5(d)). NO post deposition annealing was performed at 1200°C for 30 minutes, followed by an in-situ doped poly-Si deposition as gate electrode.

Poly-Si was patterned by dry etch, which created the sidewall gate electrode at the B'-B part of layout (Fig. 5(e)). A 550-nm-thick SiO_2 was deposited by LPCVD to act as the interlayer dielectric ILD (Fig. 5(f)). Contact hole on A'-A part of layout were defined and opened by BOE etching. Aluminum was then deposited and patterned as probing pad (Fig. 5(g)), finished with the backside aluminum electrode (Fig. 5(h)). A 400°C sintering in N_2 ambience was performed after the completion of the device structure.

V. Results and Discussions

Fig. 6 shows the cross-sectional SEM photograph of the fabricated trench array along the A'-A plane after ILD deposition. Since it was in the gate contact region, poly-Silicon covers the whole trench array. It can be seen that the TBOX was successfully formed on the bottom, which can help reduce the capacitance of the bottom oxide and protect the trench corner from early breakdown.

It was observed that during the TBOX formation, H_2O molecules likely went through the pad oxide located between nitride spacer and SiC, oxidized the top trench sidewall and cause a slight taper. The discard of the pad-oxide process can improve the overall trench profile.

For the D_{it} extraction, high-low frequency C-V method was used, with 100 kHz as high frequency signal. The C-V measurement results are shown in Fig. 7. It can be seen that a typical MOS C-V characteristics were successfully obtained. The separation of the two C-V curves in depletion region indicates high density of states. The D_{it} extracted from the C-V data in Fig. 7 is shown in Fig. 8. Using the simulation results from section II, a parasitic capacitance C_p of 11.8% of C_{ox} is assumed. The black line is the data evaluated straight from the C-V curve, while the blue line is the results after the correction by equation (3). It can be seen that C_p has caused overestimation of D_{it} as expected.

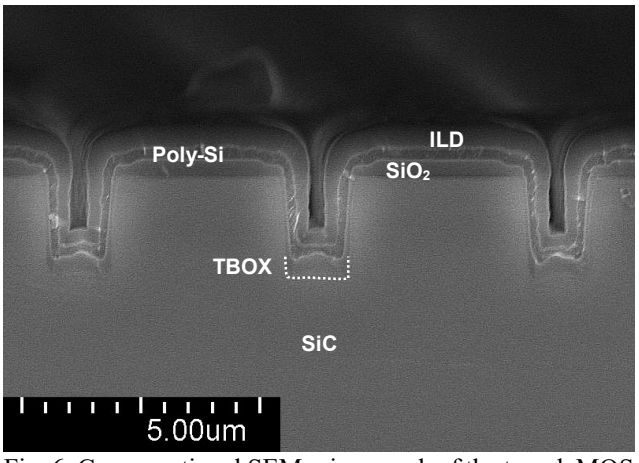


Fig. 6. Cross-sectional SEM micrograph of the trench MOS capacitor along the A'-A plane after ILD deposition.

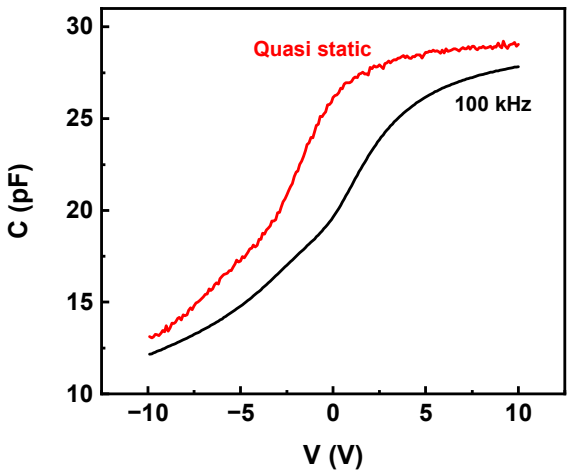


Fig. 7. Measured high frequency (100 kHz) and quasi-static capacitance-voltage (C-V) characteristics of the trench MOS capacitor.

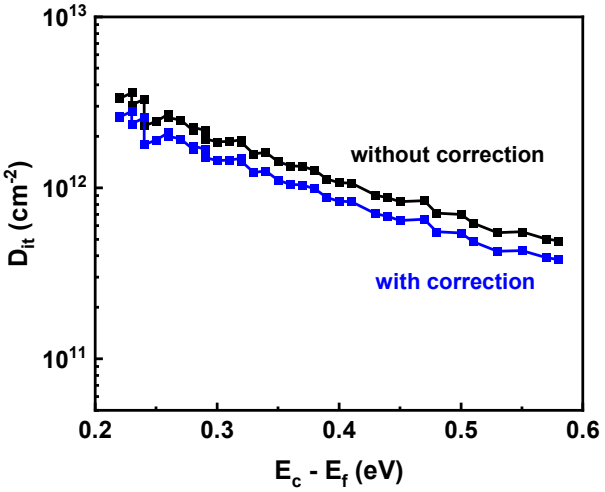


Fig. 8. Interface state density (D_{it}) at the energy level of 0.2 ~ 0.6 eV below conduction band obtained from Fig. 7.

However, even with correction, the D_{it} at the energy level of 0.2 eV below conduction band (E_c) is still around 3×10^{12} $eV^{-1}cm^{-2}$ and decreases gradually to deeper level. Even the D_{it} at Ec-0.6 eV is still higher than 2×10^{11} $eV^{-1}cm^{-2}$.

It should be noted that the same NO annealing process can reduce the D_{it} at E_c-0.2 eV to be lower than 1×10^{12} $eV^{-1}cm^{-2}$ on the (0001) face [10]. It appears that the passivation effect of NO annealing on the etched sidewall is not as good as that on the un-etched flat surfaces [11]. Furthermore, it is believed that the plasma etching induced defects plays a major role on the high D_{it} [12]. With the rather high etching bias power used for SiC trench etch, the damage induced by the process might have caused the formation of large interface defects and could not be eliminated easily by the oxidation process and NO annealing. Further improvement is required to optimize the interface quality.

The high D_{it} also explains the channel mobility on the trench sidewall being similar to that on the planar (0001) face [13]. Despite the theoretically high bulk mobility, the defects induced by etching process and the inefficiency of NO post oxidation annealing used in conventional SiC MOSFETs limit the potential of SiC UMOSFETs.

VI. Summary

A trench MOS capacitor structure for electrical characteristics measurement of gate oxide on trench sidewall is proposed in this work. Design concept and fabrication process are presented in detail. Using the TBOX and poly-Si spacer structure, as well as layout design that places probing pad directly above the trench array, parasitic capacitance is lowered to 11.8% and the sidewall characteristics can hence be obtained directly.

Furthermore, with the direct access to C-V characteristics, D_{it} can be easily extracted. The effect of the parasitic capacitance C_p on D_{it} extraction has also been discussed. It's shown that C_p can cause overestimation of D_{it} measured. By using TCAD simulation to estimate C_p, the effect of parasitic capacitance can be removed from D_{it} calculated to obtain the value that is closer to the reality.

In this paper, it is also revealed that typical NO annealing process cannot passivate the trench etching induced D_{it} on trench sidewall effectively. Interface states density is several times higher on the trench sidewall than it is on planer 4H-SiC (0001) face MOS. This lowers the channel mobility of SiC UMOSFETs and limits the advantage of the higher bulk mobility of (11$\bar{2}$0) face. Advanced etching and defect removing processes are urgently needed to improve the performance of SiC UMOSFET.

Acknowledgement

This work was supported by the National Science and Technology Council, Taiwan, under the contract No. NSTC 112-2218-E-A49-017.

References

[1] Zhiyu Guo, Jingmin Wu, Run Tian, Fengxuan Wang, Pengfei Xu, Xiang Yang, Zhongchao Fan, Fuhua Yang, and Zhi He, "Extraction of the Trench Sidewall Capacitances in an n-Type 4H-SiC Trench Metal–Oxide–Semiconductor Structure," IEEE Transactions on Electron Devices, Vol. 68, No. 6, p.2879, 2021.

[2] Judith Berens, Fabian Rasinger, Thomas Aichinger, Michael Heuken, Michael Krieger, and Gregor Pobegen, "Detection and Cryogenic Characterization of Defects at the $SiO_2$4H-SiC Interface in Trench MOSFET, " IEEE Transactions on Electron Devices, Vol. 66, No. 3, p.1213, 2019.

[3] Zhiyu Guo, Zhi He, Fengxuan Wang, Jingmin Wu, Xiang Yang, Zhongchao Fan, and Fuhua Yang, "Study of Vertical Capacitance in an n-Type 4H-SiC Stepped Thick-Oxide Trench MOS Structure," IEEE Transactions on Electron Devices, Vol. 69, No. 8, p.4617, 2022.

[4] TCAD Sentaurus Device Manual, Synopsys, Inc.: Mountain View, CA, 2020.

[5] Menghua Wang, Mingchao Yang, Weihua Liu, Jinwei Qi, Songquan Yang, Chuanyu Han , Li Geng, Yue Hao, "A Highly Efficient Annealing Process With Supercritical N_2O at 120°C for SiO_2/4H–SiC Interface," IEEE Transactions on Electron Devices, Vol. 68, No. 4, p.1841, 2021.

[6] Razvan Pascu1, Florea Craciunoiu, Mihaela Kusko, Mihai Mihaila1, Gheorghe Pristavu, Marian Badila, Gheorghe Brezeanu, "SiO_2/4H-SiC

interface states reduction by $POCl_3$ post-oxidation annealing," International Semiconductor Conference, 2015.

[7] Y. H. Tseng, C. Y. Lin, and B. Y. Tsui, "Characterization of LOCOS Field Oxide on 4H-SiC Formed by Ar Pre-amorphization Ion Implantation," IEEE Electron Device Lett. vol.38, no.6, pp.798-801, 2017.

[8] B. Y. Tsui, Y. R. Chung, and J. C. Cheng, "Breakthrough of LOCal Oxidation of SiC (LOCOSiC) Process for Sub-micron 4H-SiC Integrated Circuits," in the Extended Abstracts of the 2019 International Conference on Solid State Devices and Materials, Nagoya, 2019, pp737-738.

[9] B. Y. Tsui, Y. R. Jhuang, J. H. Lin, Y. T. Huang, T. K. Tsai, K. T. Hsu, Y. H. Su, and Y. F. Hsieh, "Device isolation process for 4H-SiC CMOS ICs," presented in the 6th IEEE Electron Devices Technology & Manufacturing Conference (EDTM), 2022, pp. 238-240

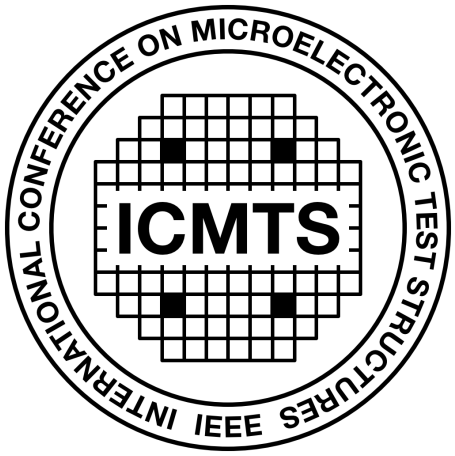
[10] B. Y. Tsui, Y. T. Huang, T. L. Wu, C. H. Chien, "Time-dependent dielectric breakdown of gate oxide on 4H-SiC with different oxidation processes," Microelectronics Reliability, vol.123, no.8, p.114186, 2021.

[11] K. Tachiki, K. Mikami, K. Ito, M. Kaneko, and T. Kimoto, "Mobility enhancement in heavily doped 4H-SiC (0001), $(11\bar{2}0)$, and $(1\bar{1}00)$ MOSFETs via an oxidation-minimizing process," Applied Physics Express, vol.15, p.071001, 2022.

[12] C. H. Wang, L. J. Lin, C. L. Hung, Y. K. Hsiao, and B. Y. Tsui, "Influence of Active Area Etching Method on the Integrity of Gate Oxide on 4H-SiC," presented in the International Conference on Silicon Carbide and Related Materials (ICSCRM) 2023.

[13] K. Kutsuki, E. Kagoshima, T. Onishi, J. Saito, N. Soejima1, and Y. Watanabe, "Experimental Investigation and Improvement of Channel Mobility in 4H-SiC Trench MOSFETs," on IEEE International Electron Devices Meeting, 2019.

SESSION 3
Cryogenic Characterisation



16th April 2024, 14:20–15:20

Session Co-Chairs: Stewart Smith, *The University of Edinburgh, UK*
Wuxia Li, *NXP Semiconductors, China*

Transistor Matrix Array for Measuring Variability and Random Telegraph Noise at Cryogenic Temperatures

Tomoko Mizutani[1], Kiyoshi Takeuchi[1], Takuya Saraya[1], Hiroshi Oka[3], Takahiro Mori[3], Masaharu Kobayashi[1,2], and Toshiro Hiramoro[1]

[1]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
Email: mizutani@nano.iis.u-tokyo.ac.jp
[2]Systems Design Lab (d.lab), The University of Tokyo, Tokyo, Japan
[3]National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki, Japan

***Abstract*— Addressable transistor arrays using 65 nm bulk technology were fabricated and tested at cryogenic temperatures. It was confirmed that variability at 1.5 K slightly degrades compared with 300 K. Random telegraph noise (RTN) was also measured and existence of extremely slow RTN at 1.5 K was confirmed using a quasi-parallel measurement technique. Such test structures will be particularly useful for enhancing cryogenic measurement efficiency.**

Keywords—Cryogenic, CMOS, variability, random telegraph noise (RTN)

I. Introduction

Quantum computers are rapidly evolving toward practical applications. To realize classical/quantum interface of a quantum computer with a large number of qubits, CMOS circuits situated close to the qubits, operating at deep-cryogenic temperatures are needed [1]-[4]. To design such circuits, it is necessary to understand not only typical characteristics of Cryo-CMOS devices [5]-[8], but also their variability [9]-[15]. It is well known that addressable arrays are very effective for efficient variability measurements, and have been applied also to cryogenic measurements [12]-[15]. It is considered that advantage of such arrays is more pronounced than at room temperature, due to limited availability of cryogenic probers, probe contact resistance issues, etc.

In this work, test chips containing transistor arrays were fabricated and tested at cryogenic temperatures. It is shown that the arrays are useful for measuring variability of low frequency noise (or random telegraph noise; RTN), in addition to that of static characteristics.

II. Experimental

Devices to be tested (i.e., DUTs) in this work are 65 nm bulk technology n and p-channel MOS transistors. Two types of unit cells were adopted, which are (a) fully isolated and (b) common source/gate types (Fig. 1). While the former allows interchanging source and drain, and can isolate defective DUTs, the latter allows applying a constant voltage to all the gates, and can be used for special noise measurements as detailed later. Lower parasitic resistance is another advantage of (b). All the selector switches are fabricated by 3.3V IO n-channel transistors, whose channel length L and width W were carefully chosen, so that accurate Kelvin (four terminal) measurements can be achieved using standard parametric testers. Note that too high access resistance degrades analog feedback operation of such testers [16]. Resistances of access wires, which tend to be rather high for suppressing heat in-flow, was taken into account. L and W of the DUTs were varied (Table I). Test chip layout comprising four 8k banks is shown in Fig. 2.

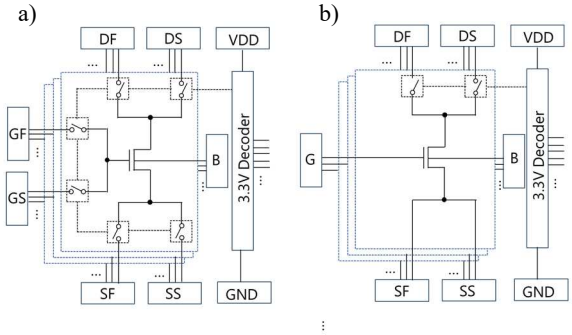


Fig. 1. Two types of addressable until cells containing a device under test (DUT): (a) c isolated, and (b) common source/gate.

TABLE I. DUT SIZES INCLUDED IN ARRAY

L/W	120 nm	200 nm	400 nm	1000 nm
65 nm	○	○	○	○
100 nm		○		
200 nm			○	
500 nm	○			○

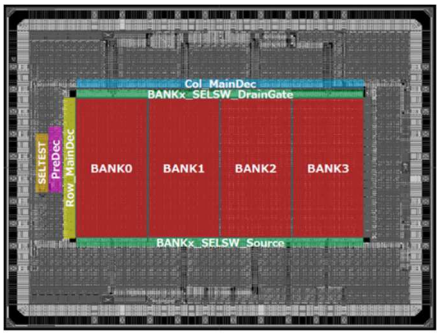


Fig. 2. Chip layout.

Fabricated chips were mounted on ceramic packages and attached to a sample rod of a refrigerator system. Force and sense terminals of the tester were directly connected to the Force/Sense pads of the chip, which is then connected to a specific DUT on-chip via switch transistors activated by the decoder circuit. The decoder address was supplied from a programmable multi-channel voltage source. First, effectiveness of the Kelvin connection was examined. Fig. 3 shows drain current (Id) vs. gate voltage (Vg) curves of both a single device with four pads and a device in the array. In order to suppress the influence of random dopant fluctuation (RDF), large-sized devices (L/W = 500 nm/1000 nm) were used. Without the sense connection, Id of the array device is substantially lower than that of the single device due to high access resistances (Fig. 3a), while if the four-terminal connection is used, Id-Vg curves of the two devices are in reasonable agreement (Fig. 3b). This confirms that the parasitic resistance of on-chip switches and long wires along the rod can be properly compensated.

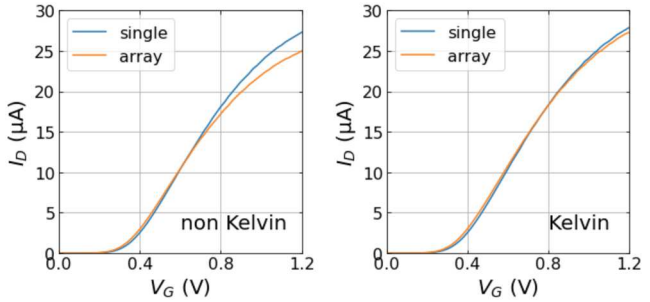


Fig. 3. Id-Vg curves of single and array devices; a) without sense connection, b) with sense connection. Vd=50 mV.

III. Results and Discuttion

A. *Variability of static characteristics*

Using the test chip, dc characteristics of n and p-channel MOSFETs were measured. Fig. 4 shows Id-Vg curves of small (L/W = 65 nm/120 nm) and large (L/W = 500 nm/1000 nm) channel n/p MOSFETs, measured at 1.5 K and 300 K. Curves of 256 devices are drawn on each figure. At 1.5 K, the lowest current level is below 1 pA. This indicates that the measurements are not appreciably interfered by the leakage of switch transistors. From Fig. 4, it is apparent that, though subthreshold current and slope are reduced at 1.5 K, horizontal spread of the current is increased compared with 300 K, suggesting increased threshold voltage (Vth) variability. Fig. 5 shows normal quantile plots of Vth extracted by constant current (0.1μA×W/L) and extrapolation methods (denoted V_{THC} and V_{THEX}, respectively). Vth is normally distributed in all cases. Fig. 6 shows Vth standard deviation vs. $1/\sqrt{LW}$ plots for both V_{THC} and V_{THEX}. The data mostly fall on straight lines passing the origin, with some exceptions. It can be confirmed that Vth variability at 1.5 K is larger than at 300 K. Interestingly, V_{THC} exhibit larger variability increase than V_{THEX}. To examine this in more detail, variability of current onset voltage (COV) V_{THEX} -V_{THC} (see Fig. 7) is plotted in Fig. 8. It can be seen that COV variability at 1.5 K is more than doubled compared with 300 K, which is consistent with an earlier report [11]. This can be attributed to increased subthreshold percolation effect caused by the loss of thermal excitation. Fig. 9 shows Id standard deviation normalized by mean Id vs. $1/\sqrt{LW}$ plots for both linear and saturation regions. It was found that linear Id variability substantially increases at 1.5 K, while saturation Id variability degrades only slightly.

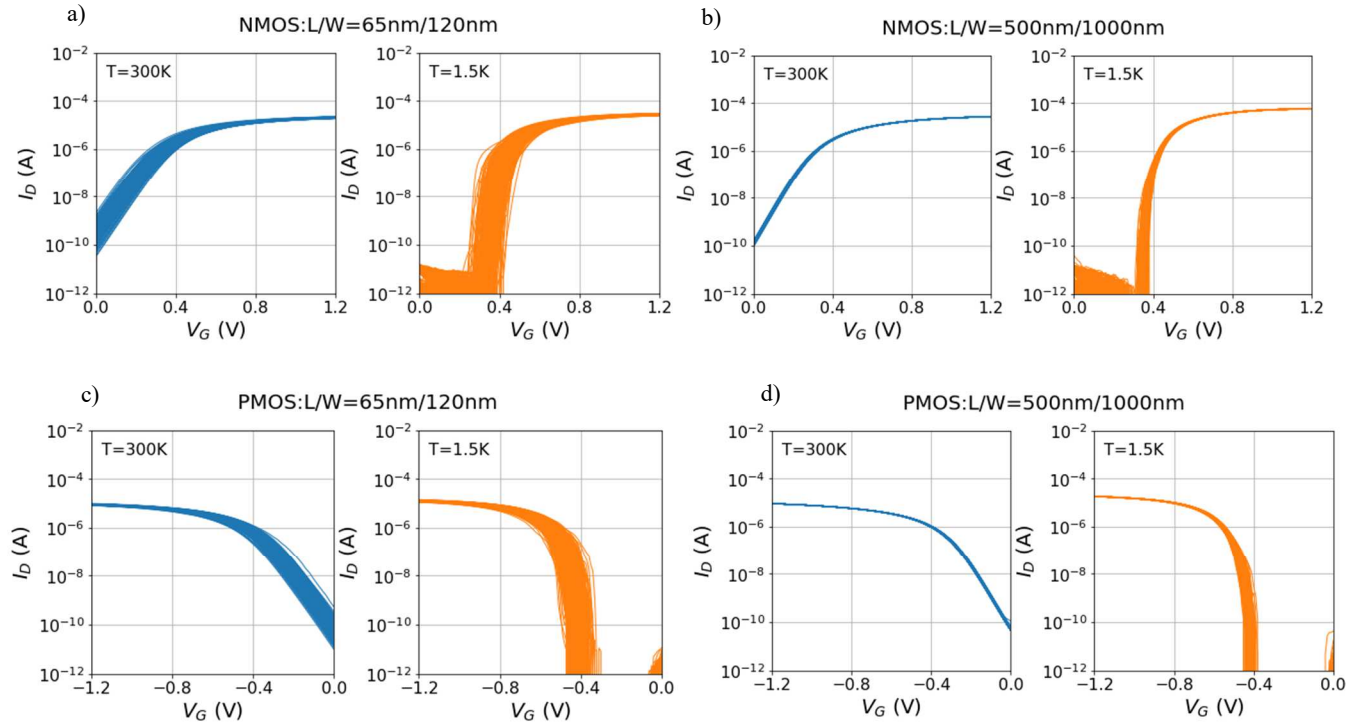


Fig. 4. Id vs. Vg curves measured using test chip at 1.5 K and 300 K. Curves of 256 devices per figure are plotted. Vd=50 mV.

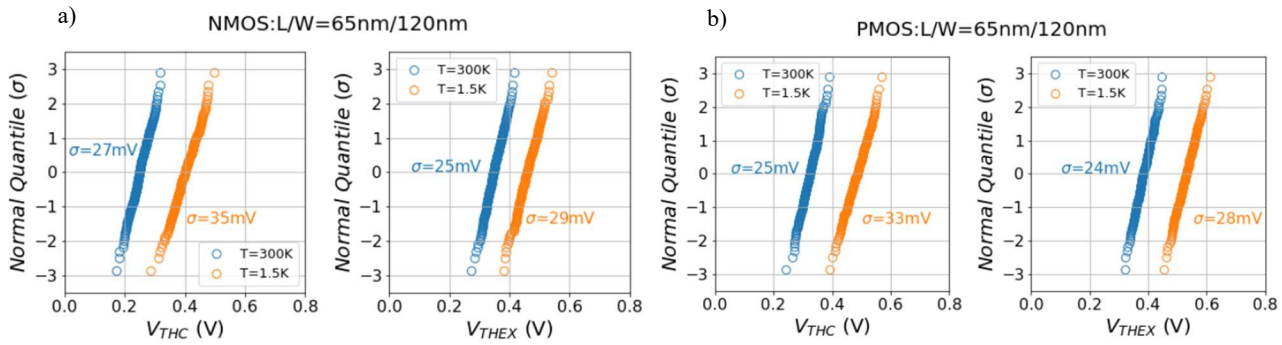


Fig. 5. Normal quantile plots of V_{THC} and V_{THEX} at 1.5 K and 300 K. Vd=50 mV.

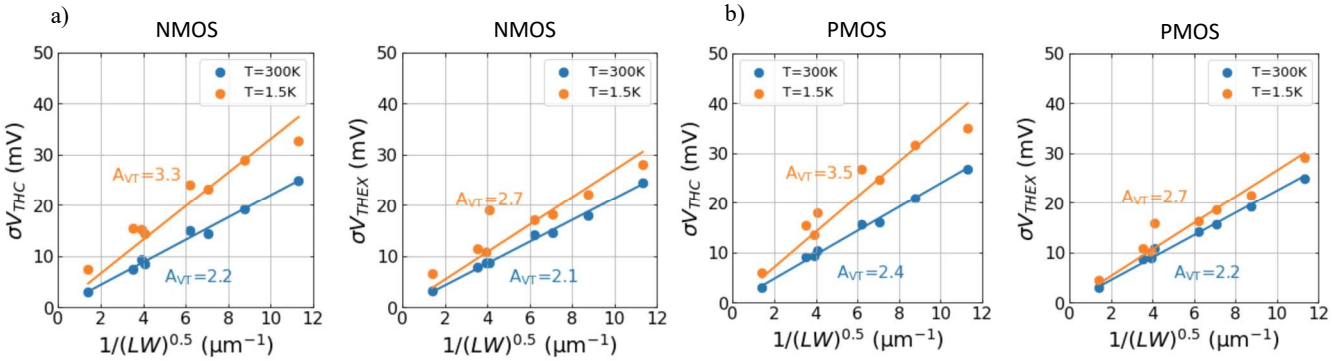


Fig. 6. Standard deviation of V_{THC} and V_{THEX} vs. $1/\sqrt{LW}$ at 1.5 K and 300 K. Vd=50 mV.

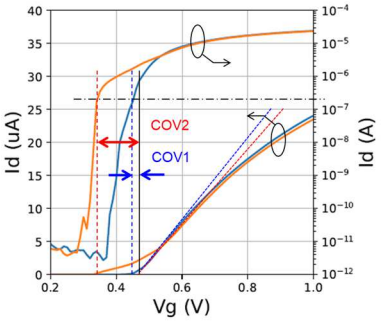


Fig. 7. Examples of COV and its variability at 1.5 K. Vd=50 mV.

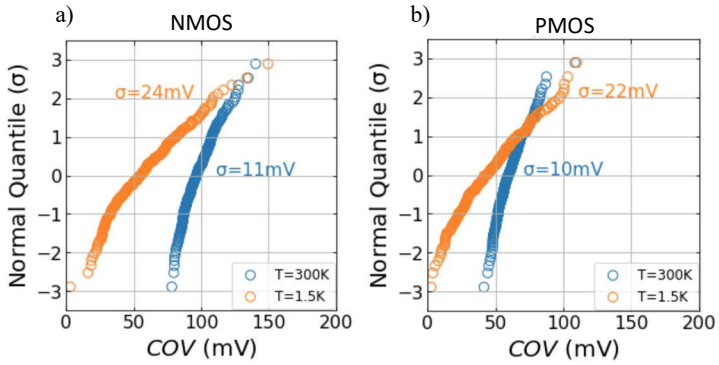


Fig. 8. Normal quantile plots of COV at 1.5 K and 300 K.

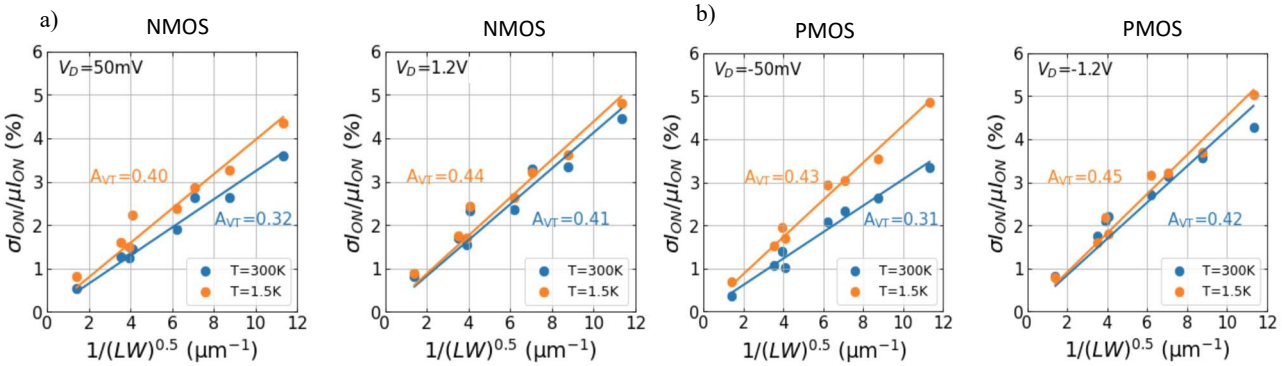


Fig. 9. Standard deviation of linear and saturation drain current vs. $1/\sqrt{LW}$ at 1.5 K and 300 K. Vg=1.2 V.

B. *Variability of low frequency noise*

In addition to static characteristics, the test chip was designed considering its application to measurements of low frequency noise (RTN) and its variability. While RTN measurements using arrays at room temperature are already reported [17-19], in this work, noise was measured at cryogenic temperatures. It is generally believed that RTN is caused by capture and emission of charge carriers by a trap site, and it does not freeze out even near 0 K [20,21]. Since there are only limited number of traps in scaled MOSFETs with small L and W, the number of active traps will statistically differ from device to device. In addition, noise characteristics caused by a trap strongly depend on its position in the device, and the trap site characteristics. As a result, noise waveforms will significantly differ between devices. Fig. 10 shows measured noise signals of three n-channel MOSFETs with the smallest channel size at 1.5 K. Variability of noise is apparent. It is considered that the silent device (Fig. 10a) happens to contain no active trap site, while the noisy device (Fig. 10c) contains many. Typical RTN signal is observed (Fig. 10b) if switching of one trap dominates. These results show that, to understand low frequency noise behavior of Cryo-CMOS, noise measurements of many individual devices (i.e., noise variability measurements) using transistor arrays would be indispensable.

At room temperature, it is reported that RTN with extremely long transition times exists [17]. To see if this is also the case at cryogenic temperatures, a quasi-parallel measurement technique reported in [17] was used, exploiting the common source/gate array (Fig. 1b). By applying a constant Vg to the entire array and reading Id intermittently, one device at a time, the measurement time range was extended from 40 s to 10 h (Fig. 11). Fig. 12 shows RTN waveforms with extremely slow transitions found by this technique. Information obtained by using addressable arrays, including the above, will be useful for understanding and modeling RTN physics at cryogenic temperatures.

IV. Conclusions

Addressable transistor arrays for cryogenic measurements were fabricated, and variability measurements of both static and noise characteristics at 1.5 K were demonstrated. Such test structures would be indispensable for enhancing cryogenic measurement efficiency, and to understand Cryo-CMOS device physics, including that of variability and noise.

Acknowledgment

This paper is based on a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization, Japan.

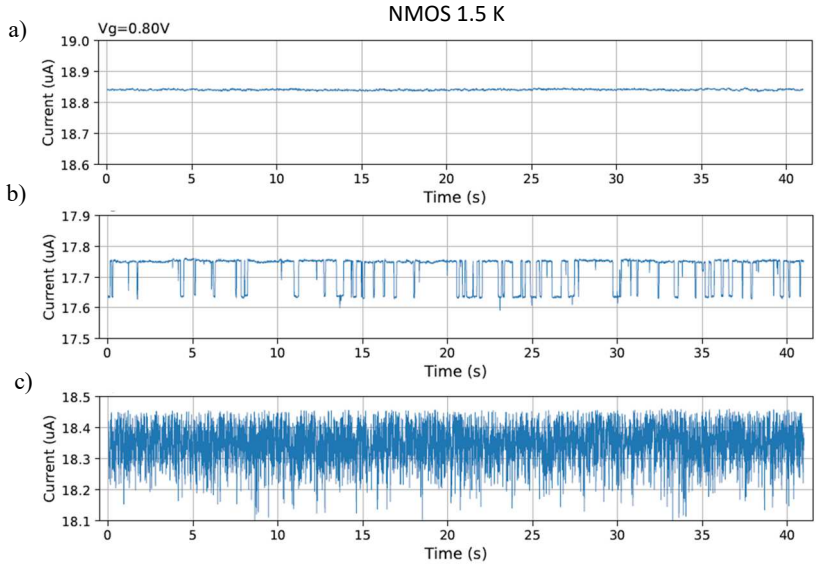


Fig. 10. Standard deviation of linear and saturation drain current vs. $1/\sqrt{LW}$ at 1.5 K and 300 K. Vd=50 mV, Vg=0.8 V.

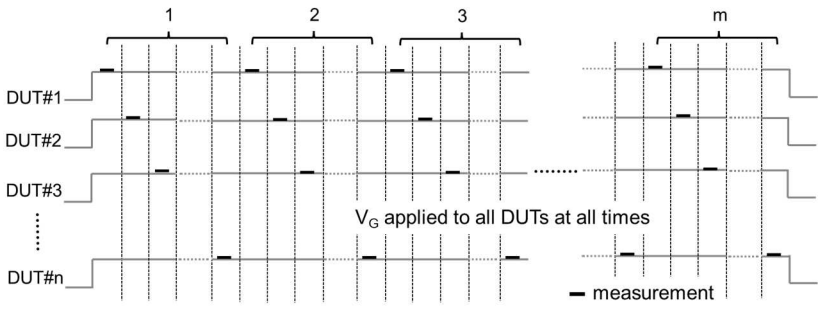


Fig. 11. Quasi-parallel measurement.

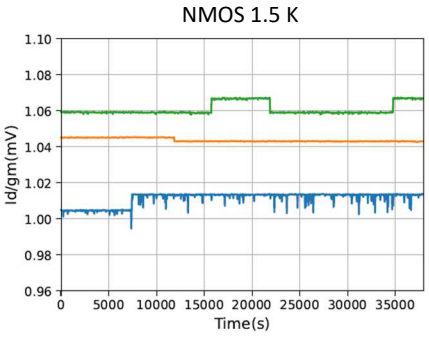


Fig. 12. RTN waveforms with extremely slow transitions.

REFERENCES

[1] S. R. Ekanayake, Torsten Lehmanne, Andrew S. Dzurak, Robert G. Clark, and, Andrew Brawley, "Characterization of SOS-CMOS FETs at LowTemperatures for the Design of Integrated Circuits for Quantum Bit Control and Readout," IEEE Trans. Electron Devices, vol. 57, no. 2, pp. 539-147, Feb. 2010.

[2] E. Charbon, F. Sebastiano, A. Vladimirescu, H. Homulle, S. Visser, L. Song, and R.M. Incandela, "Cryo-CMOS for Quantum Computing," IEEE IEDM, 2016, pp. 343-346.

[3] Joseph C. Bardin, Evan Jeffrey, Erik Lucero, Trent Huang, Ofer Naaman, Rami Barends, Ted White, Marissa Giustina, Daniel Sank, Pedram Roushan, Kunal Arya, Benjamin Chiaro, Julian Kelly, Jimmy Chen, Brian Burkett, Yu Chen, Andrew Dunsworth, Austin Fowler, Brooks Foxen, Craig Gidney, Rob Graff, Paul Klimov, Josh Mutus, Matthew McEwen, Anthony Megrant, Matthew Neeley, Charles Neill, Chris Quintana, Amit Vainsencher, Hartmut Neven, and John Martinis, "A 28nm Bulk-CMOS 4-to-8GHz <2mW Cryogenic Pulse Modulator for Scalable Quantum Computing," IEEE ISSCC, 2019, pp. 456-458.

[4] S. J. Pauka, K. Das, R. Kalra, A. Moini, Y. Yang, M. Trainer, A. Bousquet, C. Cantaloube, N. Dick, G. C. Gardner, M. J. Manfra, and D. J. Reilly, "A cryogenic CMOS chip for generating control signals for multiple qubits," Nature Electronics, vol.4, pp. 64–70, Jan. 2021.

[5] Arnout Beckers, Farzan Jazaeri, and Christian Enz, "Characterization and Modeling of 28-nm Bulk CMOS Technology Down to 4.2 K," IEEE J. Electron Devices Soc., vol. 6, pp. 1007–1018, 2018.

[6] Rosario M. Incandela, Lin Song, Harald Homulle, Edoardo Charbon, Andrei Vladimirescu, and Fabio Sebastiano, "Characterization and Compact Modeling of Nanometer CMOS Transistors at Deep-Cryogenic Temperatures," IEEE J. Electron Devices Soc., vol. 6, pp. 996–1006, 2018.

[7] Yan Liu, Lili Lang , Yongwei Chang , Yi Shan, Xiaojie Chen, and Yemin Dong, "Cryogenic Characteristics of MultinanoscalesField-Effect Transistors," IEEE Trans. Electron Devices, vol. 68, no. 2, pp. 456-463, Feb. 2021.

[8] Hung-Chi Han, Farzan Jazaeri, Antonio D'Amico, Andrea Baschirottoy, Edoardo Charbon, and Christian Enz, "Cryogenic Characterization of 16 nm FinFET Technology for Quantum Computing," IEEE ESSCIRC, 2021, pp. 71-74.

[9] Aravind Appaswamya , Partha Chakrabortya , and John D. Cressler, "Cryogenic Matching Performance of 90 nm MOSFETs," IEEE ISDRS, 2009.

[10] Nguyen Cong Dao, Abdallah El Kass, Craig T. Jin, and Philip H. W. Leong, "Impact of Series Resistance on Bulk CMOS Current Matching Over the 5–300K Temperature Range," IEEE Electron Device Lett., vol. 38, no. 7, pp. 847-850, Jul. 2017.

[11] Tomoko Mizutani, Kiyoshi Takeuchi, Takuya Saraya, Hiroshi Oka, Takahiro Mori, Masaharu Kobayashi, and Toshiro Hiramoto, "Effect of percolation path on temperature dependence of threshold voltage variability in bulk MOSFET," Jpn. J. Appl. Phys., 61, SC1006, 2022.

[12] Kushal Das, Torsten Lehmann, "Effect of deep cryogenic temperature on silicon-on-insulator CMOS mismatch: A circuit designer's perspective," Cryogenics, vol. 62, pp. 84-93, Jul.–Aug. 2014.

[13] Bruna Cardoso Paz , Loïck Le Guevel , Mikaël Cassé , Gérard Billiot, Gaël Pillonnet, Aloysius Jansen, Sebastien Haendler, Andre Juge, Emmanuel Vincent, Philippe Galy, Gérard Ghibaudo, Maud Vinet, Silvano de Franceschi, Tristan Meunier, and Fred Gaillard, "Integrated Variability Measurements of 28 nm FDSOI MOSFETs down to 4.2 K for Cryogenic CMOS Applications," IEEE ICMTS, 2020, 8.3.

[14] A. Grill, E. Bury, J. Michl, S. Tyaginov, D. Linten, T. Grasser, B. Parvais, B. Kaczer, M. Waltl, and I. Radu, "Reliability and Variability of Advanced CMOS Devices at Cryogenic Temperatures," IEEE IRPS, 2020.

[15] P. A. T. Hart, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Subthreshold Mismatch in Nanometer CMOS at Cryogenic Temperatures," IEEE J. Electron Devices Soc. Vol. 8, pp. 797-806, 2020.

[16] Tetsuo Chagawa, Kazuo Terada, Jianyu Xiang, Katsuhiro Tsuji, Takaaki Tsunomura, and Akio Nishida, "Measurement of the MOSFET Drain Curent Variation Unser High Gate Voltage," IEEE ICMTS, 2008, pp. 86-89.

[17] K. Takeuchi, T. Nagumo, S. Yoko gawa, K. Imai and Y. Hayashi, "Single-Charge-Based Modeling of Transistor Characteristics FluctuationsBased on Statistical Measurement of RTN Amplitude," IEEE Symp. VLSI Tech. 2009, pp. 54-55.

[18] Simeon Realov and Kenneth L. Shepard, "Analysis of Random Telegraph Noise in 45-nm CMOS Using On-Chip Characterization System," IEEE Trans. Electron Devices, vol. 60, no. 5, pp. 1716-1722, May. 2013.

[19] Amy Whitcombe, Scott Taylor, Martin Denham, Vladimir Milovanovi´c, and Borivoje Nikoli´c, "On-Chip I-V Variability and Random Telegraph Noise Characterization in 28 nm CMOS," IEEE ESSDERC, 2016, pp. 248-251.

[20] D. H. Cobden and M. J. Uren, "Random telegraph signals from liquid helium to room temperature," Microelectronic Eng., vol. 22, pp. 163-170, 1993.

[21] J. Michl, A. Grill, B. Stampfer, D. Waldhoer, C. Schleich, T. Knobloch, E. Ioannidis, H. Enichlmair, R. Minixhofer, B. Kaczer, B. Parvais, B. Govoreanu, I. Radu, T. Grasser, and M. Walt, "Evidence of Tunneling Driven Random Telegraph Noise in Cryo-CMOS," IEEE IEDM, 2021, pp. 665-668.

Gaussian process-based device model toward a unified current model across room to cryogenic temperatures

Michihiro Shintani[1], Tetsuro Iwasaki[1], and Takashi Sato[2]
[1]Graduate School of Science and Technology, Kyoto Institute of Technology
Matsugasaki, Sakyo-ku, Kyoto, 606-8585 Japan
Email: shintani@kit.ac.jp, tiwasaki@vlsi.es.kit.ac.jp
[2]Graduate School of Informatics, Kyoto University
Yoshida-hon-machi, Sakyo-ku, Kyoto 606-8501, Japan
Email: takashi@i.kyoto-u.ac.jp

***Abstract*—While the typical temperature range supported by standard MOSFET models is from −55 to 125 °C, recent emerging applications require to operate at a wider temperature range, particularly towards lower temperatures. To address these demands, we apply the sparse Gaussian process to build a unified and compact device model for simulating CMOS circuits operating from room to cryogenic temperatures. Unlike neural networks, the sparse Gaussian process prevents overfitting issues. The evaluation of nMOS and pMOS transistors with various dimensions fabricated using a 65 nm node demonstrates that the I-V characteristics are efficiently and accurately modeled from 4 K to 300 K.**

I. Introduction

The use of cryogenic temperatures to improve transistor performance has led to an expansion of integrated circuit applications in recent years [1]. Cryogenic operation, such as at 77 K, has been shown to significantly enhance performance in high-performance computing (HPC) compared to room temperature operation [2,3]. Furthermore, emerging applications require support for operation across wider temperature ranges. The electronics of a satellite designed for deep space must operate within a temperature range of approximately 77 K to 300K [4]. The operating temperature for magnetic resonance imaging (MRI) equipment is as low as 23 K [5].

Among these emerging applications, control circuits for quantum computers deserve significant interest. Quantum computing can potentially offer substantial speed enhancements [6] compared to conventional Neumann architectures in specific areas. At the heart of quantum computers, especially superconducting quantum computers, lies an array of quantum bits (qubits) that necessitate control through classical silicon integrated circuits. All gate operations are carried out by the control systems made of CMOS circuits.

Currently, most control systems are typically designed for room temperature operation, while qubits operate at cryogenic temperatures, usually below 10 mK. The complexity of the cables connecting between room temperature and cryogenic environments in a refrigerator increases as the number of qubits increases, as each qubit requires a connection to its respective control circuit. This creates fundamental challenges in realizing large-scale quantum computers. To improve the scalability of quantum computers, a CMOS circuit system operating at cryogenic temperature is employed for qubit control, mitigating the increase in cabling requirements [7,8].

Cryogenic temperatures are typically outside the supported range of standard compact models. Since the device behavior tends to be different at cryogenic temperatures, designers often need to create their own transistor models that are applicable to their target temperature. While several models have been proposed for specific low and cryogenic temperatures, such as 70 K or 4 K [9]–[11], a unified model capable of spanning cryogenic to room temperatures is still under intensive study. In reference [12], it was suggested that a physics-based model, such as PSP [13] or HiSIM [14], could serve as the basis for a comprehensive model that covers a wide of temperature range, from room temperature to cryogenic temperatures. However, such a model is not yet widely accessible.

This study extends the sparse Gaussian process (SGP)-based model, first introduced in [15] for power device modeling, to concisely model the measured current characteristics of CMOS devices over a wide temperature range. Specifically, we develop a simulation model for transistors of a 65 nm technology node with different dimensions, measured from 3 K up to room temperature. Unlike existing work using artificial neural networks, such as [16], SGP mitigates overfitting problems based on Bayes' theorem, thus facilitating accurate circuit simulations. The evaluation using a commercial SPICE simulator with Verilog-A implementation demonstrates that the compact SGP-based model accurately reproduces the I-V characteristics over a wide temperature range, from 4 K to 300 K.

The remainder of this paper is organized as follows. Section II provides an overview of Gaussian process regression, a fundamental component of SGP-based modeling; Section III describes the SGP-based modeling method; Section IV applies the SGP-based method to the current-voltage (I-V) characteristics of nMOS and pMOS transistors measured from 3 K to 300 K. Finally, Section V concludes the paper.

II. Sparse Gaussian process regression

In this section, we briefly review the Gaussian process (GP) [17], a fundamental component of the SGP-based model [15].

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

The relationship between $\mathbf{X}$ and $\mathbf{y}$ is referred to as a Gaussian process, where the joint probability $p(\mathbf{y})$ of the output $\mathbf{y} = (y_1, y_2, \ldots, y_N)$ corresponding to inputs $\mathbf{X} = (x_1, x_2, \ldots, x_N)$ follows a multivariate Gaussian distribution $N(\boldsymbol{\mu}, \mathbf{K})$. The kernel matrix $\mathbf{K}$ is given by $f_{\mathrm{kern}}(\mathbf{x}_n, \mathbf{x}'_n)$, where f_{kern} is a kernel function, and $\mathbf{x}_n$ and $\mathbf{x}'_n$ are the pair of inputs. As $\mathbf{K}$ is a covariance matrix, $f_{\mathrm{kern}}(\mathbf{x}_n, \mathbf{x}'_n)$ takes a large value when the two inputs and the outputs are correlated, i.e., when close inputs $\mathbf{x}_n$ and $\mathbf{x}'_n$ are given, their outputs y_n and y'_n are also close.

The GP regression problem is to estimate the function $f(\cdot)$ that satisfies $\mathbf{y} = f(\mathbf{X})$, where $\mathbf{X}$ is the input variable and $\mathbf{y}$ is the corresponding output variable. To this end, the training data $(\mathbf{X}_{\mathrm{train}}, \mathbf{y}_{\mathrm{train}}) = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_N, y_N)\}$ comprising input $\mathbf{X}$ and output $\mathbf{y}$ pairs, and f_{kernel} are provided. The kernel function f_{kernel} includes the radial basis function (RBF) kernel, exponential kernel, and polynomial kernel [18]. Among these, the RBF kernel stands out for its versatility and widespread use. This kernel is written as

$$f_{\mathrm{kern}}(\mathbf{x}, \mathbf{x}') = \theta_1 \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^2}{\theta_2}\right), \tag{1}$$

where θ_1 and θ_2 are the fitting parameters. When obtaining y^*_m corresponding to unknown $\mathbf{x}^*_m$, GP computes the function $f(\cdot)$ as a multidimensional Gaussian distribution using the kernel matrix $\mathbf{K}$ as follows:

$$\begin{aligned} &p(y^*_m | \mathbf{x}^*_m, \mathbf{X}_{\mathrm{train}}, \mathbf{y}_{\mathrm{train}}) \\ &\quad = \mathcal{N}(\mathbf{k}_*^T \mathbf{K}^{-1} \mathbf{y}_{\mathrm{train}}, \mathbf{k}_{**} - \mathbf{k}_*^T \mathbf{K}^{-1} \mathbf{k}_*), \end{aligned} \tag{2}$$

where $\mathbf{k}_*$ and $\mathbf{k}_{**}$ are the covariances of the training and test datasets, respectively. According to Eq. (2), the predicted value y^*_m is represented as a probability distribution rather than a fixed value. A narrow concentration of this distribution indicates high confidence in the predicted value, while a wide distribution suggests lower prediction accuracy.

Although GP is a versatile model with broad applicability, it has one significant drawback: the direct implementation of the above equation increases the computational time associated with matrix inversion. Specifically, the prediction requires the inversion of $\mathbf{K}$, as shown in Eq. (2). Inverting a dense matrix takes time that scales cubically with the size of the matrix, i.e., with the training data (N), which hampers computational efficiency.

To address this issue, SGP, a sparse approximation method of GP [19], is employed in [15] to reduce computational time. SGP regression (SGPR) introduces inducing points $\mathbf{Z} = (z_1, z_2, \ldots, z_M)$ to produce a representative subset of all input data, where M is the number of the inducing points ($M \ll N$). The output value $\mathbf{u} = f(\mathbf{Z})$ at the inducing point in SPGR, Eq. (2) is deformed as follows:

$$\begin{aligned} &p(y^*_m | \mathbf{x}^*_m, \mathbf{X}_{\mathrm{train}}, \mathbf{y}_{\mathrm{train}}, \mathbf{Z}) \\ &\quad = \mathcal{N}(\mathbf{k}_{M*}^\top \mathbf{K}_{MM}^{-1} \widehat{\mathbf{u}}, \mathbf{k}_{**} - \mathbf{k}_{M*}^\top \widehat{\boldsymbol{\Sigma}}_{\mathbf{u}}^{-1} \mathbf{k}_{M*}), \end{aligned} \tag{3}$$

where $\widehat{\mathbf{u}}$ and $\widehat{\Sigma}_{\mathbf{u}}$ are the mean and variance of the posterior probability of $\mathbf{u}$, i.e., $p(\mathbf{u})$. Since $\mathbf{K}_{MM}^{-1}$ and other matrices subject to matrix calculation are $M \times M$ matrix, the computation cost of the inverse matrix, which is $O(M^3)$, can be reduced by keeping M small.

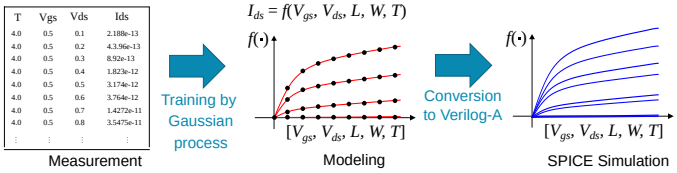


Fig. 1. Overview of the SGP-based modeling.

III. Sparse Gaussian Process-based Device Modeling

We apply the SGP-based device modeling, which was originally proposed for power device modeling in [15], to model the current characteristics of CMOS transistors. This model takes into account temperature characteristics ranging from room temperature to cryogenic temperatures. Fig. 1 illustrates an overview of the SGP-based modeling method. This method utilizes measurement data as the training set and then translates it into a Verilog-A language description [20].

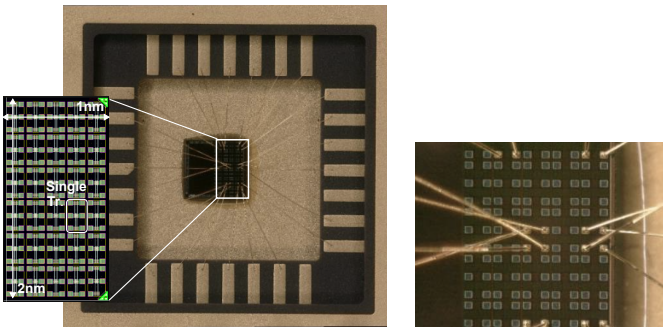
The drain current I_{ds} is represented by the function $f(V_{\mathrm{gs}}, V_{\mathrm{ds}}, L, W, T)$, where V_{gs} is the gate-source voltage, V_{ds} is the drain-source voltage, L is the channel length, W is the channel width, and T is the temperature. We would like to stress that the $f(\cdot)$ is a single function that can express I_{ds} as a function of V_{gs}, V_{ds}, L, W, and T, covering all bias voltages, transistor sizes, and temperatures. This is in contrast to the binning approach. The current model $f(\cdot)$ is incorporated into Verilog-A for SPICE simulations after the training phase. The Verilog-A syntax allows for efficient coding of the inverse algorithm used in SGP by computing the kernel matrix $\mathbf{K}$ using arrays and loops.

The prediction of the drain current by SGPR is performed using Eq. (3), where $\mathbf{X}_{\mathrm{train}}$ contains V_{gs}, V_{ds}, L, W, and T, which are the input to the prediction model, and $\mathbf{y}_{\mathrm{train}}$ corresponds to I_{ds} of each input condition. The inducing points $\mathbf{Z}$ are set equally spaced from each condition according to M. Learning by SGPR involves optimization of the hyperparameters of the kernel function and the inducing points. After optimization, three learning results are obtained: the optimized inducing point Z, the kernel matrix f_{kern}, and the parameter of the kernel function, e.g., θ_1 and θ_2 in Eq. (1), and the predictive distribution of the output value y^* for a new input $\mathbf{x}^*$ is obtained by Eq. (3).

In order to perform SPICE simulation using the predicted model, we construct a Verilog-A model. The SGPR inference algorithm is implemented in Verilog-A to predict drain current at specific bias conditions and temperature. The predicted distribution of y^* is obtained as illustrated in Eq. (3). Here, the average value of the predicted distribution, $\mathbf{k}_{M*}^T \mathbf{K}_{\mathbf{MM}}^{-\mathbf{1}} \widehat{\mathbf{u}}$, is returned as the prediction result. The implementation in Verilog-A incorporates the optimized inducing point $\mathbf{Z}$, the kernel function f_{kern}, and its hyperparameters using the `` `include `` statement. The computations of the inverse matrix are effectively implemented by combining arrays and the `for` statement.

IV. Experiments

To validate the proposed SGP-based unified device model for room to cryogenic temperatures, we performed an evaluation using the measured data of nMOS and pMOS transistors



(a) Layout (b) Bonding

Fig. 2. Fabricated transistors in a DIP package for the measurement in a cryostat.

TABLE I
MEASURED TRANSISTOR SIZES

Gate length L	Gate width W
60 nm	200 nm
60 nm	800 nm
120 nm	200 nm
120 nm	800 nm

fabricated using a 65 nm bulk process node. The generated current model using the measured I-V characteristics by the SGP-based modeling was implemented in Verilog-A language and evaluated using a commercial SPICE simulator [21]. All programs used in this evaluation were implemented in Python, utilizing the Pyro package to construct the SGPR model [22]. The RBF kernel was employed as f_{kern}. Runtimes for the SGP modeling and SPICE simulation were obtained using a server with an AMD Ryzen Threadripper 3990X 2.2 GHz CPU and NVIDIA RTX A4000 GPU.

A. Fabricated transistors and measurement results

Fig. 2 shows the transistor layout that was designed for the validation of the proposed model. Table I summarizes the dimensions of the measured nMOS and pMOS transistors. Measurements were performed for different transistor sizes, each with a combination of two channel lengths L (60 nm and 120 nm) and two channel widths W (200 nm and 800 nm). The fabricated chip is bonded to a DIP package using Au wires. The DIP package is then mounted on a PCB attached at the tip of a dipstick. This dipstick is inserted into a probe station in a cryostat with liquid Helium dewar and a temperature controller.

A semiconductor device analyzer [23] was used to measure currents by sweeping V_{gs} from 0 V to 1.2 V with a 0.1 V interval and V_{ds} from 0 V to 1.2 V with a 0.02 V interval. Fig. 3 shows a representative I_{ds} characteristic of a measured nMOS transistor with L=60 nm and W=200 nm at different temperatures. The I-V characteristics were measured at 56 different temperatures in total, ranging from 3 K to 300 K, as summarized in Fig. 4. The proposed model was trained using the I-V characteristics of 30 temperatures, while the I-V characteristics of other temperatures were used for validation. It is important to note that the training and validation temperatures are exclusive. From Fig. 3, it can be seen that lower

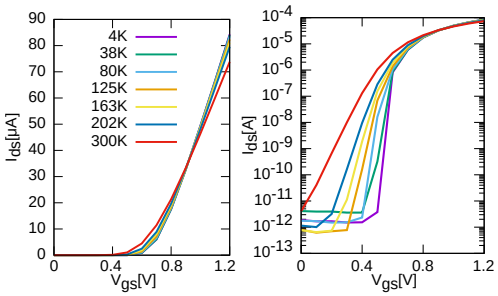


Fig. 3. Measured I_{ds} characteristics of a 65 nm nMOS transistor in linear and log scale (L =60 nm and W =200 nm). All measurements were conducted in a cryostat.

temperatures result in higher threshold voltages and steeper subthreshold slopes due to bandgap widening, intrinsic carrier-density scaling, and incomplete ionization, where the same observation was presented in [9,24]. The resulting temperature-dependent changes in the I-V characteristics are crucial for the modeling.

Based on the above setup, the training data pairs consist of four different transistor sizes, each with 13 V_{gs} conditions, 61 V_{ds} conditions, and 30 temperatures, resulting in a total of 95,160 pairs. Applying the naive GP approach is impractical due to the large kernel matrix size, which is 95,160 × 95,160, which would require more than a week to optimize parameters even with the help of a GPU. Therefore, the use of SGP is crucial to reduce computation time and make the proposed model feasible for practical applications.

B. Modeling results

1) Training: Fig. 5 shows the training and prediction results. For clarity and due to the page limitation, only the results for the nMOS and pMOS transistors with the size of L=60 nm and W=200 nm are presented. Though not included in this manuscript, similar results were obtained for the other transistor sizes and temperatures using a single model.

The proposed unified SGP-based model, simulated on a commercial SPICE simulator, accurately fits for both the training data (open) and the non-training data (solid) I-V curves. The training data points were taken at V_{gs} = 0.0 V to 1.2 V with a 0.1 V interval, excluding data points at V_{gs} = 0.1 V to 1.1 V with a 0.2 V interval. The number of the training data N was $(13 \times 61 \times 30 \times 2 \times 2)$ for V_{gs}, V_{ds}, T, L, and W, while that of the inducing points M was $(13 \times 13 \times 8 \times 2 \times 2)$. For V_{ds}, the 13 points were selected equally spaced from 1 V to 1.2 V. For T, the 8 points were selected equally spaced from a list of the 30 temperatures.

The proposed unified model accurately predicts the full voltage range of I-V characteristics and across a wide temperature range. The model accurately predicts the measured currents for the bias voltages that were not included in the training dataset. However, as depicted in Fig. 5(a), there is a slight discrepancy in the currents at low V_{gs}. This issue arises from the error definition during the model parameter optimization, defined as the absolute difference between the predicted and measured values, which does not consider exponential magnitude differences in current in the subthreshold region. It would be fixed with the change of the definition of loss function.

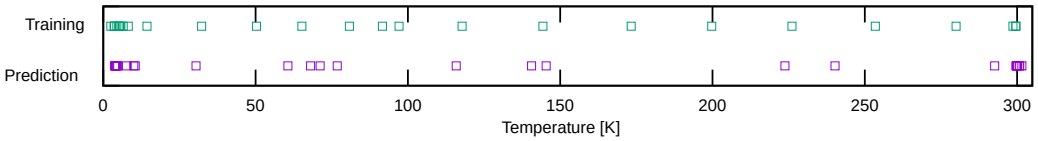
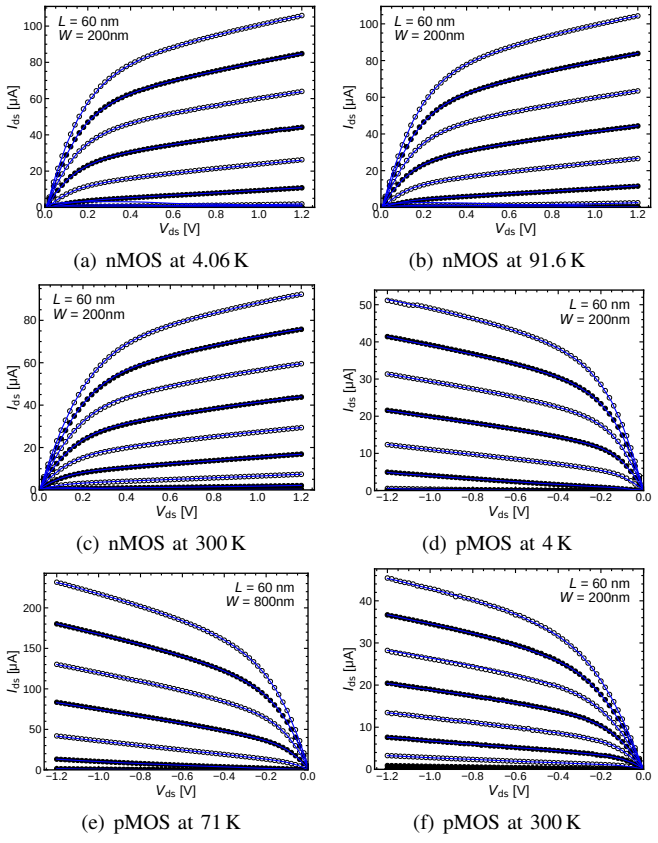
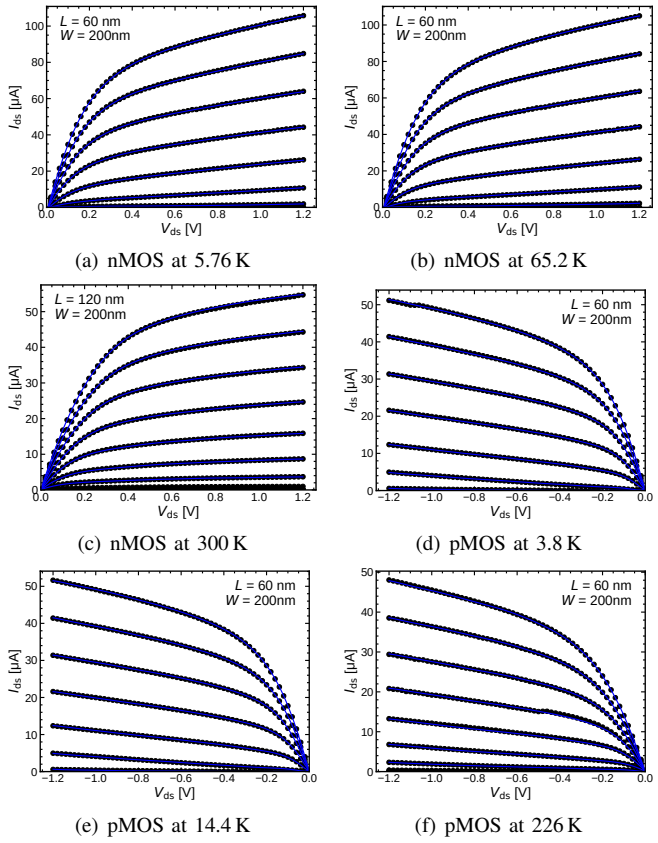


Fig. 4. Points of 56 measured temperatures. The 30 temperatures used for the training and the 26 temperatures for the prediction are marked.



(a) nMOS at 4.06 K (b) nMOS at 91.6 K
(c) nMOS at 300 K (d) pMOS at 4 K
(e) pMOS at 71 K (f) pMOS at 300 K

Fig. 5. Training and prediction results of nMOS and pMOS transistors with L=60 nm and W=200 nm.



(a) nMOS at 5.76 K (b) nMOS at 65.2 K
(c) nMOS at 300 K (d) pMOS at 3.8 K
(e) pMOS at 14.4 K (f) pMOS at 226 K

Fig. 6. Prediction results of the nMOS and pMOS transistors of L=60 nm and W=200 nm for temperatures without I-V measurement data in the training set. The trained model accurately predicts I-V characteristics across various bias voltages and temperatures.

Our SGP-based modeling approach for the nMOS transistors, which encompasses all sizes and temperatures, is accurate and efficient. The generation of the SGP-based model for the nMOS transistors, which covers all sizes and temperatures, took approximately 2 hours and 28 minutes to run. This is a realistic time frame compared to GP, which is difficult to apply in practice. Compared with the physics-based models, the training time can be considered faster, further emphasizing its practicality.

2) Prediction: We then evaluated the accuracy of the predictions for temperatures outside of the training dataset. Fig. 6 shows the predicted I-V curves for the temperatures not present in the training dataset. The results for the nMOS and pMOS transistors with dimensions $L = 60\,\text{nm}$ and $W = 200\,\text{nm}$ are provided as an example. The total SPICE simulation time, including the proposed model evaluations, to generate Fig. 6(a) was 4 seconds. This simulation performed a DC analysis at 793 ($= 13 \times 61$) bias points by sweeping V_{gs} from 0 V to 1.2 V with a 0.1 V step, and V_{ds} from 0 V to 1.2 V with a 0.02 V step. The proposed SGP-model accurately predicted the I-V measurements. The root mean squared errors (RMSEs) for the nMOS transistor in Figs. 6(a), 6(b), and 6(c) are 0.684 µA, 0.096 µA, and 0.105 µA, respectively. The RMSEs for the pMOS transistor in Figs. 6(d), 6(e), and 6(f) are 0.069 µA, 0.050 µA, and 0.116 µA, respectively. The average RMSE for all sizes of the nMOS transistors was 0.523 µA, while that for the pMOS transistors was 0.212 µA. We can conclude that the unified function, created using SGP-based modeling, is capable of efficiently and precisely simulating the drain current characteristics at a wide range of temperatures, including those temperatures that were not included in the training.

V. Conclusion

We applied the SGP-based compact modeling approach, originally proposed in [15], to the simulation of I-V characteristics of 65 nm nMOS and pMOS transistors measured from 3 K to 300 K. Our evaluations confirmed the accuracy of the SGP-based model in reproducing and predicting the I-V characteristics across a wide temperature range. This modeling framework allows for the creation of accurate and flexible compact models, even when the physical mechanism of the target device is not fully understood. However, the usage of our SPICE model is still partially limited due to

a discrepancy shown in Fig. 5(a). We intend to address this issue while maintaining the model's efficiency to establish a more comprehensive current model.

ACKNOWLEDGMENT

This work was supported by JST Moonshot R&D Grant Number JPMJMS226A. The authors also acknowledge support from VDEC, the University of Tokyo, in collaboration with Synopsys, Inc.

REFERENCES

[1] W. F. Clark, B. El-Kareh, R. Pires, S. Titcomb, and R. L. Anderson, "Low temperature CMOS—a brief review," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. 15, no. 3, pp. 397–404, 1992.

[2] I. Byun, D. Min, G.-H. Lee, S. Na, and J. Kim, "CryoCore: A fast and dense processor architecture for cryogenic computing," in *Proceedings of ACM/IEEE Annual International Symposium on Computer Architecture*, 2020.

[3] ——, "A next-generation cryogenic processor architecture," *Proceedings of IEEE/ACM International Symposium on Microarchitecture*, vol. 41, no. 3, pp. 80–86, 2021.

[4] R. Patterson, A. Hammoud, J. Dickman, S. Gerber, M. Elbuluk, and E. Overton, "Electronics for deep space cryogenic applications," in *Proceedings of IEEE Workshop on Low Temperature Electronics*, 2002.

[5] T. C. Cosmus and M. Parizh, "Advances in whole-body MRI magnets," *IEEE Transactions on Applied Superconductivity*, vol. 21, no. 3, pp. 2104–2109, 2010.

[6] R. P. Feynman, "Quantum mechanical computers," *Foundations of Physics*, vol. 16, pp. 11–20, 1985.

[7] B. Patra, R. M. Incandela, J. P. G. van Dijk, H. A. R. Homulle, L. Song, M. Shahmohammadi, R. B. Staszewski, A. Vladimirescu, M. Babaie, F. Sebastiano, and E. Charbon, "Cryo-CMOS circuits and systems for quantum computing applications," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 309–321, 2018.

[8] J. P. G. V. Dijk, B. Patra, S. Subramanian, X. Xue, N. Samkharadze, A. Corna, C. Jeon, F. Sheikh, E. Juarez-Hernandez, B. P. Esparza, H. Rampurawala, B. R. Carlton, S. Ravikumar, C. Nieva, S. Kim, H.-J. Lee, A. Sammak, G. Scappucci, M. Veldhorst, L. M. K. Vandersypen, E. Charbon, S. Pellerano, M. Babaie, and F. Sebastiano, "A scalable Cryo-CMOS controller for the wideband frequency-multiplexed control of spin qubits and transmons," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 15, pp. 2930–2946, 2020.

[9] R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and compact modeling of nanometer CMOS transistors at deep-cryogenic temperatures," vol. 6, pp. 996–1006, 2018.

[10] C. Enz, A. Beckers, and F. Jazaeri, "Cryo-CMOS compact modeling," in *IEEE International Electron Devices Meeting Technical Digest*, 2020, pp. 25.3.1–25.3.4.

[11] E. Schriek, F. Sebastiano, and E. Charbon, "A Cryo-CMOS digital cell library for quantum computing applications," *IEEE Solid-State Circuits Letters*, vol. 3, pp. 310–313, 2020.

[12] M. B. Yelten, "Holistic device modeling: Toward a unified MOSFET model including variability, aging, and extreme operating conditions," *IEEE Transactions on Circuits and Systems-II*, vol. 69, no. 6, pp. 2635–2640, 2022.

[13] G. Gildenblat, X. Li, W. Wu, H. Wang, A. Jha, R. van Langevelde, G. D. J. Smit, A. J. Scholten, and D. B. M. Klaassen, "PSP: An advanced surface-potential-based MOSFET model for circuit simulation," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 1979–1993, 2006.

[14] M. Miura-Mattausch, N. Sadachika, D. Navarro, G. Suzuki, Y. Takeda, M. Miyake, T. Warabino, Y. Mizukane, R. Inagaki, T. Ezaki, H. J. Mattausch, T. Ohguro, T. Iizuka, M. Taguchi, S. Kumashiro, and S. Miyamoto, "HiSIM2: Advanced MOSFET model valid for RF circuit simulation," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 1994–2007, 2006.

[15] K. Shimozato and T. Sato, "dGPLVM: A nonparametric device model for statistical circuit simulation," in *Proceedings of IEEE International Conference on Microelectronic Test Structures*, 2022.

[16] P. Hart, J. van Staveren, F. Sebastiano, J. Xu, D. E. Root, and M. Babaie, "Artificial neural network modelling for Cryo-CMOS devices," in *Proceedings of IEEE Workshop on Low Temperature Electronics*, 2021.

[17] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006. [Online]. Available: http://gaussianprocess.org/gpml/chapters/RW.pdf

[18] D. Duvenaud, "The kernel cookbook." [Online]. Available: https://www.cs.toronto.edu/~duvenaud/cookbook/

[19] M. K. Titsias, "Variational learning of inducing variables in sparse Gaussian processes," in *Proceedings of International Conference on Artificial Intelligence and Statistics*, 2009.

[20] C. C. McAndrew, G. J. Coram, K. K. Gullapalli, J. R. Jones, L. W. Nagel, A. S. Roy, J. Roychowdhury, A. J. Scholten, G. D. J. Smit, X. Wang, and S. Yoshitomi, "Best practices for compact modeling in Verilog-A," *IEEE Journal of the Electron Devices Society*, vol. 3, no. 5, pp. 383–396, 2015.

[21] *HSPICE User Guide: Basic Simulation and Analysis Version P-2019.06*, Synopsys, Inc., 2019.

[22] E. Bingham *et al.*, "Pyro: Deep universal probabilistic programming," *Journal of Machine Learning Research*, vol. 20, pp. 28:1–28:6, 2019.

[23] Keysight Technologies, Inc., *B1500A Semiconductor Device Analyzer*, 2023.

[24] A. Beckers, F. Jazaeri, and C. Enz, "Cryogenic MOSFET threshold voltage model," in *Proceedings of European Solid-State Device Research Conference*, 2020.

A Testbed for Cryogenic On-wafer Noise Measurement Using Cold Source Method with Temperature-Dependent Loss Correction

Guo-Wei Huang, Bo-Yuan Chen, Yu-Shao Shiao, Chia-Wei Chuang, Liang-Chung Shen, Kun-Ming Chen and Chia-Sung Chiu
Taiwan Semiconductor Research Institute
No. 26 Prosperity Road I, Hsinchu Science Park, Hsinchu 300091, Taiwan, R.O.C.
gwhuang@narlabs.org.tw, Tel: 886-3-5773693, Fax: 886-3-5713403

***Abstract*—On-wafer noise measurement at cryogenic temperature is challenging due to complexity of temperature gradient distributions in cryostats. As Keysight Technologies introduce their on-wafer cold source method for stable noise temperature measurement at room temperature, we propose our correct formulas as loss correction to exploit the method to work properly for cryogenic noise temperature measurement, especially as a good fit for those which involve a large number of samples, e.g., Known-Good-Die testing. In this paper, we demonstrate the feasibility of the correction with 16nm, 65nm CMOS, and two InP pHEMT transistors at 4-14GHz. The correction significantly compensates the underestimated noise temperatures by more than 20K. As the results, we suggest that the correction is convenient and more accurate while using the cold source method for cryogenic on-wafer noise temperature measurement.**

Keywords—cold source method, cryogenic noise temperature measurement, cryogenic CMOS, cryogenic InP pHEMT.

I. INTRODUCTION

Cryogenic noise measurement is critical for cryogenic amplifiers in radio astronomy [1] and superconductor physics. However, cryogenic on-wafer noise measurement at ultra-low temperatures is practically difficult. Traditional Y-factor methods, such as the hot/cold load method, cold-attenuator method, variable-temperature load method, become problematic in wafer testing for a large number of samples [2][3].

First, commercial noise sources must function and calibrate at room temperature. Second, temperature gradients of chambers and cables are difficult to determine. On-wafer variable-temperature load and cold-attenuator methods may not be available for more than a single sample, because wafer to wafer interconnection is inevitably troublesome. The cold source method with a termination resistor as the cold thermal source at cryogenic temperature requires S-parameter measurement that requires cryogenic switches [4]. Therefore on-wafer noise measurement for several samples at cryogenic temperatures without thermal cycles seems impractical.

Keysight Technologies had announced their enhanced cold source method at room temperature in an application note for years [5]. Therefore, instead of switching to a cryogenic cold source in [4], we proposed to calculate the cold source noise temperature and compensate the noise temperature measurement with loss correction formulas, and carry out a series of measurement as validation of the proposed correction.

II. COLD SOURCE METHOD

The cold source method generally requires a vector network analyzer (VNA) for precise gain measurement of the device under test (DUT), in addition to a low noise receiver to be calibrated initially using a noise source. To measure the noise temperature of a DUT, T_{DUT}, with a given noise source as an input, T_{IN}, one needs the output noise temperature detected by the noise receiver, T_{RX}, and the gain of the DUT, G_{DUT}, determined by the VNA. T_{DUT} can be obtained as

$$T_{DUT} = \frac{T_{RX}}{G_{DUT}} - T_{IN}. \quad (1)$$

Here T_{IN} is usually the room temperature, 293K. All noise temperature power unit is Kelvin (K) while G_{DUT} is unit-less. The noise figure of the DUT therefore is

$$NF_{DUT} = 10\log(1+\frac{T_{DUT}}{290}), \quad (2)$$

where 290 is the 290K reference temperature.

The enhanced cold source method [5] proposed by Keysight Technologies, exploits their newly VNA model PNA-X series (hereafter PNA) with built-in low noise receivers. The method minimizes the impact of mismatch error using vector-error correction. Noise-parameter effects of both inputs of the DUT and the low noise receiver in a PNA cause noise power measurement mismatch uncertainty and error, which therefore need to be corrected by vector noise calibration in Keysight's procedure as shown in Fig. 1.

In this paper, the pictures of our cryogenic on-wafer noise measurement setup are shown in Fig. 2. It includes a Keysight N5242B PNA-X network analyzer, a N4691A ECal module, a 346A noise source with 5.6dB ENR, and a Lake Shore CRX-4K probe station. The noise receiver in the PNA is calibrated from 4 to 14 GHz with hot/cold temperatures as 1343K and 293K.

Fig. 3 illustrates the system block diagram of the setup. L1 and L4 are blocks that include low loss coaxial cables and connectors between the PNA and the cryogenic chamber. L2 and L3 include the short cables and probes in the cryogenic chamber at both ends of the DUT. L1 and L2 are on the input signal path, while L3 and L4 are on the output. Their loss factors are noted as $L1$, $L2$, $L3$ and $L4$, respectively. Loss

factors are equal or greater than 1. These temperature and frequency dependent loss factors are measured with a standard cable loss method provided by Keysight at room temperature and cryogenic temperature. After the vector noise calibration, the DUT gain and detected output noise power reference plane by the PNA is at position A in Fig. 3.

At room temperatures, DUT noise temperature is shown as

$$T_{DUT} = \frac{T_{RX,A}}{G_{DUT}} - T_{293K}. \quad (3)$$

T_{IN} in Eqn. (1) is now the 293K room temperature, T_{293K}. The PNA port and cables work as a 293K thermal noise source. Since the calibrated reference plane is at position A, we assume that the real detected noise power by the noise receiver in the PNA at position B is

$$T_{RX,B} = \frac{T_{RX,A}}{L3 \cdot L4}. \quad (4)$$

In other words, the losses of L3 and L4 have been compensated when PNA software calculates the noise temperature of the DUT. As the result, $T_{RX,A}$ is the true output noise power in the noise temperature measurement.

III. Cryogenic Loss Correction

As the chamber is cooled down to cryogenic temperature, the T_{IN} is no longer the room temperature, 293K, due to the loss of the input cold cable block L2. The output noise power, $T_{RX,A}$, has also changed due to the effect of the cold output block, L3. Since the vector noise calibration procedures is performed at room temperature, therefore we propose a method of correction formulas for the given cold source method to conveniently resolve the problem.

To accurately calculate the thermal power, it is critical choosing proper functions for the temperature gradients. However without knowledge of the gradients of the block L2 and L3, here they are assumed to be linear in this paper. The thermal noise power contribution of the cable blocks is estimated as other previously published methods [1][3][7]. At cryogenic temperature, the thermal noise contribution of a cable block is as

$$T_L = \frac{(T_1 + T_2)}{2} \cdot (1 - \frac{1}{L}), \quad (5)$$

where T_1 and T_2 are the physical temperatures of the cable block ends. L is the cable loss factor. Therefore the new input cold source noise temperature is

$$T'_{IN} = \frac{T_{293K}}{L2} + T_{L2}, \quad (6)$$

where

$$T_{L2} = \frac{T_{293K} + T_{4K}}{2}(1 - \frac{1}{L2}). \quad (7)$$

The first term on the right hand side of Eqn. (6) is the attenuated T_{293K} and the second term is the thermal noise contribution from the block L2. The block L1 suffers no temperature gradient and works with the PNA port termination as a 293K thermal noise source.

The corrected output noise temperature should be

$$T'_{RX,A} = T_{RX,A} - \frac{T_{293K} + T_{4K}}{2}(L3 - 1), \quad (8)$$

where the second term on the right hand side is the thermal noise power from the block L3. $T_{RX,A}$ can be obtained by reversing the detected noise temperature by the PNA as Eqn. (3), however, which incorrectly includes the thermal noise power contribution of L3 and assumes the input source noise temperature is room temperature. As the noise measurement at room temperature mentioned before, as Eqn (4), the calibrated noise power is presumably the uncalibrated power multiplied by $L3 \cdot L4$ by the PNA software. Therefore

$$T_{RX,A} = T_{RX,B} \cdot L3 \cdot L4 = T'_{RX,A} + T_{L3} \cdot L3 \quad (9)$$

where the thermal noise of L3 is

$$T_{L3} = \frac{(T_{293K} + T_{4K})}{2} \cdot (1 - \frac{1}{L3}), \quad (10)$$

and the uncalibrated noise power is

$$T_{RX,B} = \frac{T'_{RX,A}}{L3 \cdot L4} + \frac{T_{L3}}{L4}. \quad (11)$$

It is worth mentioning that T_{L3} is not attenuated by its own loss, $L3$, while the PNA software compensates anyway when calculating noise figure as at room temperature. Thus the thermal noise power from L3 added at between L3 and L4 should be re-corrected and multiplied by $L3$ as Eqn (9). Finally T_{DUT} can be obtained by Eqn. (12) with T'_{IN} and $T'_{RX,A}$, as below

$$T_{DUT} = \frac{T'_{RX,A}}{G_{DUT}} - T'_{IN}. \quad (12)$$

IV. Measured Results with Loss Correction

First we measure an attenuator at room temperature. The detected output noise power is about 293K as room temperature. The attenuator equivalent noise temperature can be calculated from its gain. However, when the attenuator is cooled down to cryogenic temperature, in addition to cold cable losses, its output noise temperature is much less than room temperature. For example, a 10dB attenuator only let 29.3K noise power pass through to the noise receiver. Block L3 and L4 become to dominate the output noise temperature. As shown in Fig. 4, the noise figures are no longer the same as the attenuation at 4K and 77K. The PNA software miscalculated the attenuator noise temperature.

Four transistors from two different CMOS RF and one InP pHEMT technologies were measured following the on-wafer two-port calibration in the same cooldown cycle. Using our loss correction method based on Eqn. (6) and (8), Fig. 5 and Fig. 6 show the results of 65nm and 16nm CMOS transistors. Modern CMOS transistors are known for their mediate noise temperatures. The amount of the noise temperature correction is greater in the frequency regions due to lower gains and higher cable losses. The amount is about 20K to 60K. When we examine the values from Eqn. (6) and (8), block L2 and L3 both comparably effect the noise temperature, while L2 slightly causes a larger impact on T'_{IN} by decreasing tens of K.

InP pHEMT transistors are known as very low noise devices. Fig. 7 and Fig. 8 show the results of a 2×25μm and 4×62.5μm transistors. Although they already show very low noise temperatures at room temperature, the amount of correction at cryogenic temperature is similar with the CMOS

transistors. The block L2 and L3 loss factors and temperature gradients are very steady as functions of the room and cryogenic temperatures. However assuming linear profiles of temperature gradients for the probes and cables in the cryogenic chamber may be over-simplified and cause other uncertainty. Moreover the input and output impedances of these transistors are unlikely 50Ω, so that the results may not reflect their minimum of noise temperatures.

V. Conclusion

Cryogenic on-wafer noise measurement has been challenging for scientists and engineers for a very long time. Using the Keysight enhanced cold source method with the proposed loss correction as an experimental expedient, we had presented noise temperature measurement results of various devices at cryogenic temperature. It seems promising that a large number on-wafer noise measurement at cryogenic temperatures can be carried out with the correction formulas. However, the cable temperature gradient profiles for proposed loss correction and loss uncertainty may still cause some deviations of DUT noise temperatures. The further calculation of cold cable thermal noise power contribution is still under development.

Acknowledgment

We specially thank Prof. Robert Hu for insightful advice and Kenny Liao with Keysight Technologies for technical support. This work was supported in part by the "Center for Advanced Semiconductor Technology Research" from the Featured Areas Research Center Program within the Framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan; and in part by the National Chung-Shan Institute of Science and Technology, Taiwan, under Grant NCSIST-702-V204(113).

References

[1] R. Hu, "Analysis of the input noise contribution in the noise temperature measurements," in *IEEE Microwave and Wireless Components Letters*, vol. 15, no. 3, pp. 141-143, March 2005.

[2] F. Heinz, F. Thome, A. Leuther and O. Ambacher, "A Cryogenic On-Chip Noise Measurement Procedure With ±1.4-K Measurement Uncertainty," *2022 IEEE/MTT-S International Microwave Symposium*, Denver, CO, USA, 2022, pp. 233-236.

[3] D. Bruch et al., "A single chip broadband noise source for noise measurements at cryogenic temperatures," 2011 *IEEE MTT-S International Microwave Symposium*, Baltimore, MD, USA, 2011, pp. 1-4.

[4] A. Sheldon, L. Belostotski, H. Mani, C. E. Groppi, and K. F. Warnick, "Cryogenic Noise-Parameter Measurements", *IEEE Microwave Magazine*, pp 52-64, August 2021.

[5] Keysight Technologies, "High-Accuracy Noise Figure Measurements Using the PNA-X Series Network Analyzer", Application note 1408-20.

[6] Lake Shore Cryotronics, Inc., "Microwave Measurement with Lake Shore Probe Stations", June 2007.

[7] J. E. Fernandez, "A noise-temperature measurement system using a cryogenic attenuator," JPL TMO, Pasadena, CA, USA, Prog. Rep. 42-135, Nov. 1998.

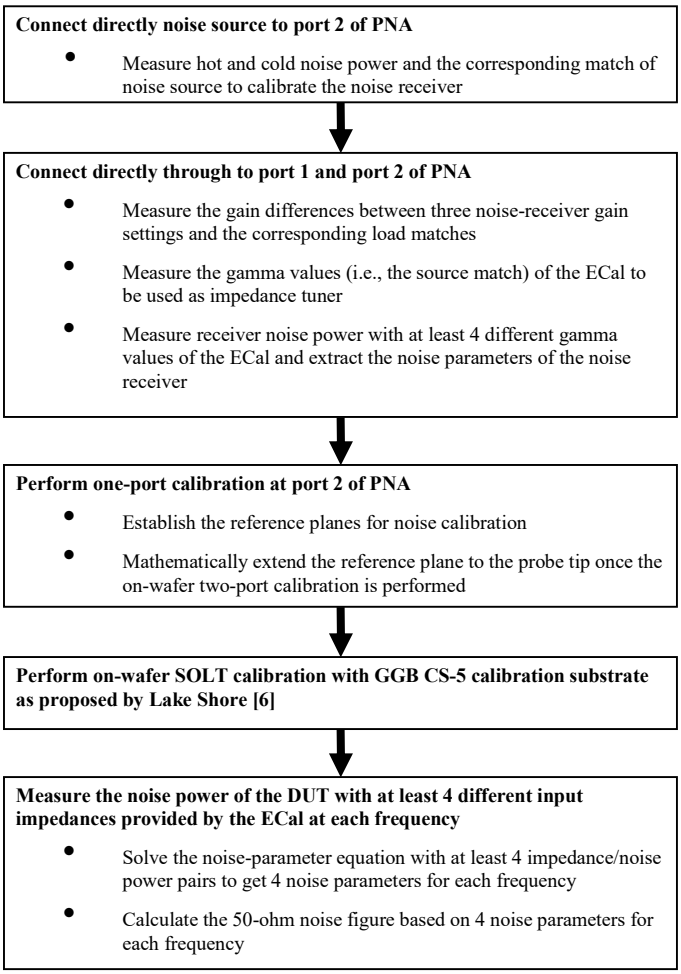


Fig. 1. Flowchart of vector noise calibration and measurement procedure

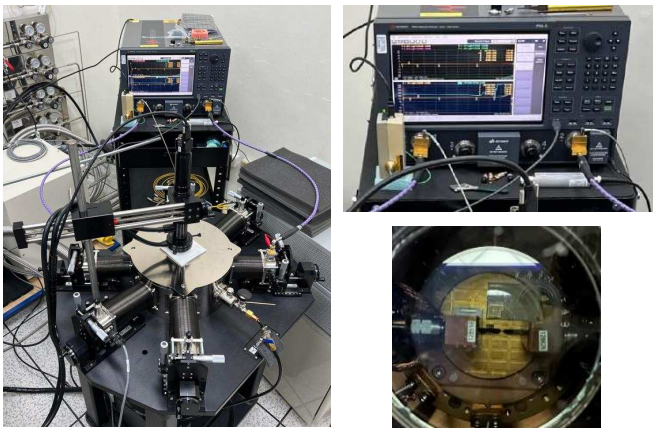


Fig. 2. Pictures of our on-wafer cryogenic noise measurement setup. The upper right is the PNA and ECal module. The lower right is the probes and holder.

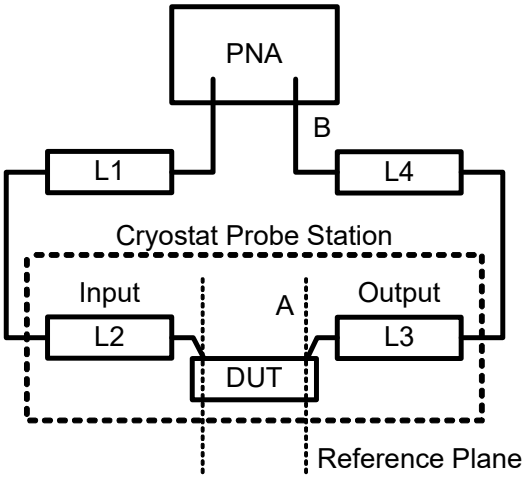


Fig. 3. System block diagram of the noise measurement setup. When the probe station is at 4K temperature, block L1 and L4 are cables and connectors at room temperature while L2 and L3 are with temperature gradients from 293K to 4K. Wires between the blocks are lossless.

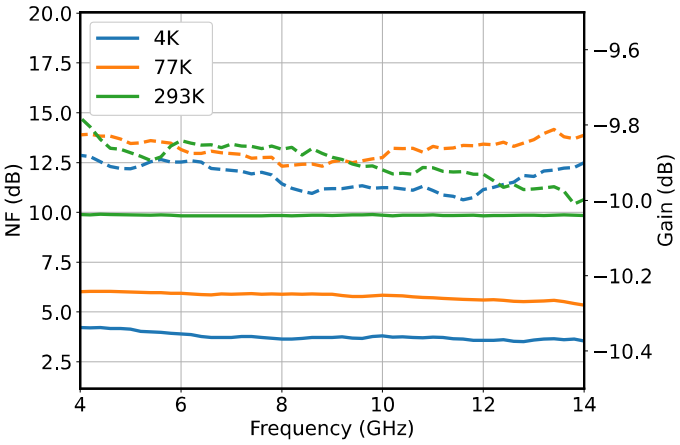


Fig. 4. 10dB attenuator gains and noise figures at 4K, 77K and 293K. The dashed lines are gains while solid lines are noise figures.

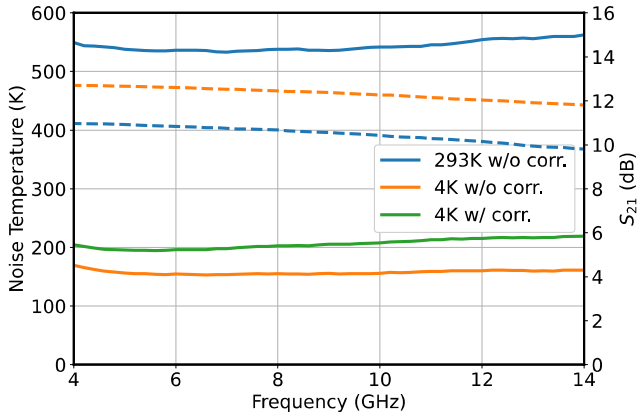


Fig. 5. 65nm CMOS transistor gains and noise temperatures at 4K and 293K. The dashed lines are gains while solid lines are noise temperatures. The green solid line shows noise temperatures after correction.

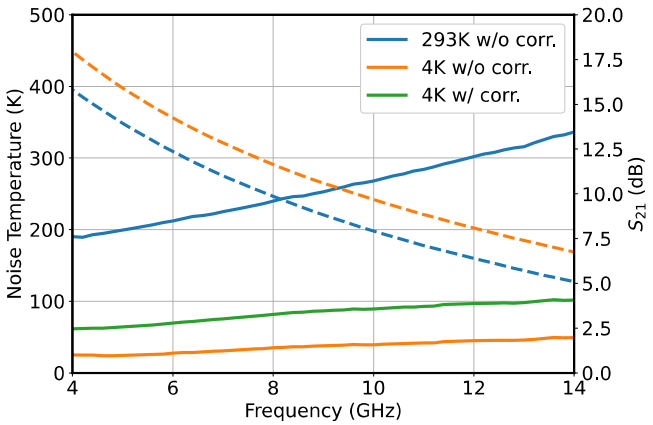


Fig. 6. 16nm CMOS transistor gains and noise temperatures at 4K and 293K. The dashed lines are gains while solid lines are noise temperatures. The green solid line shows noise temperatures after correction.

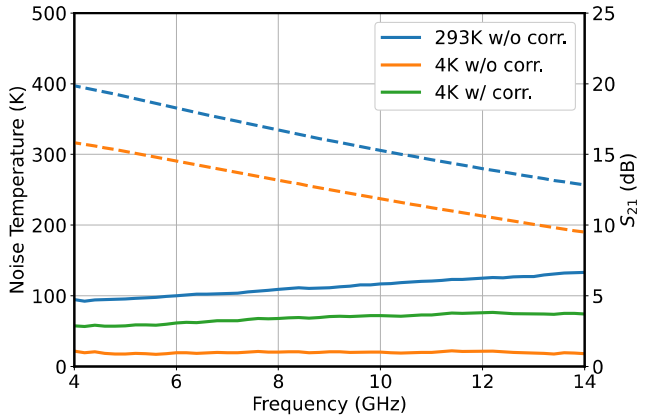


Fig. 7. 2×25µm InP pHEMT transistor gains and noise temperatures at 4K and 293K. The dashed lines are gains while solid lines are noise temperatures. The green solid line shows corrected noise temperatures.

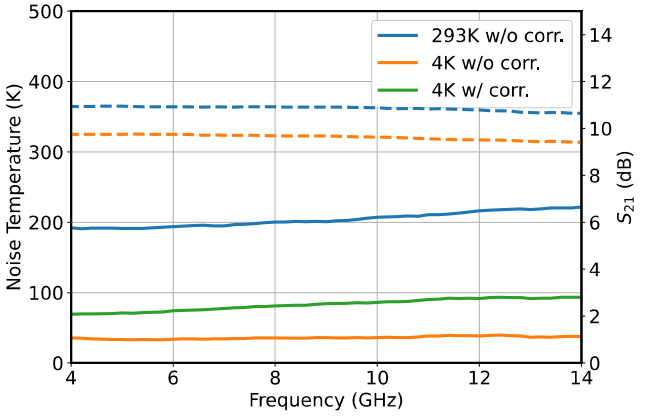
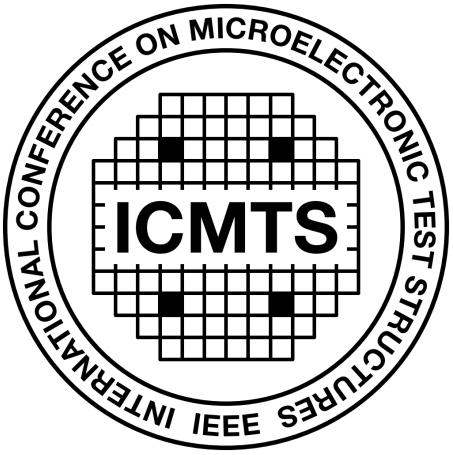


Fig. 8. 4×62.5µm InP pHEMT transistor gains and noise temperatures at 4K and 293K. The dashed lines are gains while solid lines are noise temperatures. The green solid line shows corrected noise temperatures

SESSION 4
Dielectrics and Ferroelectrics



16th April 2024, 16:30–17:30

Session Co-Chairs: Kjell Jeppson *Chalmers University of Technology, Sweden*
Chadwin Young, *University of Texas at Dallas, USA*

Analysis and Compensation of the Series Resistance Effects on the Characteristics of Ferroelectric Capacitors

Marco Massarotto[1], Francesco Driussi[1], Mattia Bucovaz[1], Antonio Affanni[1], Suzanne Lancaster[2], Stefan Slesazeck[2], Thomas Mikolajick[2,3] and David Esseni[1]
[1]DPIA, University of Udine, Udine, Italy; Email: massarotto.marco001@spes.uniud.it
[2]NaMLab gGmbH, Dresden, Germany
[3]Chair of Nanoelectronics, IHM, TU–Dresden, Germany

***Abstract*—The energy efficiency of ferroelectric-based devices makes them interesting for many applications. However, their optimization requires a dependable characterization of the ferroelectric (FE) material. In this work, we show and investigate how the series resistance (R_S) can strongly impact the current-voltage (I-V) characteristics of Metal-Ferroelectric-Metal (MFM) stacks and distorts the hysteresis curves, which can lead to an inaccurate extraction of the FE parameters and a misleading interpretation of FE switching dynamics. The complex R_S effect on the I-V curves cannot be easily compensated, so here we propose, for the first time to our knowledge, a procedure for an improved extraction of the FE parameters even in the presence of a non-negligible series resistance.**

I. Introduction

$HfZrO_4$ (HZO) is a CMOS-compatible, highly scalable ferroelectric (FE) that is raising interest for many applications ranging from non-volatile memories to neuromorphic computing [1]. However, the FE device optimization requires an in-depth HZO characterization and a dependable extraction of its remnant polarization (P_R) and coercive voltage (V_C).

In this work, we report an extensive experimental study of Metal-Ferroelectric-Metal (MFM) stacks carried out at different frequencies of the applied signals and for samples with different areas. Our measurements evidenced features and distortions that we ascribe to the presence of a non-negligible series resistance (R_S). After verifying this assumption, we analyzed the impact of such extrinsic R_S on the extraction of the main HZO parameters, namely the P_R and V_C values. Finally, we quantitatively assess the influence of R_S on the MFM characteristics and propose a new compensation procedure for a dependable extraction of P_R and V_C even in the presence of significant R_S values.

II. Devices and Experimental Setup

Metal-Ferroelectric-Metal (MFM) stacks are fabricated by ALD of 10.5 nm HZO onto W (30 nm)/TiAlN (22 nm) bottom electrode (BE, Fig. 1). PVD is used for the 22 nm TiAlN top contacts. Circular areas with diameter Φ ranging from 110 to 450 μm are defined by shadow mask evaporation of Ti/Pt and etching. The BE is reached through a lateral broken MFM device (BD, Fig. 1).

To measure the samples, the previously verified setup in Fig. 1 is used [2]. FE polarization P is switched by a triangular voltage wave (V_{IN}, Fig. 2a) and studied by measuring the switching current I_{MFM} during time through an I–V converter (Fig. 1). Then, the integral over time of I_{MFM} provides the switching charge, which is typically interpreted as the polarization P [3].

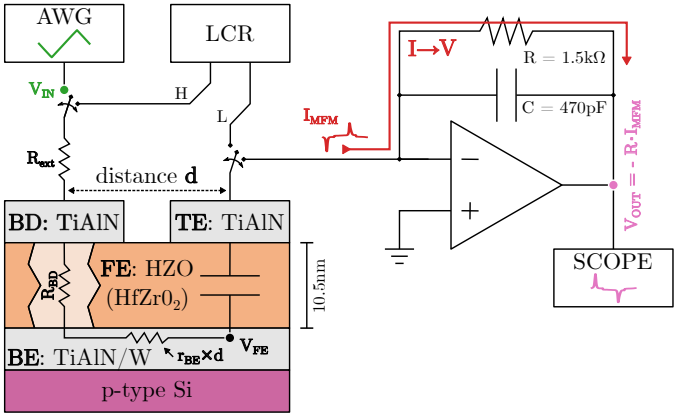


Fig. 1: Experimental setup used for the MFM characterization. The setup is used as an I→V converter to probe the switching current I_{MFM}. The Arbitrary Waveform Generator (AWG) drives the MFM sample and V_{OUT} is monitored through an oscilloscope. The setup includes also an LCR meter for AC characterization [4], as well as the possibility to insert an external resistance R_{ext} in series with the MFM sample.

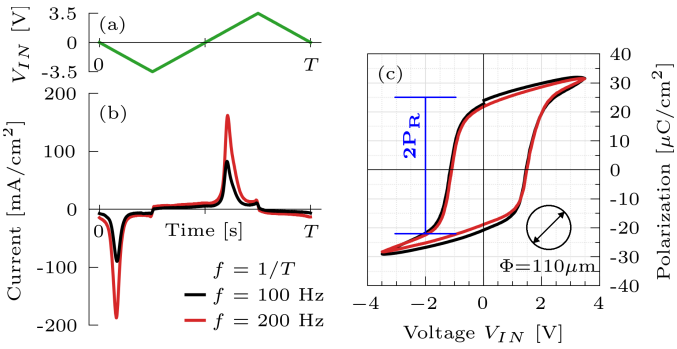


Fig. 2: (a) Input triangular voltage used to switch the MFM stacks. (b) Pulses at different frequencies f (ranging from 100 Hz to 10 kHz) are used to obtain the switching current of the MFM samples. Note that I_{MFM} is proportional to f at the reported frequencies. (c) P-V characteristics for an MFM with $\Phi = 110$ μm obtained by integrating the current curves in (b). $2P_R$ is defined in the figure.

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

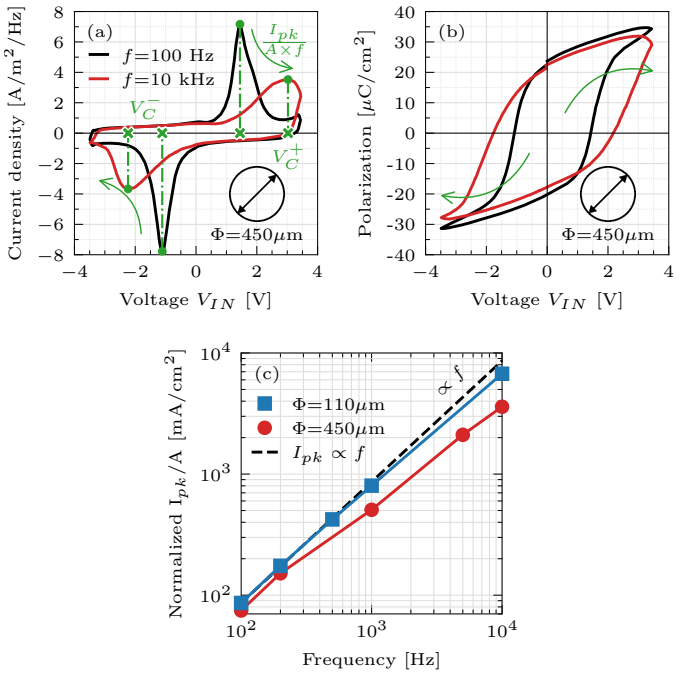


Fig. 3: (a) I_{MFM} normalized to frequency and sample area as a function of V_{IN} for a MFM with $\Phi = 450$ μm. At $f = 10$ kHz the I-V curve is largely distorted. (b) P-V curves obtained from the I-V characteristics in (a), also showing the distortion at $f = 10$ kHz. (c) Current peaks normalized to the sample area of two devices with diameter Φ=110μm (blue squares) and Φ=450μm (red circles). The peaks correctly scale linearly with both the area and frequency at low f, while they deviate from the linearity (dashed black) at high f due to the R_S-induced distortions, as demonstrated in the following of the manuscript.

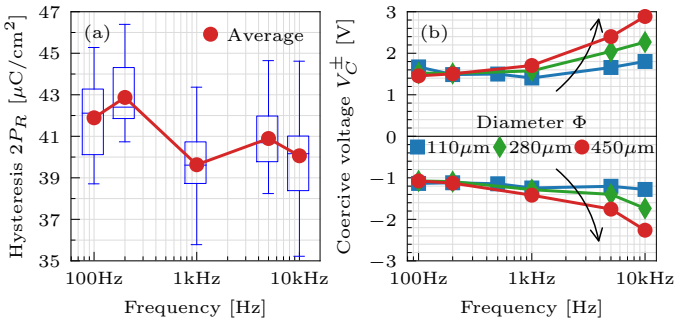


Fig. 4: (a) Box plot of extracted $2P_R$ values for all the measured samples (also with different areas). The average $2P_R$ seems independent of f, indicating a full switching of all devices. (b) Average coercive voltages V_C^+ (top) and V_C^- (bottom) versus f for different device areas. Their magnitude increases with f suggesting a slow response of the MFM at large f. The dependence on the device area is also evident.

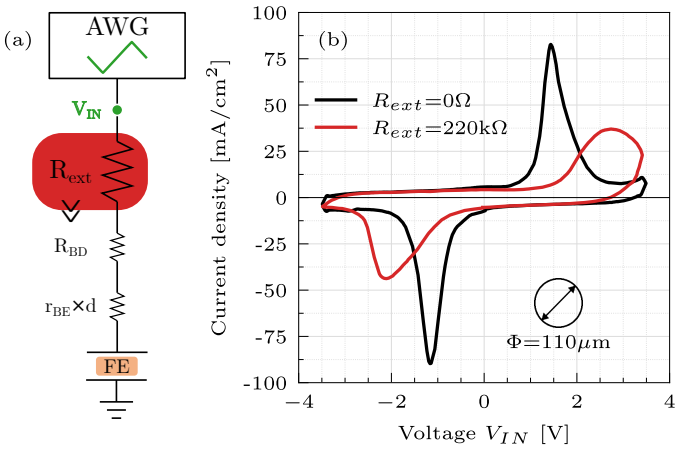


Fig. 5: I_{MFM} measured for a MFM sample with $\Phi = 110$ μm at $f = 100$ Hz. This experiment ensures the lowest possible current. (a) I_{MFM} is measured with and without a large external series resistance R_{ext}. (b) The series resistance effect is evident, with an I_{MFM} distortion (red) very similar to those observed in large samples measured at high f (Fig. 3a).

III. Experiments

Triangular V_{IN} pulses result in I_{MFM} peaks due to P switching, which adds up to the V_{IN}-independent dielectric response of the stack (Fig. 2b) [3]. Figures 2(b) and (c) show the I_{MFM}-V_{IN} and the P-V_{IN} curves of a sample with diameter Φ=110 μm. Two fairly low frequencies (f) for V_{IN} are shown, and we verified that I_{MFM} is proportional to f in this frequency range. Thus, in the following, we will report I_{MFM} normalized to both f and device area (A) to compare different samples and measurement conditions. Instead, P is independent of f as clearly shown in Fig. 2(c).

Figures 3(a) and (b) show results for a 450 μm sample. At 10 kHz, the normalized I_{MFM} (red) is largely distorted compared to 100 Hz (black). In particular, although the current peak I_{pk} is higher at 10 kHz than it is at 100 Hz (see Fig. 3c), the I_{pk} normalized to f in Fig. 3(a) is greatly reduced at 10 kHz and it is shifted in voltage. Consequently, the resulting P-V loop at 10 kHz tends to widen as the coercive voltages seem to increase (Fig. 3b).

This may suggest that the ferroelectric dynamics cannot track the V_{IN} time evolution [5]. To study this point, we measured many samples with different Φ and over a frequency range between 100 Hz and 10 kHz. We then defined the coercive voltages $V_C^{\pm}$ at the current peaks (Fig. 3a), while the $2P_R$ value was obtained from the P-V loops (see Fig. 2c). Figure 4(a) shows the box plot of all $2P_R$ values, which are rather constant in frequency, thus indicating a complete FE switching irrespective of f and Φ. Figure 4(b) reports instead the average $V_C^{\pm}$ grouped by device area, showing that $|V_C^{\pm}|$ increases with f and especially for the larger Φ, thus indicating that this trend could be an extrinsic, spurious effect. In this last respect, also I_{pk} increases with both Φ and f (Fig. 3c), suggesting a possible impact of a series resistance R_S.

IV. Series Resistance Effect

In order to verify a possible influence of R_S on the results of Figs. 3 and 4, we devised the experiment illustrated in Fig. 5. A 110 μm MFM is first measured at f=100 Hz (black), resulting in the lowest possible I_{MFM} (small Φ, low f) and thus in the minimum possible influence of R_S. Then, we repeated the measurement by inserting an external resistance R_{ext} in series with the MFM. The I-V curve (Fig. 5, red line) is now distorted, showing a decrease of the I_{MFM} peak and a $|V_C^{\pm}|$ modulation very similar to those observed in larger samples measured at high f (Fig. 3a). This confirms that such distortion of the current vs. time measurements in Fig. 3(a) can be at least partly due to R_S.

We estimated the contributions to R_S in our devices by using LCR meter-based measurements and inspecting the resistance between several broken devices at different locations on the wafer, as in Fig. 6. The contributions to R_S are also depicted and extracted in Fig. 6, and comprise the linear resistance r_{BE} of the bottom electrode (BE) and the

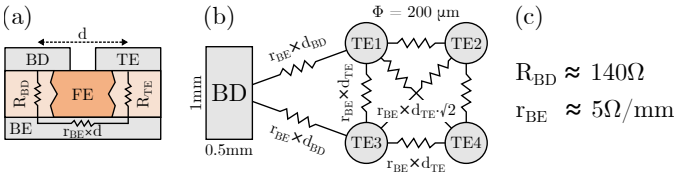


Fig. 6: The contributions to R_S (a) are estimated by measuring the resistance between several broken MFM samples (b) using the LCR meter in Fig. 1. R_{BD} is the resistance associated with the access device generally used for the MFM characterization, while R_{TE} is the resistance of the specific broken MFM. r_{BE} is the resistance per unit length of the bottom electrode. The detailed analysis of the different resistances measured in the network of (b) allows us to obtain the values in (c) that are then used to calculate the R_S value of the specific MFM under test as $R_S = R_{BD} + r_{BE} \cdot d$, where d is the distance between the access device BD and the MFM sample.

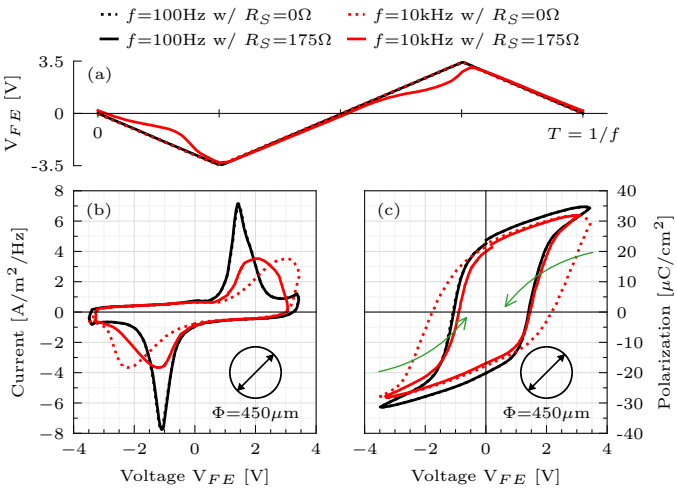


Fig. 7: Experiments on a Φ=450μm sample measured at two frequencies (*colors*). Measured data are plotted after calculating the voltage drop V_{FE} across the MFM by using two different R_S values (*linestyles*) in (2). (a) V_{FE} time evolution, which coincides with V_{IN} when $R_S = 0$ is considered in (2). (b) Measured I_{MFM}-V_{FE} characteristics. (c) Hysteretic P-V_{FE} curves.

resistance R_{BD} due to the broken device/access contact. Then, we calculate the series resistance for each device under test (DUT) as:

$$R_S = R_{BD} + r_{BE} \cdot d \tag{1}$$

where d is the distance of the DUT from the access contact. For the different MFM samples, we obtained R_S values in the order of 200 Ω. Given the I_{MFM} currents at play, such R_S values impact the I_{MFM} measurements even for the smallest samples when measuring at 10 kHz.

So, R_S induces a voltage drop that distorts the V_{FE} waveform actually delivered to the MFM (Fig. 7a, red line), especially at the switching peaks of I_{MFM}. Of course, the voltage drop on R_S depends on I_{MFM}, thus on Φ and f. In this last respect, now, we can estimate the actual voltage drop across the MFM stack, V_{FE}, from the measured I_{MFM} and the R_S calculated with (1) and the parameters in Fig. 6 as:

$$V_{FE} = V_{IN} - R_S \cdot I_{MFM} \tag{2}$$

Hence, we have re-plotted I_{MFM} and P versus V_{FE} in Figs. 7(b) and (c) (solid lines), respectively. The corrected P-V loop at 10 kHz (Fig. 7c, red solid line) is now very consistent with the one measured at 100 Hz (black solid line), thus indicating that the P-V distortion at 10 kHz was an artifact due to R_S.

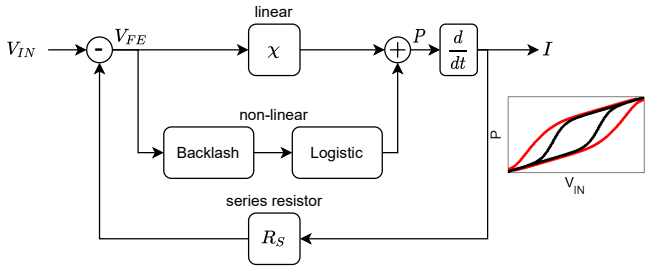


Fig. 8: Block diagram of the *Simulink* model used to simulate the MFM behaviour and the impact of R_S. The inset shows the obtained, qualitative P versus V_{IN} curves with R_S=0 (black) and with R_S≠0 (red).

However, even if such re-scaling of the voltage x-axis seems to restore the correct P-V hysteresis, it is not sufficient to match the I-V curves at different f (Fig. 7b) [6]. Indeed, the coercive voltages at 10 kHz (solid red) are now much closer to those at 100 Hz (black), but the I_{MFM} peaks (normalized to Φ and f) still remain lower at high f, since the y-axis of this plots is not affected by the correction in (2).

V. Compensation Procedure of the R_S Effect

The FE switching and thus the related I_{MFM} also depend on the actual shape of the V_{FE} waveform delivered to the MFM stack, which differs significantly for the two f values reported in Fig. 7(a). Hence, we devised a procedure to eliminate the R_S influence even on the I-V characteristics. Figures 7(a) and (c) indicate that P does not depend on the V_{FE} slew rate. So, we can write the switching current as [4]:

$$I_{MFM} = \frac{dP(V_{FE})}{dt} = \frac{dP}{dV_{FE}} \cdot \frac{dV_{FE}}{dt} = C_{LS} \cdot \frac{dV_{FE}}{dt} \tag{3}$$

where we have defined the large-signal capacitance C_{LS} linking P to V_{FE}. Then, by inverting (3) and recalling (2), we obtain:

$$\begin{aligned} C_{LS}(V_{FE}) &= I_{MFM} \cdot \left(\frac{dV_{FE}}{dt}\right)^{-1} = \\ &= I_{MFM} \cdot \left(\frac{dV_{IN}}{dt} - R_S \frac{dI_{MFM}}{dt}\right)^{-1} \end{aligned} \tag{4}$$

Equation (4) now returns C_{LS}, which is invariant w.r.t. the V_{FE} waveform and unaffected by the R_S distortion. Hence, C_{LS} should be used to extract dependably the relevant HZO parameters.

We validated our method by simulating a ferroelectric capacitor with different in-series R_S values with *Simulink* from *MathWorks*®. The block diagram of the model (Fig. 8) accepts the voltage applied to the MFM stack as the input and then returns the polarization of the device as the output. In order to simulate the ferroelectricity, the model sums a linear function, which models the linear dielectric behaviour of the polarization and is proportional to the input voltage, to a hysteresis function which, in turn, models the spontaneous polarization and is implemented by the cascade of a backlash model and of a logistic function.

In particular, the backlash emulates the hysteresis when the input voltage increases beyond the positive coercive voltage and then returns below the negative coercive voltage [7]. On the other hand, the logistic function models the strong increase

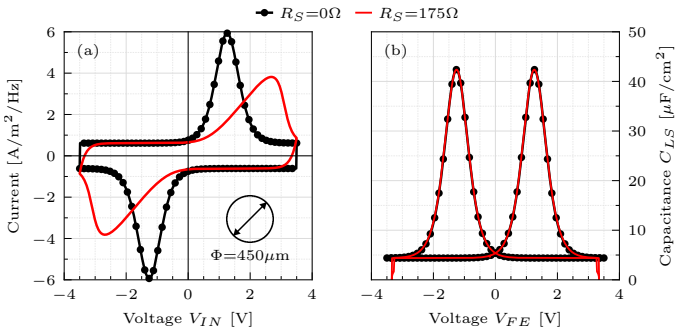


Fig. 9: Circuit simulation of a hysteretic capacitor with (red) or without (black) a series resistance R_S. (a) Simulated I-V curves highlight the R_S impact on the MFM characteristics (shift and lowering of switching peaks), which is qualitatively similar to the experimental observations in Figs. 3(a) and 5. (b) By using (4), we obtain a unique large-signal capacitance for the MFM stack, irrespective of R_S, thus validating the proposed correction method.

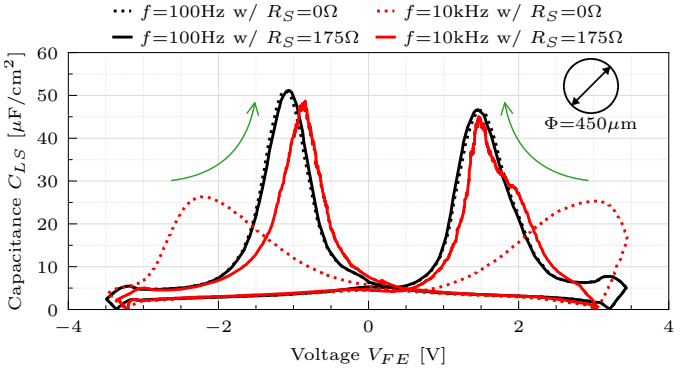


Fig. 10: Large-signal capacitance C_{LS} calculated with (4) applied to the I_{MFM} measured on a 450 μm sample for two frequencies f (*color*) and R_S (*linestyle*) values. *Dashed*: non-corrected C_{LS}-V_{FE} curves obtained by using R_S=0 in (2) and (4). *Solid*: corrected C_{LS}-V_{FE} curves obtained for R_S=175 Ω (estimated from Fig. 6). By accounting for R_S, a unique C_{LS}-V_{FE} curve for the different f values is obtained. Note that the C_{LS} peak can be used to monitor $V_C^{\pm}$ as well as to obtain the correct P-V loops.

of the polarization when V_{FE} is close to $V_C^{\pm}$ through a sigmoid function.

Then, we differentiate the polarization with respect to time to obtain the current through the device, from which we simulate the voltage drop across R_S and subtract it from V_{IN} to finally obtain V_{FE}. As an example, the inset in Fig. 8 shows the qualitative results of simulated polarization when R_S=0 (black curve) and when R_S≠0 (red curve).

Figure 9(a) shows the I_{MFM} simulated by assuming R_S=0 (black) or R_S=175 Ω (red), in agreement with the values in Fig. 6 and (1). When including R_S, the simulated I_{MFM} shows features similar to those observed in the experiments of Fig. 3(a). The application of (4) to the simulated I_{MFM} provides a unique C_{LS}-V_{FE} curve (see Fig. 9b), cleared from the undesired effects of R_S, thus validating our method.

Consequently, we applied (4) to the experiments of Fig. 3. Figure 10 shows that, by accounting for R_S in (2) and (4), we obtain the same C_{LS}-V_{FE} curve, irrespective of f, thus suggesting that, in the explored f range, the frequency and area dependence of I_{MFM} are mainly due to R_S.

The polarization and current can now be cleared out from the R_S-induced distortions as:

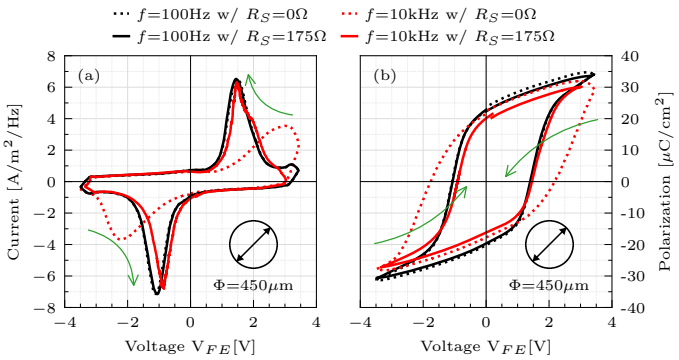


Fig. 11: I-V_{FE} (a) and P-V_{FE} (b) characteristics measured for two different frequencies (*color*) and R_S values plugged in (2), (5) and (6) (*linestyle*). R_S = 0 Ω results in V_{IN} = V_{FE} which returns the original, non-corrected characteristics, while R_S = 175 Ω yields the corrected ones. Note how the corrected current in (a) completely recovers from the distortions due to R_S, leading to a unique I-V_{FE} curve, irrespective of f. Also the P-V_{FE} loops in (b) recover a unique MFM characteristic, irrespective of the measurement frequency, after the correction procedure.

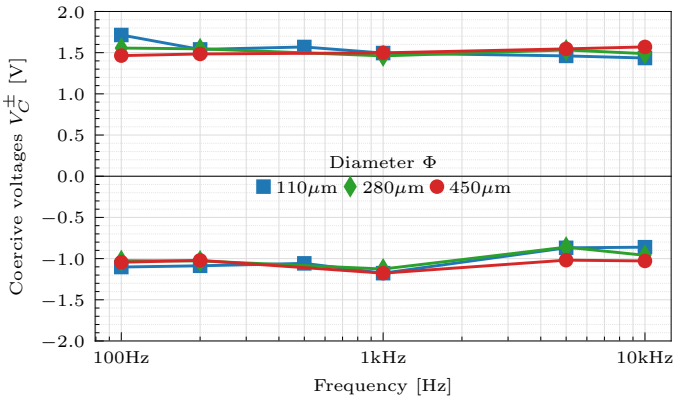


Fig. 12: Average V_C^+ and V_C^- values obtained after applying the correction procedure to all measurements performed on the MFM samples. Thanks to our correction, the $V_C^{\pm}$ are now rather constant in frequency and independent of the device area. For each MFM sample, the proper R_S value calculated through (1) and the parameters in Fig. 6 have been accounted for.

$$P = \int C_{LS}(V_{FE})dV_{FE} \quad (5)$$

$$I_{corr} = C_{LS} \cdot \frac{dV_{IN}}{dt} = I_{MFM} \cdot \left(1 - R_S \frac{dI_{MFM}}{dV_{IN}}\right)^{-1} \quad (6)$$

Figure 11 shows the final P-V and I-V characteristics obtained with (5) and (6), from which we can extract again the correct $V_C^{\pm}$ and $2P_R$ values. Note that the corrected I-V curves in Fig. 11(a, solid lines) completely recover from the distortions seen in the measured I_{MFM} (dashed). The correction method is finally applied to all our measurements and the results are summarized in Fig. 12, which shows that the $V_C^{\pm}$ values are now rather independent of both Φ and f.

VI. Conclusion

In this work, a statistical analysis of MFM remnant polarization and coercive voltages is reported. The extracted $2P_r$ values are rather constant irrespective of the device area and measurement frequency. Instead, the obtained $V_C^{\pm}$ highlight possible series resistance-induced distortions, which are then

confirmed and demonstrated. The effects of a series resistance R_S on the electrical characterization of MFM ferroelectric capacitors are then extensively assessed through dedicated experiments and simulations.

Then, for the first time, a rigorous procedure to compensate for R_S is presented. A large-signal capacitance is calculated and used to extract both $2P_r$ and $V_C^{\pm}$, whose values result rather frequency-independent in the explored frequency range [8]. This study hints that HZO has a high switching speed and the increase of the coercive voltages at high frequencies is to a large extent due to spurious effects due to undesired series resistances [9].

ACKNOWLEDGEMENTS

Funded by the European Union (Horizon Europe, FIXIT, GA no. 101135398; Horizon 2020, BeFerroSynaptic, GA no. 871737). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them.

REFERENCES

[1] S. Slesazeck and T. Mikolajick, "Nanoscale resistive switching memory devices: a review," *Nanotechnology*, vol. 30, no. 35, p. 352003, jun 2019. [Online]. Available: https://dx.doi.org/10.1088/1361-6528/ab2084

[2] M. Massarotto, F. Driussi, A. Affanni *et al.*, "Versatile experimental setup for ftj characterization," *Solid-State Electronics*, vol. 194, p. 108364, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038110122001368

[3] M. Massarotto, F. Driussi, A. Affanni *et al.*, "Novel experimental methodologies to reconcile large- and small-signal responses of hafnium-based ferroelectric tunnel junctions," *Solid-State Electronics*, vol. 200, p. 108569, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038110122003410

[4] M. Massarotto, M. Segatto, F. Driussi *et al.*, "Bridging large-signal and small-signal responses of hafnium-based ferroelectric tunnel junctions," in *2023 35th International Conference on Microelectronic Test Structure (ICMTS)*, 2023, pp. 1–6.

[5] X. Lyu, M. Si, X. Sun *et al.*, "Ferroelectric and anti-ferroelectric hafnium zirconium oxide: Scaling limit, switching speed and record high polarization density," in *2019 Symposium on VLSI Technology*, 2019, pp. T44–T45.

[6] Z. Gao, S. Lyu, and H. Lyu, "Frequency dependence on polarization switching measurement in ferroelectric capacitors," *Journal of Semiconductors*, vol. 43, no. 1, p. 014102, jan 2022. [Online]. Available: https://dx.doi.org/10.1088/1674-4926/43/1/014102

[7] Mathworks, "Model behavior of system with play," 2023. [Online]. Available: https://it.mathworks.com/help/simulink/slref/backlash.html

[8] P. Hao, S. Zheng, B. Zeng *et al.*, "Highly enhanced polarization switching speed in hfo2-based ferroelectric thin films via a composition gradient strategy," *Advanced Functional Materials*, vol. 33, no. 31, p. 2301746, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/adfm.202301746

[9] S. Boyn, A. M. Douglas, C. Blouzon *et al.*, "Tunnel electroresistance in BiFeO3 junctions: size does matter," *Applied Physics Letters*, vol. 109, no. 23, p. 232902, 12 2016. [Online]. Available: https://doi.org/10.1063/1.4971311

Impedance Measurement Platform for Statistical Capacitance and Current Characteristic Measurements of Arrayed Cells with Atto-order Precision

Koga Saito[1], Tatsuhiko Suzuki[1], Hidemi Mitsuda[1], Tsubasa Nozaki[1], Takezo Mawaki[1,2], and Rihito Kuroda[1,2]
[1]Graduate School of Engineering, Tohoku University, 6-6-05 Aza-Aoba, Aramaki, Aoba-ku, Sendai 980-8579, Japan
[2]New Industry Creation Hatchery Center, Tohoku University, 6-6-10 Aza-Aoba, Aramaki, Aoba-ku, Sendai 980-8579, Japan
Telephone: +81-22-795-4833, Fax: +81-22-795-4834
Email: koga.saito.s3@dc.tohoku.ac.jp

***Abstract*—A novel impedance measurement platform that enables high-precision statistical measurements of C-V and I-V characteristics is presented. The platform consists of $366^{H} \times 228^{V}$ cell array and a common readout circuit. The pre-formed device under test (DUT) cells of metal-oxide-silicon field-effect-transistors (MOSFETs) and metal-insulator-metal (MIM) capacitors were measured to verify its measurement performance. The results demonstrated that for about 5,000 DUTs, fF-order capacitance with 0.05 % precision and fA-order current with 0.46 % precision were achieved by C-V and I-V measurements, respectively.**

***Keywords*—Capacitor, current, statistical measurement, array test circuit, platform**

I. Introduction

The measurement of electrical characteristics using a device under test (DUT) is essential for developing semiconductor devices. For memory materials development such as for dynamic random-access memory (DRAM) [1] and ferroelectric memories [2], it is crucial to statistically measure and evaluate capacitance and current characteristics in detail, as these characteristics serve as indicators of performance. In materials development, using discrete test element groups (TEGs) with DUTs placed in parallel and measuring them with a wafer prober has become a widely adopted method. However, measuring electrical characteristics of a single DUT size is challenging due to limitations in the accuracy of the measurement system as well as, measuring a large number of TEGs is time-consuming. Therefore, it is essential to develop a platform that can measure electrical characteristics such as capacitance and current for a variety of arrayed DUTs. As a test circuit to realize capacitance measurement, the floating gate capacitance measurement (FGCM) method has been proposed to be useful for measuring capacitance with a simple circuit [3-5]. This method measures the change in the output voltage of the source follower circuit due to the charge conservation between the DUT and the reference capacitor. Furthermore, methods based on FGCM using correlated double sampling (CDS) to reduce the measurement errors [5] and a CMOS proximity capacitance image sensor that detects proximity capacitance in a two-dimensional array using CDS [6-7] have been proposed. In addition, an array test circuit technology has been proposed to achieve large-scale measurement of a single characteristic such as random telegraph noise (RTN), current, or resistance, from DUTs arranged in a cell array [8-12]. In this work, we developed an impedance measurement platform that can be formed with various materials as DUTs on the platform using a simple additional process. The cell circuit is configured to measure fA-order current as well as, fF-order capacitance using CDS, with high-precision. This paper presents the proposed circuit architecture and measurement results of the pre-formed DUTs.

II. Experimental

A. Platform structure

Fig. 1 (a-b) shows the cross-sectional diagram of the developed platform before and after the additional process. The chips were fabricated by a 0.18 µm 1-Poly-Si 5-Metal layers CMOS technology. As shown in Fig. 1 (a), the measurement circuits excluding the DUTs are formed in the platform and the circuits consist of a cell array with each cell contains a metal electrode using metal 5, and a common readout circuit. The entire platform is covered with a passivation film. DUTs are formed by an additional process that opens a hole on the passivation film in each cell as in Fig. 1 (b) and forms a material to be measured and the top electrode. Therefore, arbitrary DUTs can be formed or stacked on the platform using a simple additional process. In this paper, C-V and I-V measurements are to be performed using the pre-formed DUTs to verify its measurement performance.

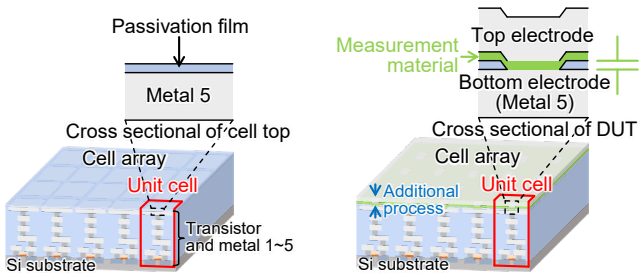


Fig. 1. Cross-sectional diagram of the developed platform (a) before the additional process and (b) after the additional process.

B. *Circuit architecture*

Fig. 2 shows the photograph and the overall circuit schematic diagram of the developed platform. The chip size is 4.8^{H} mm × 4.8^{V} mm and it consists of $366^{H} \times 228^{V}$ cell array and a common readout circuit. The basic circuit configuration was developed based on previously reported array test circuits [8-12], vertical and horizonal shift registers sequentially select rows and columns by giving pulses, and signals from cells can be read out at high-speed. The unit cell pitch is 10^{H} µm × 10^{V} µm and the types of pre-formed DUTs are divided by rows.

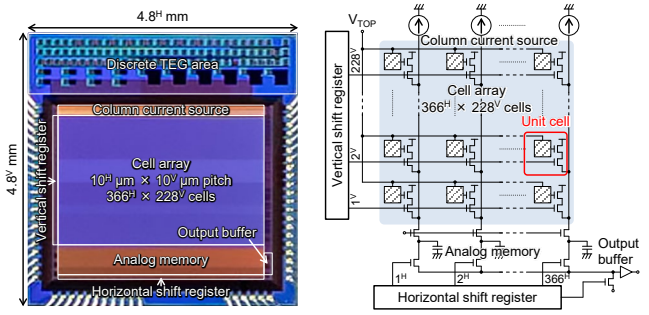


Fig. 2. (a) photograph and (b) overall circuit schematic diagram of the developed platform. The chip size is 4.8^{H} mm × 4.8^{V} mm and it consists of 366^{H} × 228^{V} cell array and a common readout circuit. The unit cell pitch is 10^{H} µm × 10^{V} µm and the types of pre-formed DUTs are divided by rows.

Fig. 3 shows the circuit schematic diagram of the unit cell. A DUT is formed in each cell, and signals are output to a common readout circuit via two source followers. The pre-formed DUTs are metal-oxide-silicon field-effect-transistors (MOSFETs) and metal-insulator-metal (MIM) capacitors formed simultaneously during the fabrication of the cell circuit. The circuit is configured to realize C-V and I-V measurements for each DUT cell.

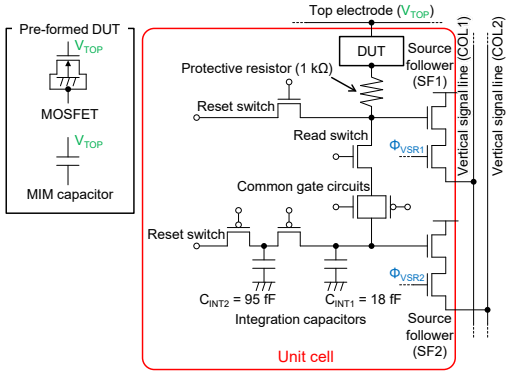


Fig. 3. Circuit schematic diagram of the unit cell. The pre-formed DUTs are formed simultaneously during the fabrication of the cell circuit. The circuit is configured to realize C-V and I-V measurements for each DUT cell.

C. *Capacirance measurement operation*

Fig. 4 shows the circuit operation of the C-V measurement. In the C-V measurement, a pulse voltage is applied to the top electrode (V_{TOP}) and the difference of signals (V_{OUT1} and V_{OUT2}) is read out with CDS. Fig. 4 (a) shows circuit schematic for the C-V measurement. The common readout circuit has two analog memories in the column, allowing for the switching of which memory holds the signal. Noise is eliminated by two output signals from the unit cell that are held in the analog memories and read out simultaneously. As shown in Fig. 4 (b-c), when a pulse voltage is applied to the V_{TOP}, the difference of the output signals (ΔV_{OUT}) is shown in (1) based on the charge conservation between the DUT capacitance (C_{DUT}) and the parasitic capacitance of the bottom electrode (C_C).

$$\Delta V_{OUT} = (C_{DUT} / (C_{DUT} + C_C)) \cdot \Delta V_{TOP} \cdot G_{SF} \quad (1)$$

Here, G_{SF} is the source follower gain and is measured for each cell to eliminate transistor variation. The parasitic capacitance C_C is calculated by the simulation based on the cell layout design; the value is 23.8 fF common to all cells excluding the DUT. From (1), C_{DUT} is expressed by (2).

$$C_{DUT} = C_C \cdot (\Delta V_{OUT} / (\Delta V_{TOP} \cdot G_{SF} - \Delta V_{OUT})) \quad (2)$$

Also, the applied bias voltage of the DUT (V_{DUT}) in the C-V measurement is the average voltage between the top and bottom electrodes as shown in (3). Here, V_R is the reset voltage of the bottom electrode.

$$V_{DUT} = (V_{TOP1} + V_{TOP2}) / 2 - V_R \quad (3)$$

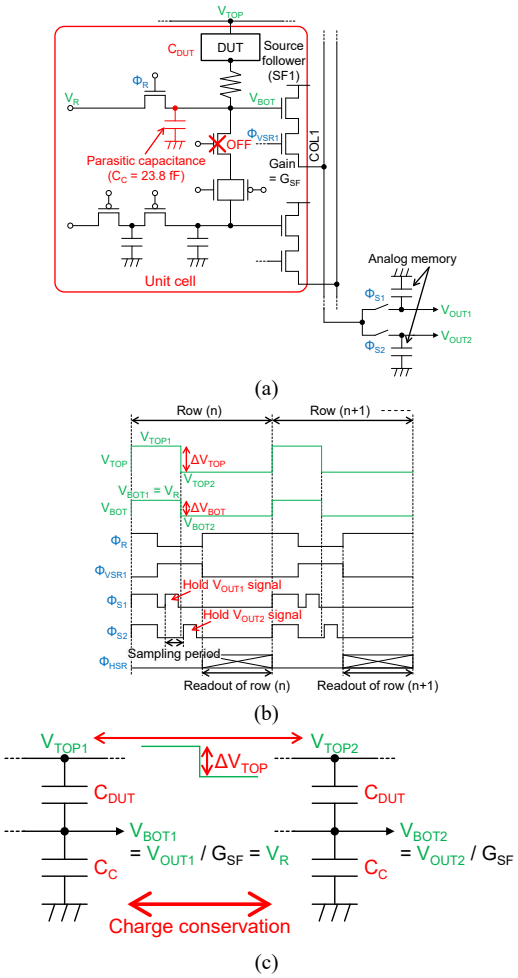


Fig. 4. Circuit operation of the C-V measurement. (a) Circuit schematic for measurement (b) Pulse timing diagram (c) Principle of measurement. Pulse voltage is applied to the top electrode (V_{TOP}) and the difference of signals is read out with CDS.

D. Current measurement operation

Fig. 5 shows the circuit operation of the I-V measurement. In the I-V measurement, a constant voltage is applied to the V_{TOP}, and the generated current (I_{DUT}) is measured from the time change of the signals. Fig. 5 (a) shows circuit schematic for the I-V measurement. The current passes through the common gate circuits and accumulates in the integration capacitor (C_{INT}), causing the potential change of the integration capacitor (V_{INT}) from the reset voltage (V_R). The common gate circuits have the function of keeping the bottom electrode voltage (V_{BOT}) even though V_{INT} is changed. As shown in Fig. 5 (b-c), the current can be calculated based on the accumulation of C_{INT} and is expressed by (4).

$$I_{DUT} = C_{INT} \cdot (dV_{INT} / dt) \quad (4)$$

When the V_{INT} change is large and saturated during the integration period, the calculation is performed with up to the signal saturation. Two integration capacitors (C_{INT1}, C_{INT2}) are formed and the capacitance of C_{INT} can be switched. The capacitance is calculated by the simulation based on the cell design; the values are 18 fF and 95 fF common to all cells. Also, the V_{BOT} is read out by another source follower, the applied voltage of the DUT (V_{DUT}) in the I-V measurement is the difference between the top and bottom electrode voltages as shown in (5).

$$V_{DUT} = V_{TOP} - V_{BOT} \quad (5)$$

E. Measurement conditions

Table I shows the measured samples and measurement conditions. Two types of MOSFETs with different dielectric thicknesses and two types of MIM capacitors with different electrode areas were measured on the platform. Also, the discrete TEGs consist of multiple DUTs placed in parallel were measured to compare the measurement results. The discrete TEG samples were measured using a semiconductor device parameter analyzer (Keysight B1500A) as references. For the platform, capacitance was calculated as the average of 1,000 frames at each voltage and current was calculated by the signal change of 200 frames at each voltage. The parasitic capacitance C_C of the MOSFETs increases due to the depletion capacitance between the source-drain and substrate. Therefore, in the C-V measurement of the platform, the capacitances calculated by the simulation based on each transistor design were added to C_C. For the I-V measurement, only MOSFET samples with Tox of 4.4 nm were measured for their gate leakage current.

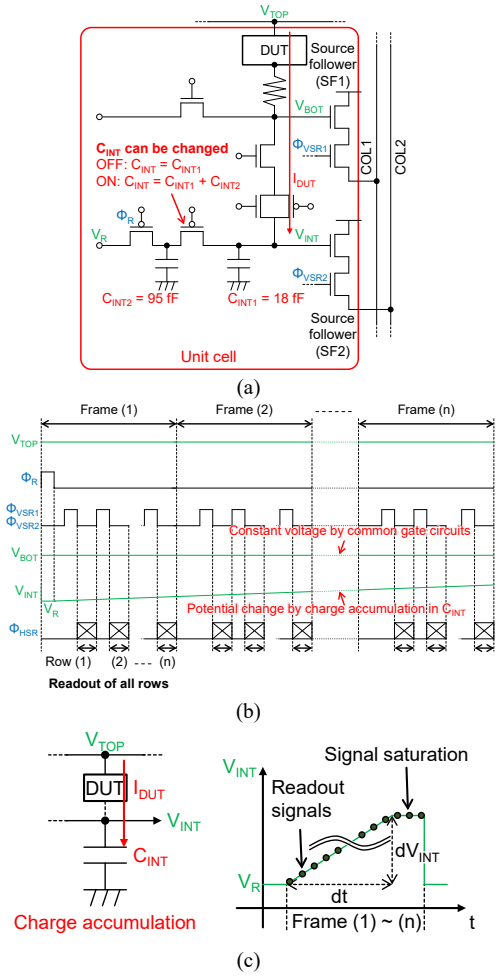


Fig. 5. Circuit operation of the I-V measurement. (a) Circuit schematic for measurement (b) Pulse timing diagram (c) Principle of measurement. The current is calculated from the time change of the signals by the integration capacitor.

TABLE I. SUMMARY OF MEASURED SAMPLES AND MEASUREMENT CONDITIONS

Device type	Platform				Discrete TEG			
Method	Platform circuit (Array test circuit)				Semiconductor device parameter analyzer (Keysight B1500A)			
DUT type	MOSFET (Tox: 4.4 nm)	MOSFET (Tox: 7.7 nm)	MIM cap. (Area: 16 μm²)	MIM cap. (Area: 40 μm²)	MOSFET (Tox: 4.4 nm)	MOSFET (Tox: 7.7 nm)	MIM cap. (Area: 16 μm²)	MIM cap. (Area: 40 μm²)
Electrode area	4 μm²	4 μm²	16 μm²	40 μm²	4 μm² × 40 parallel	4 μm² × 40 parallel	16 μm² × 30 parallel	40 μm² × 30 parallel
Number of DUTs	1,464 ($366^H \times 4^V$)	1,464 ($366^H \times 4^V$)	1,098 ($366^H \times 3^V$)	1,098 ($366^H \times 3^V$)	1	1	1	1
Measurement temperature	Room temperature							
C-V measurement conditions	Applied voltage = -2 – 2 V, 0.05 V step, ΔV_{TOP} = 0.5 V, V_R = 2.3 V, Sampling period = 15 μs, C_C = 23.8 fF, 1,000 frames				-2 – 2 V, 0.05 V step, V_{OSC} = 25 mV, Frequency = 20 kHz			
I-V measurement conditions	V_{TOP} = 0 – 5.0 V (Applied voltage ≈ 0 – 2.5 V), 0.05 V step, C_{INT} = 113 fF, 200 frames	-	-	-	0 – 2.5 V (Source-drain bias = 2.4 V), 0.05 V step	-	-	-

III. Measurement Results and Discussions

A. C-V measurement results

Fig. 6 shows measured signals of platform cells with MOSFETs samples in the C-V measurement. Plotted cells were selected randomly. It can be confirmed that the difference in output signals varies according to the applied voltage. The capacitance is calculated for all cells at all voltages using (2) from the average of 1,000 frames.

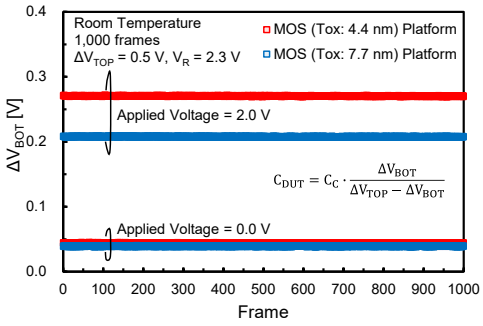


Fig. 6. Measured signals of platform cells with MOSFETs samples in the C-V measurement. Plotted cells were selected randomly. The capacitance is calculated from the average of 1,000 frames.

Fig. 7 shows C-V characteristics of all samples. For the platform, the average of all measurement cells is plotted. The voltage dependence of MOSFETs and MIM capacitors were confirmed in the platform samples. The obtained values agree well between the discrete TEG and platform.

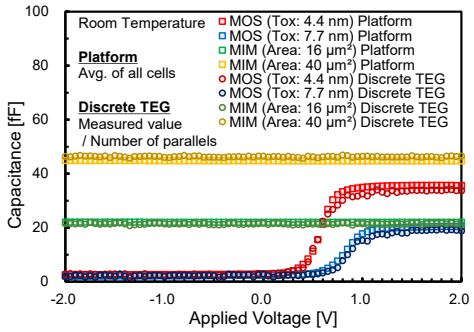


Fig. 7. C-V characteristics of all samples. For the platform, the average of all measurement cells is plotted.

Fig. 8 shows histogram of the measured capacitance variation in the platform samples at the applied voltage is 2.0 V. The result confirms that the capacitance variation is within a few percent. From (2), the variation of C_C directly affects the capacitance calculation, but the pre-formed DUTs were formed simultaneously during the fabrication of the cell circuit; thus, the variation of C_C is also expected to be within a few percent.

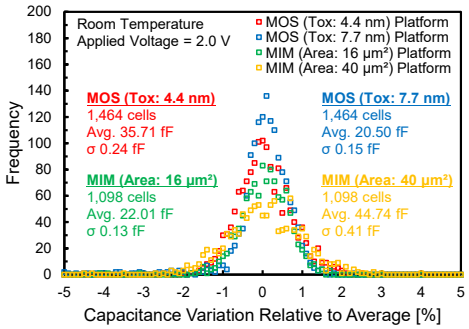


Fig. 8. Histogram of the measured capacitance variation at applied voltage is 2.0 V.

B. I-V measurement results

Fig. 9 shows measured signals of a platform cell in the I-V measurement. A plotted cell was selected randomly. It can be confirmed that the common gate circuits keep the V_{BOT} constant, and the signal change increases as the applied voltage increases. The current and applied voltage are calculated for all cells at all voltages using (4) and (5).

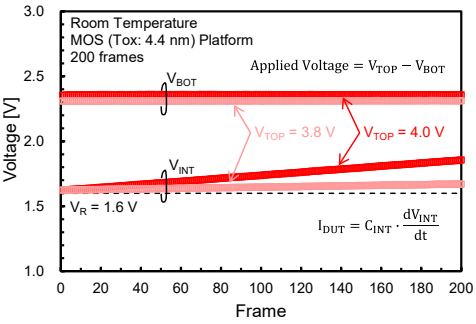


Fig. 9. Measured signals of a platform cell in the I-V measurement. A plotted cell was selected randomly. The current is calculated from the V_{INT} change and the applied voltage is calculated from the average of V_{BOT}.

Fig. 10 shows I-V characteristics of MOSFET (Tox: 4.4 nm) samples. For the platform, the average of all measurement cells is plotted. For the discrete TEG, a bias voltage was applied to the source-drain to align with the platform measurement conditions. At the room temperature, the measured current of the platform is affected by the background current originated from junction leakage at diffusion regions, but its I-V characteristics is close to the discrete TEG. The effect of the background current can be reduced by lowering the measurement temperature [12].

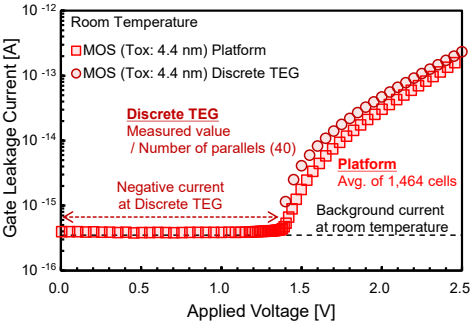


Fig. 10. I-V characteristics of MOSFET (Tox: 4.4 nm) samples. For the platform, the average of all measurement cells is plotted.

Fig. 11 shows cumulative probability of the measured current at the applied voltage is 2.0 V. In the I-V measurement, the applied voltage is not a common value for all cells since it is calculated by (5) for each cell; thus, the results are plotted for V_{TOP}, where the average applied voltage of all cells is 2.0 V. It can be confirmed that the current is exponentially distributed.

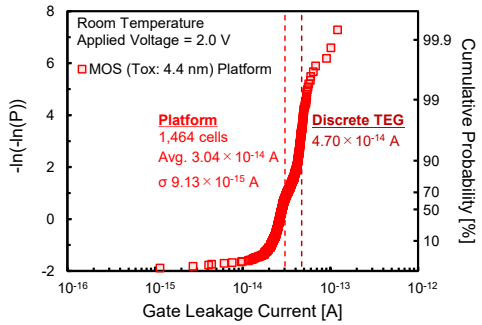


Fig. 11. Cumulative probability of the measured current at applied voltage is 2.0 V.

C. Measurement precision

In order to verify the measurement precision of the platform, the applied voltage was fixed, and measurements were repeated for 50 times with the same measurement conditions. The average and standard deviation of all measurement cells were obtained. The correlation between standard deviation and average is shown in Fig. 12 for the capacitance and Fig. 13 for the current. The dotted line represents the calculated variation relative to the average. The average precision of all samples at the voltages of the 50 repetitions measurements was 0.05 % for the C-V measurements and 0.46 % for the I-V measurements; thus, the precision is atto-order for femto-order measurements.

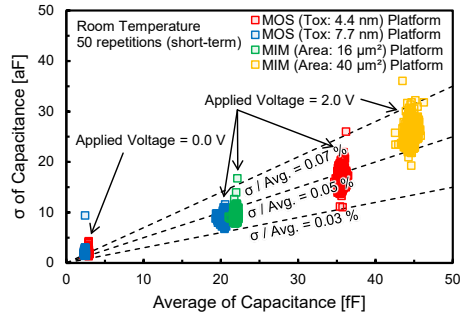


Fig. 12. The correlation between standard deviation (σ) and average of the capacitance for the 50 repetitions measurements in short-term.

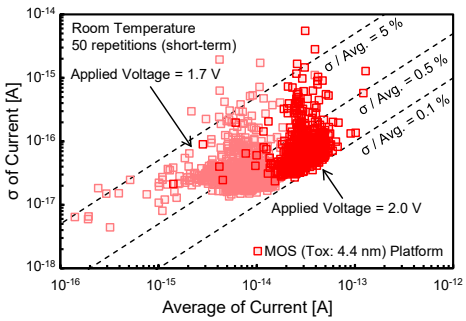


Fig. 13. The correlation between standard deviation (σ) and average of the current for the 50 repetitions measurements in short-term.

For the C-V measurement, the effect of signal noise on precision is discussed below. Equation (2) differentiated with respect to the output signal (ΔV_{OUT}) is expressed by (6).

$$C_{DUT}' = C_C \cdot (\Delta V_{TOP} \cdot G_{SF}) / (\Delta V_{TOP} \cdot G_{SF} - \Delta V_{OUT})^2 \quad (6)$$

Equation (6) shows that the increase in capacitance is not linear with respect to the increase in output signal, indicating that the effect of signal noise depends on the measured capacitance. Assuming that the measurement precision is due to signal noise (V_{RMS}), the calculation using (2) and (6) is the dotted line shown in Fig. 14. The correlation between the standard deviation and the average follows the dotted line, indicating that the precision of capacitance measurement is affected by the V_{RMS}, which is less than 100 μV. This result shows the effect of noise is well reduced by readout with CDS.

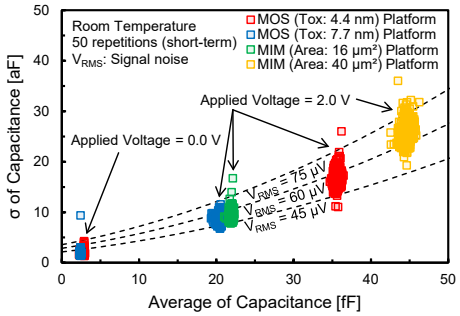


Fig. 14. The correlation between standard deviation (σ) and average of the capacitance with additional dotted line calculated by signal noise.

D. Performance summary

Table II shows the design specification and Table III shows the performance summary of the developed platform. Frame period depends on the number of measurement cells; thus, reducing the number of measurement cells is effective to realize more high-speed measurements. The measurement range is calculated by changing the applied voltage of the C-V measurement and the sampling time of the I-V measurement using the calculation formulas. For the I-V measurement, the influence of the background current can be reduced by measuring at low temperatures. These results demonstrate that the developed impedance measurement platform enables high-precision and high-speed statistical measurements of C-V and I-V characteristics.

TABLE II. DESIGN SPECIFICATION OF THE DEVELOPED PLATFORM

Process technology	0.18 μm 1-poly-Si 5-Metal layers CMOS technology	
Chip size	4.8^H mm × 4.8^V mm	
Cell pitch	10^H μm × 10^V μm	
Number of cells	$366^H \times 228^V$	
Number of cells for additional process	$366^H \times 116^V$	
Number of pre-formed DUT cells (Measured samples)	MOSFET (Tox: 4.4 nm): $366^H \times 4^V$ = 1,464 cells	MOSFET (Tox: 7.7 nm): $366^H \times 4^V$ = 1,464 cells
	MIM cap. (Area: 16 μm²): $366^H \times 3^V$ = 1,098 cells	MIM cap. (Area: 40 μm²): $366^H \times 3^V$ = 1,098 cells

TABLE III. PERFORMANCE SUMMARY OF THE DEVELOPED PLATFORM

	C-V measurement	I-V measurement
Measurement method	Charge conservation with CDS	Charge accumulation
Signal output frequency	300 kHz	
Frame period (Depends on number of measurement cells)	30 ms/frame (Measurement cells = $366^{H} \times 24^{V}$ cells)	
Measured sample characteristics	1 – 50 fF	10^{-16} – 10^{-13} A
Measurement range (Calculated value)	0.1 – 100 fF (Calculated at C_C = 23.8 fF)	10^{-17} (at 243 K) – 10^{-11} A (Calculated at C_{INT} = 113 fF)
Measurement precision (σ / Avg. of the repeatability)	0.05 % (at 1 – 50 fF)	0.46 % (at 10^{-16} – 10^{-13} A)

IV. CONCLUSION

An impedance measurement platform that can measure C-V and I-V characteristics of DUTs in a cell array with high-precision is developed. Statistical measurements of capacitance and current were shown using about 5,000 pre-formed DUTs, and the C-V measurement can measure fF-order capacitance with 0.05 % precision with 1,000-times averaging and the I-V measurement can measure fA-order current with 0.46 % precision with 200-times sampling under 300 kHz signal output frequency and 30 ms frame period. In both C-V and I-V measurements, the precision is atto-order for femto-order measurements. In the calculated value, capacitance measurement range is 0.1 – 100 fF and current measurement range is 10^{-17} – 10^{-11} A. Since a DUT on the platform can be formed arbitrarily with an additional process, statistical evaluations of capacitance and current characteristics for various semiconductor devices will be realized.

ACKNOWLEDGMENT

We would like to thank Dr. Akinobu Teramoto of Hiroshima University, for his useful advice. This work was partly supported by JSPS KAKENHI under Grant Numbers 22J20102 and 22KJ0283.

REFERENCES

[1] A. Spessot and H. Oh, "1T-1C Dynamic Random Access Memory Status, Challenges, and Prospects," IEEE Trans. Electron Devices, vol. 67, no. 4, pp. 1382-1393, April 2020.

[2] T. Mikolajick, U. Schroeder, and S. Slesazeck, "The Past, the Present, and the Future of Ferroelectric Memories," IEEE Trans. Electron Devices, vol. 67, no. 4, pp. 1434-1443, April 2020.

[3] C. Kortekaas, "On-chip Quasi-static Floating-gate Capacitance Measurement Method," in IEEE ICMTS, vol. 3, pp. 109-113, March 1990.

[4] H.P. Tuinhout, H. Elzinga, J.T.H. Brugman, and F. Postma , "The Floating Gate Measurement Technique for Characterization of Capacitor Matching," IEEE Trans. Semicond. Manuf., vol. 9, no. 1, pp. 2-8, February 1996.

[5] Y.-C. Kwon and O.-K. Kwon, "A Precision Mismatch Measurement Technique for Integrated Capacitor Array Using a Switched Capacitor Amplifier," IEEE Trans. Semicond. Manuf., vol. 26, no. 2, pp. 226-232, May 2013.

[6] M. Yamamoto, et al., "A CMOS Proximity Capacitance Image Sensor with 16μm Pixel Pitch, 0.1aF Detection Accuracy and 60 Frames Per Second," In Proceedings of the 2018 IEEE Int. Electron Devices Meeting, pp. 660-663, December 2018.

[7] Y. Sugama, et al., "Two High-Precision Proximity Capacitance CMOS Image Sensors with Large Format and High Resolution," Sensors, vol. 22, no. 2770, pp. 1-20, April 2022.

[8] S. Watabe, S. Sugawa, A. Teramoto, and T. Ohmi, "New Statistical Evaluation Method for the Variation of Metal–Oxide–Semiconductor Field-Effect Transistors," Jpn. J. Appl. Phys, vol. 46, no. 4B, pp. 2054-2057, April 2007.

[9] S. Watabe, et al., "A Simple Test Structure for Evaluating the Variability in Key Characteristics of a Large Number of MOSFETs," IEEE Trans. Semicond. Manuf., vol. 25, no. 2, pp. 145-154, May 2012.

[10] K. Abe, et al., "A Test Circuit for Statistical Evaluation of p–n Junction Leakage Current and Its Noise," IEEE Trans. Semicond. Manuf., vol. 25, no. 3, pp. 303-309, August 2012.

[11] T. Maeda, Y. Omura, R. Kuroda, A. Teramoto, T. Suwa, and S. Sugawa, "Resistance Measurement Platform for Statistical Analysis of Emerging Memory Materials," IEEE Trans. Semicond. Manuf., vol. 33, no. 2, pp. 232-239, May 2020.

[12] K. Saito, et al., "A high-precision current measurement platform applied for statistical measurement of discharge current transient spectroscopy of traps in SiN dielectrics," Jpn. J. Appl. Phys, vol. 60, pp. 086501-1-086501-10, July 2021.

Compact expression to model the effects of dielectric absorption on analog-to-digital converters

Simone Saro[1], Pierpaolo Palestri[2], Enrico Caruso[3], Paolo Toniutti[3],
Rocco Calabro[3], Semen Terokhin[3], Francesco Driussi[1]
[1]DPIA, University of Udine, Italy
[2]was with University of Udine, now with University of Modena and Reggio Emilia, Italy
[3]Infineon Technologies Austria, Villach, Austria
Email: saro.simone001@spes.uniud.it

***Abstract*—An analytical model for the dielectric absorption on capacitors and its impact on the errors induced in ADC conversion is here proposed. The reported simulations are consistent with the results of the R-C model widely used in the literature and well reproduce a large set of capacitance versus frequency and current versus time experiments. We also measured ADC conversion errors due to the dielectric absorption and we demonstrated that such errors can be reproduced by our model just calibrating one single technology parameter.**

I. Introduction

The literature widely reports that the dielectric constant of insulators depends on the frequency and it has a complex value [1]. While its real part is responsible for the capacitance value of the metal-insulator-metal (or semiconductor-insulator-metal) capacitors, the imaginary part results in an AC conductance in parallel to the capacitance. This effect is known as *dielectric absorption* or *dielectric relaxation* [1].

In the time domain, the dielectric absorption (DA) manifests itself as a slow current transient going roughly as $1/t$ (inverse of the time) after the application of a voltage step. Of course, this spurious current adds to the expected Dirac-delta current peak associated to the charging/discharging of the capacitor during the voltage step [2].

This long current transient has a negative impact on the performance of several electronic circuits/systems. Among the others, the analog-to-digital converters (ADCs) are affected by the detrimental impact of the DA [3]–[5], since, during the hold phase, the DA induced current partially discharges/charges the sampling capacitors, lowering/increasing the voltage to be converted with respect to the voltage applied as an input during the sampling phase.

II. Modeling framework

In [6] the complex relative permittivity for FR-4 substrates is described as

$$\varepsilon_r(\omega) = \varepsilon_\infty + \sum_{i=1}^{N} \frac{\Delta\varepsilon_i}{1+j\frac{\omega}{\omega_i}} - j\frac{\sigma}{\omega\varepsilon_0} \tag{1}$$

where ε_0 is the permittivity in vacuum, ε_∞ is the real part of the relative permittivity at high frequencies, $\Delta\varepsilon_i$ is the i^{th} variation of the permittivity and ω_i is the angular frequency where the i^{th} variation is centered. The last term in (1) is related to the volume conductivity (σ) of the dielectric. This term can be neglected in this analysis because it would result in a resistor in parallel to the capacitor and it is significant only at very low frequencies [6].

By assuming constant $\Delta\varepsilon_i$ for ω_i evenly distributed on a logarithmic scale between the angular frequencies $\omega_1 = 10^{m_1}$ and $\omega_2 = 10^{m_2}$ and by taking an infinite number of i terms inside such interval [6], the sum in (1) becomes:

$$\begin{aligned}\sum_{i=1}^{N} \frac{\Delta\varepsilon_i}{1+j\frac{\omega}{\omega_i}} &= \frac{\Delta\varepsilon}{m_2-m_1}\int_{x=m_1}^{m_2} \frac{dx}{1+j\frac{\omega}{10^x}} \\ &= \frac{\Delta\varepsilon}{m_2-m_1}\frac{\ln\left(\frac{\omega_2+j\omega}{\omega_1+j\omega}\right)}{\ln 10}\end{aligned} \tag{2}$$

where $\Delta\varepsilon$ is the total variation of ε_r between the ω_1 and ω_2. Furthermore, for $\omega_1 \ll \omega \ll \omega_2$ we can write:

$$\ln\left(\frac{\omega_2+j\omega}{\omega_1+j\omega}\right) \simeq \ln\left(\frac{\omega_2}{j\omega}\right) = \ln\left(\left|\frac{\omega_2}{\omega}\right|\right) - j\frac{\pi}{2} \tag{3}$$

Substituting (2) and (3) in (1) the complex permittivity becomes:

$$\varepsilon_r(\omega) = \varepsilon_\infty + \frac{\Delta\varepsilon}{\ln\left(\frac{\omega_2}{\omega_1}\right)}\left[\ln\left(\left|\frac{\omega_2}{\omega}\right|\right) - j\frac{\pi}{2}\right] \tag{4}$$

By using the definition of the admittance we have:

$$\begin{aligned}Y(\omega) &= j\omega\varepsilon_r\varepsilon_0\frac{A}{d} \\ &= \frac{\omega\varepsilon_0 A}{d}\left[j\varepsilon_\infty + j\frac{\Delta\varepsilon}{\ln\left(\frac{\omega_2}{\omega_1}\right)}\ln\left(\left|\frac{\omega_2}{\omega}\right|\right) + \frac{\pi\Delta\varepsilon}{2\ln\left(\frac{\omega_2}{\omega_1}\right)}\right]\end{aligned} \tag{5}$$

where d is the insulator thickness and A is the capacitor area. Now, starting from the definition of the loss tangent/dissipation factor of the capacitor, we derive:

$$\tan\delta = \frac{\Re(Y)}{\Im(Y)} = \frac{\pi\Delta\varepsilon}{2\ln\left(\frac{\omega_2}{\omega_1}\right)} \cdot \frac{1}{\varepsilon_\infty + \frac{\Delta\varepsilon}{\ln\left(\frac{\omega_2}{\omega_1}\right)}\ln\left(\left|\frac{\omega_2}{\omega}\right|\right)} \tag{6}$$

Typically, the variation $\Delta\varepsilon$ is much smaller than ε_∞, so we approximate (6) as:

$$\tan\delta = \frac{\pi\Delta\varepsilon}{2\varepsilon_\infty \ln\frac{\omega_2}{\omega_1}} \tag{7}$$

From (7) we obtain $\Delta\varepsilon$ to be substituted in (4), finally obtaining the relative permittivity as:

$$\varepsilon_r(\omega) = \varepsilon_\infty\left[1 + \tan\delta\frac{2}{\pi}\left[\ln\left(\left|\frac{\omega_2}{\omega}\right|\right) - j\frac{\pi}{2}\right]\right] \tag{8}$$

979-8-3503-2987-2/24/$31.00 ©2024 IEEE

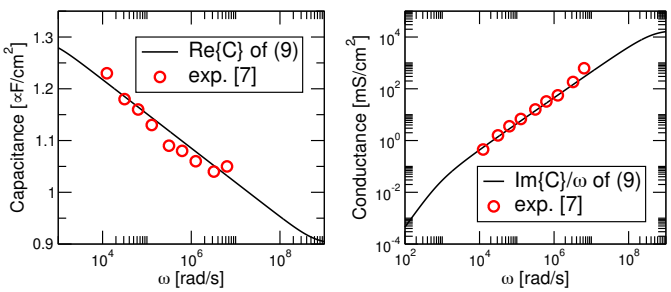


Fig. 1. Comparison between the capacitance (left) and the conductance (right) calculated with (9) and the experiments in [7]. The model parameter values used to fit the measurements are $C_\infty = 0.9$ μF/cm^2, $f_1 = \frac{\omega_1}{2\pi} = 100$ Hz, $f_2 = \frac{\omega_2}{2\pi} = 100$ MHz and $\tan\delta = 0.0502$.

that is essentially the model proposed in [6] for FR-4 substrates, just recast to display explicitly $\tan\delta$, that is the constant tangent of the loss-angle in the angular frequency interval between ω_1 and ω_2. Then, the complex capacitance C can be easily expressed as:

$$C(\omega) = \frac{\varepsilon_r(\omega)\varepsilon_0}{d} = C_\infty\left[1+\tan\delta\frac{2}{\pi}\left[\ln\left(\left|\frac{\omega_2}{\omega}\right|\right)-j\frac{\pi}{2}\right]\right] \quad (9)$$

where C_∞ is the capacitance at infinite frequency.

From the capacitance versus frequency equation, one can also derive the current versus time relationship. Applying a voltage step (with amplitude V) to the capacitance in (9), the current response would be:

$$\begin{aligned} I(\omega) &= V(\omega)j\omega C(\omega) \\ &= C_\infty V\left[1+\tan\delta\frac{2}{\pi}\left[\ln\left(\left|\frac{\omega_2}{\omega}\right|\right)-j\frac{\pi}{2}\right]\right] \end{aligned} \quad (10)$$

Through the inverse Fourier transform, the DA current over time is given by:

$$\frac{I(t)}{P} = \frac{I(t)}{C_\infty V_{IN}} = \frac{2\tan\delta}{\pi}\cdot\frac{1}{t} \quad (11)$$

which indeed results in a $1/t$ trend of the current.

Figure 1 shows that the complex capacitance derived in (9) is able to reproduce well experimental results for integrated capacitors in terms of both capacitance and conductance versus frequency f [7]. Furthermore, in Fig. 2 and 3 , our model reproduces well both the current versus t experiments and the capacitance versus f data for MIM capacitors featuring different dielectrics reported in [4], [5] by using $\tan\delta$ values in the order of 10^{-3}.

Therefore, in this work we propose the use of (9) to model the dielectric absorption, as an alternative to the widely used R-C model [5], [8] employing many R-C dipoles in parallel to the main capacitor (in Fig. 4(a) it is reported only the case with two of them).

Although phenomenological, the R-C model is however helpful to understand the effect of dielectric absorption on the ADC performance. In Fig. 4(b), we see that during the sampling phase the capacitor C (completely discharged in the idle phase) is set to a voltage V. This is not the case for C_1 and C_2 that slowly charge. During the hold phase, the current flowing

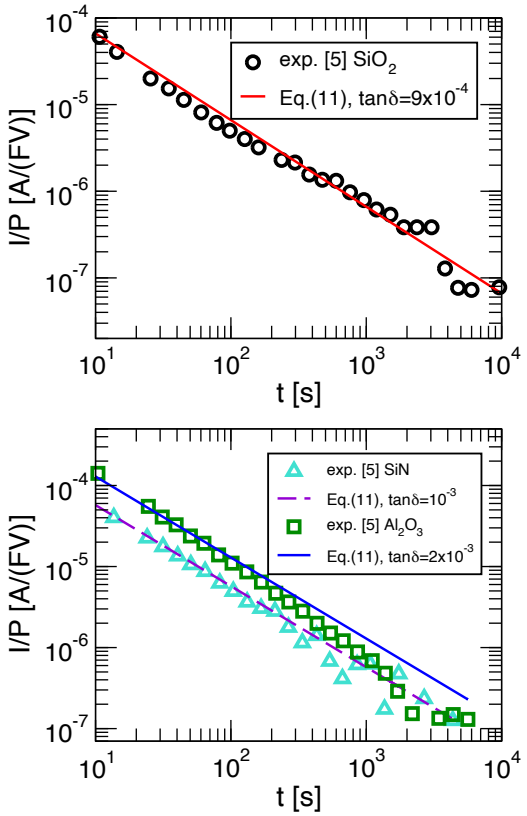


Fig. 2. Comparison between our model and the experiments in [5]. Equation (11) is used to fit the current vs. time experiments in [5]. In this case the sole model parameter is $\tan\delta$.

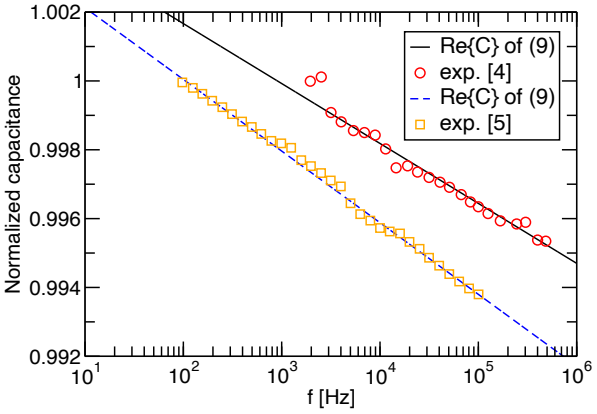


Fig. 3. Comparison between the results of (9) and the C vs. f curves measured in [4], [5]. Experiments are provided normalized to the capacitance measured at the lowest measurement frequency [4], [5]. In (9), we used $f_1 = \frac{\omega_1}{2\pi} = 10$ mHz and $f_2 = \frac{\omega_2}{2\pi} = 500$ MHz. For the black solid line we used $\tan\delta = 1.2\cdot10^{-3}$, while for the dashed blue line we used $\tan\delta = 1.44\cdot10^{-3}$.

in the R-C branches discharges the capacitor C, lowering the voltage.

In the R-C model, the capacitance C_i of the i-th R-C branch is typically defined through the parameter α that represents the relative capacitance decrease per frequency decade [see Fig. 4(a)]. Instead R_i is obtained defining the time constant $\tau_i = R_iC_i$ characterizing the i-th R-C branch.

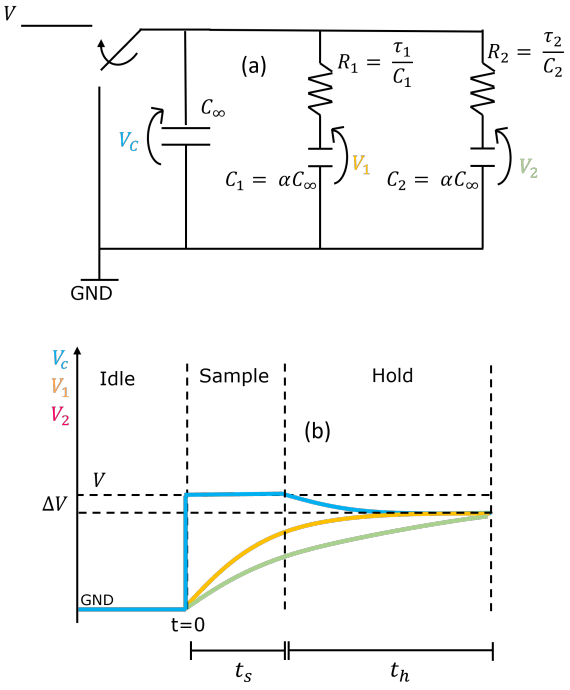


Fig. 4. (a) sketch of the sample-and-hold circuit in which the dielectric absorption is modeled through discrete R-C circuits. The switch can assume three different positions: in the idle phase, C is connected to ground resulting in its complete discharge; in the sample phase, the voltage V charges C during the sample time (t_s); in the hold phase the switch disconnects C from the input and the measurement of the stored voltage occurs after the hold time (t_h). The hold time includes also the time required for data conversion (that is however much shorter). (b) time evolution of the voltages across the capacitors. During t_h, the charge stored in C redistributes on C_1 and C_2, decreasing the voltage V_C on the main capacitor. $\Delta V = V - V_C$ after t_h [see also (12)].

Now, when considering a single R-C branch with constant $\tau_1 = R_1 C_1$, the capacitor C_1 is α times the value of the main capacitance C_∞ [Fig. 4(a)]; at the beginning of the conversion phase, the reduction of the stored voltage with respect to V is given by:

$$\Delta V_i = \alpha V \left(e^{-\frac{t_s}{\tau_i}} - e^{-\frac{t_s + t_h}{\tau_i}} \right) \tag{12}$$

where $i = 1$ in this case and t_s and t_h are the sampling and hold times of the S/H system, respectively. More in general, ΔV_i in (12) is the voltage decrease due to the sole i-th R-C branch in Fig. 4(a). Each additional branch would add its own ΔV_i following (12). Hence, the total voltage loss is obtained just adding all the ΔV_i contributions.

Instead, in our model, since we modelled the DA current through (11), just simply integrating the current and assume that C_∞ dominates over the various C_i we can obtain the total voltage loss after the first hold phase, that is:

$$\Delta V = \frac{2V \tan\delta}{\pi} \ln\left(\frac{t_s + t_h}{t_s} \right) \tag{13}$$

As it can be appreciated, this new approach links ΔV directly to $\tan\delta$, thus avoiding the definition of a set of R-C branches and the addition of all their contributions as from (12) to obtain the total voltage loss ΔV in the capacitor.

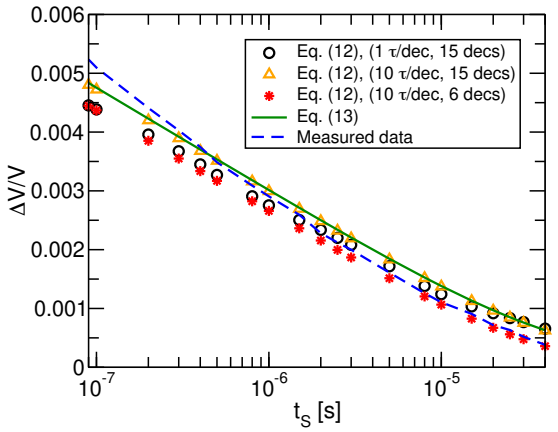


Fig. 5. Normalized voltage error (after a step with voltage V=5 V) vs. the sample time measured after t_h = 51.6μs. We fit the experiment with different models: the green solid line is (13) with $\tan\delta = 1.2 \cdot 10^{-3}$; symbols are simulations with the R-C model assuming different number of branches (see Fig. 4), i.e. different number of $\tau_i = R_i \cdot C_i$ distributed over different number of time decades. For 1 τ/dec, $\alpha = 1.76 \cdot 10^{-3}$, consistently with (16); for 10 τ/dec, $\alpha = 1.76 \cdot 10^{-4}$. Increasing the number of τ_i, the R-C model tends to (13). The coverage of more t_s decades improves the matching between the models. This is consistent with the hypothesis that (13) is the R-C model in the case of an infinite number of time constants, with τ_i distributed over the whole time range.

III. Simulations and Experiments

In this section, we first compare the predictions of (13) with those of the R-C model of (12). In the latter case, different R-C branches are considered, assuming τ_i values extending over different number of time decades and different numbers of τ_i per single decade (τ/dec). All R-C pairs share the same α value, which is related to the capacitance decrease per time decade.

Given the similarity of (12) and (13), it can be understood that α and $\tan\delta$ of the capacitor are linked together [5]. To obtain a quantitative relationship between α and $\tan\delta$, we see in (9) that the decrease of the capacitance ΔC with respect to C_∞ divided by the frequency decades where this decay takes place is equal to the term $\frac{2\tan\delta}{\pi}$, namely:

$$\frac{2\tan\delta}{\pi} = \frac{\Delta C}{C_\infty} \frac{1}{\ln(\frac{\omega_2}{\omega_1})} \tag{14}$$

Instead, since by the definition α is the relative capacitance decrease per frequency decade, we can write:

$$\alpha = \frac{\Delta C}{C_\infty} \frac{1}{\log_{10}(\frac{\omega_2}{\omega_1})} \tag{15}$$

The comparison of (14) and (15) readily gives the relation between α and $\tan\delta$:

$$\alpha = \frac{2\ln(10)\tan\delta}{\pi} \tag{16}$$

This is obtained when assuming that only one τ_i value is present in each single time decade (1 τ/dec). Otherwise the α value in (16) must be divided by the assumed number of τ/dec.

Figure 5 shows that the R-C model perfectly matches (13) if a span of 15 time decades is considered, with 10 τ_i values per each decade. Otherwise, small differences arise. This is

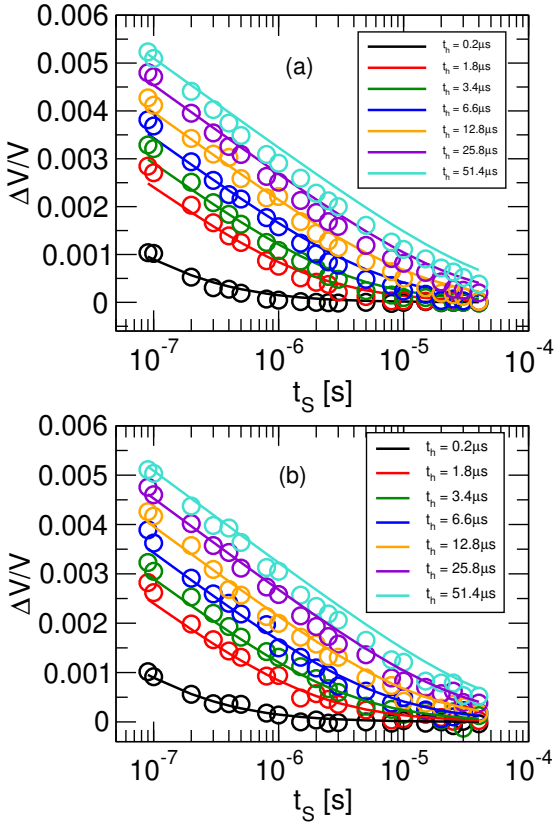


Fig. 6. Comparison between experimental data measured for different hold times (symbols) and (13) (lines). For these experiments, the input voltage are (a) $V = 5$ V and (b) $V = 3$ V. The fitting procedure provides $\tan\delta = 1.28 \cdot 10^{-3}$ which best fits the experiments. Equation (13) seems to precisely predict the normalized voltage error for all t_s and t_h.

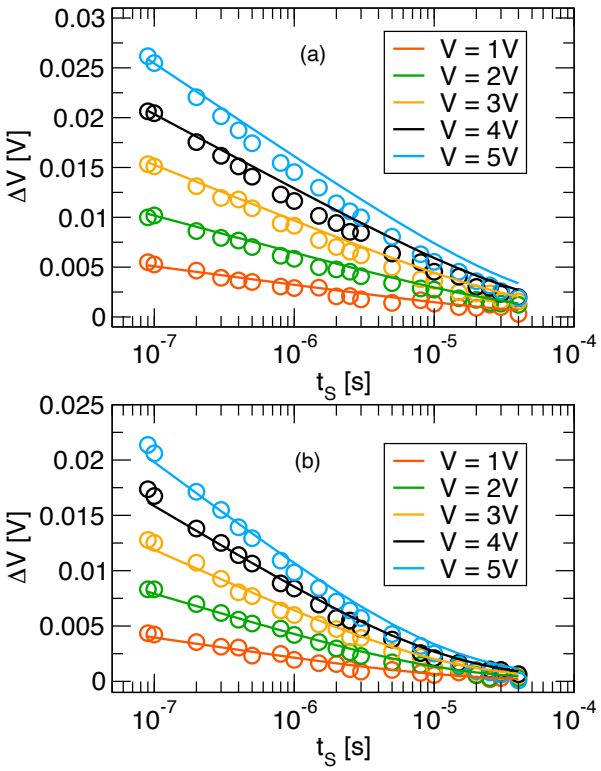


Fig. 7. Comparison between experimental data measured for (a) $t_h = 51.4$ μs and (b) $t_h = 12.8$ μs (symbols) and (13) (lines) for different input voltages. We use $\tan\delta = 1.28 \cdot 10^{-3}$, which is the same used in Fig. 6.

consistent with the assumption that (13) is the R-C model when considering an infinite number of τ_i.

To verify our model, we also performed experiments on a sample and hold (S/H) ADC featuring a sampling MIM capacitor fabricated in the BEOL. In particular, the characterized system is meant to measure the actual voltage stored in the capacitor after sampling for a time t_s and holding the data for a time t_h. The system allows to configure both t_s and t_h. The measured conversion error of the ADC is reported in Fig. 5 (dashed blue line) and one can note the good agreement between experiments and the two models when using $\tan\delta = 1.2 \cdot 10^{-3}$ and an α value computed with with (16).

Figure 6 extensively compare the predictions of (13) against experiments performed by varying the sample and hold times and for two different input voltages V. It is evident that our model, with very few parameters (in practice only $\tan\delta$ significantly affects the results), can nicely reproduce all the experiments. Figure 7 summarize all the performed measurements, showing that the conversion error scales linearly with the input voltage V and, thus, further validating (13), in which the absolute conversion error ΔV is indeed proportional to V.

The best fit of the whole set of experiments provides $\tan\delta = 1.28 \cdot 10^{-3}$, which is in good agreement with the $\tan\delta$ values extracted through fitting of experiments available in the literature [4], [5] (see Fig. 2), although the $\tan\delta$ depends on both the technology and layout used for the capacitor.

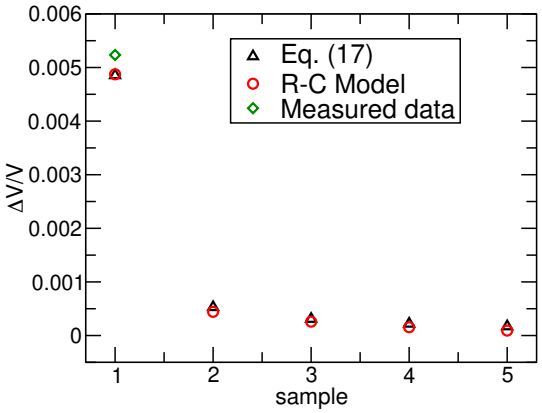


Fig. 8. Normalized voltage error predicted by (17) with $\tan\delta = 0.0012$ (black triangles) and the R-C model (red circles) for successive samples of the same input data $V = 5$ V when using t_s = 0.1 μs and t_h = 51.2 μs. For the R-C model we used $\alpha = 1.8 \cdot 10^{-3}$ and $\tau_1 = 0.1$ μs, $\tau_2 = 1$ μs, $\tau_3 = 10\mu$s and $\tau_4 = 100\mu$s (1 τ/dec). The measurement after the first sample (green diamond) is reported in figure as a reference.

The effects of dielectric absorption propagates also to the future sampling instants. Assuming that the ADC samples again the same V, it can be easily demonstrated that the error on the n-th sample for the model of (9) is given by

$$\Delta V^{(n)} = \frac{2V\tan\delta}{\pi} \ln\left(\frac{n(t_s+t_h)}{(n-1)(t_s+t_h)+t_s}\right) \qquad (17)$$

$\Delta V^{(n)}$ is the voltage reduction after the n-th sample-and-hold phases of the same V value.

Figure 8 compares the results of (17) with circuit simulations using the R-C model. At present time experimental data are available only for the first sample and this latter is reported in the figure as a reference. The simple equation (17) nicely reproduces the predictions of more complex circuit simulations with the discrete R-C model, thus demonstrating the usefulness of the proposed model.

IV. Conclusion

We developed a model to predict the errors induced by the dielectric absorption to the ADC conversion of a sampled voltage. The proposed equations are able to reproduce a wide set of experiments through the calibration of just a single parameter, namely $\tan\delta$. Although much simpler than the widely used R-C model, the developed analytical equations are also able to easily predict the errors induced on subsequent sampling.

This new model allows to directly link a single technological parameter ($\tan\delta$) to the DA impact on the ADC operation. In this way, the measurement of the capacitor characteristics allows the designer to directly estimate the DA effects on the ADC and to devise an adequate compensation procedure.

We demonstrated the application of this model to a simple ADC exploiting a single sampling capacitor. However, the model can be easily extended to different ADC architectures. In perspective, the compensation of the DA impact on the ADC may be also automatized after knowing just the $\tan\delta$ of the sampling capacitor and the topology of the ADC.

On the other way, we have also demonstrated that our model can allow the extraction of the $\tan\delta$ of sampling capacitors just by fitting the experimental conversion errors measured on an ADC.

Finally, the model can be easily extended also to predict the impact of dielectric absorption on other electronic applications.

References

[1] Kwan Chi Kao, “Electric Polarization and Relaxation”. *Dielectric Phenomena in Solids*, Academic Press, 2004, p. 41–114.

[2] H. Reisinger et al., ”A comparative study of dielectric relaxation losses in alternative dielectrics”, *International Electron Devices Meeting Technical Digest*, pp. 12.2.1-12.2.4, 2001.

[3] M. Kropfitsch, P. Riess, G. Knoblinger and D. Draxelmayr, ”Dielectric absorption of low-k materials: extraction, modelling and influence on SAR ADCs”, *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 4, 2006.

[4] T. Matsui, K. Sento, T. Ebata and A. Ishibashi, ”A Capacitor Dielectric Relaxation Effect Cancellation Circuit in a 12-Bit, 1-MSps, 5.0-V SAR ADC on a 28-nm Embedded Flash Memory Microcontroller”, *Proc. of IEEE 45th European Solid State Circuits Conference (ESSCIRC)*, pp. 95-98, 2019.

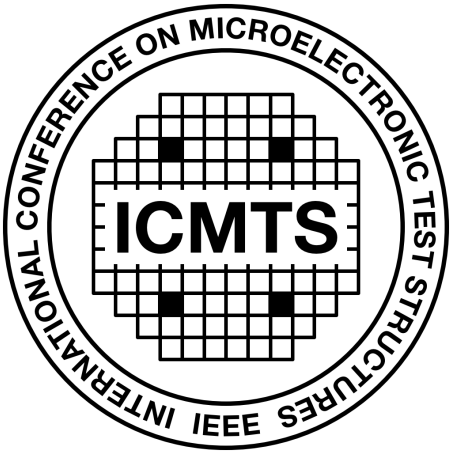
[5] K.-H. Allers, P. Brenner, M. Schrenk, ”Dielectric reliability and material properties of Al2O3 in metal insulator metal capacitors (MIMCAP) for RF bipolar technologies in comparison to SiO_2, SiN and Ta_2O_5.” *IEEE BCTM*, vol. 3, pp. 35-38, 2003.

[6] Djordjevic, A. R. and Biljie, R. M. and Likar-Smiljanic, V. D. and Sarkar, T. K., ”Wideband frequency-domain characterization of FR-4 and time-domain causality”, *IEEE Trans. on Electromagnetic Compatibility*, vol. 43, n. 4, pp. 662-667, 2001.

[7] Peng Zhao et al., ”Evaluation of border traps and interface traps in HfO2/MoS2 gate stacks by capacitance–voltage analysis”, *2D Materials*, vol. 5, p. 031002, 2018.

[8] K. Kundert, ”Modeling Dielectric Absorption in Capacitors”, *The designer’s guide community*, vol. 18, pp. 1-19, 2008.

SESSION 5
MEMs and Sensors



17th April 2024, 10:40–12:00

Session Co-Chairs: Johan Klootwijk, *Philips Research, Netherlands*
Yoshio Mita, *University of Tokyo, Japan*

Test structure for chemiluminescence measurement in aqueous solutions: initial design

A.A. Moreno-Guerrero, Camelia Dunare, Anthony J. Walton, Peter Lomax, Jonathan G. Terry, Ian Underwood
School of Engineering
The University of Edinburgh, United Kingdom
Email: alberto.moreno@ed.ac.uk

***Abstract*—There is an unmet need for on-chip integration of sensitive chemiluminescence sensors due to their extensive use in chemical analysis. This manuscript introduces a novel test structure for ease of opto-electrical integration in microfluidic systems. The MOS-based structure functions as an optical sensor for Luminol-based analysis and validates the suitability of the semiconductor fabrication process.**

Keywords—MOS tunnelling, Luminol, Photodetector, on-chip microfluidics, TCAD simulation.

I. Introduction

Laboratory analysis is fundamental in clinical diagnosis, food quality control, and environmental monitoring [1], [2]. For example, the concentration of medical markers or active substances in drugs is commonly assessed using different technologies such as liquid chromatography and nuclear magnetic resonance. These methods provide a high degree of selectivity and sensitivity but at the expense of reduced throughput and lack of portability [3]. These downsides have motivated the development of, for example, novel microfluidic-based techniques offering rapid detection and improved portability [4].

Most experimental microfluidic analytical systems still rely on external detectors to provide the measurand concentration [4], [5]. For example, optical detection systems integrating off-the-shelf components such as miniature silicon and conventional photomultiplier tubes (PMT) have been proposed [6], [7], [8]. Despite these systems having advanced significantly in portability and power consumption, on-chip integration still represents a more reliable approach due to advantages such as further miniaturisation, direct measurement, and integrated electronics. Kim et al. reported a portable small photodiode for HRP-catalysed luminol luminescence, addressing the high energy consumption of large systems such as PMT-based detectors [9]. Similarly, Ho et al. presented a miniature photosensor for HRP-catalysed luminol luminescence. That group aimed to replace the PMT in conventional optical chemiluminescence (CL) detection systems. They proposed a photodiode fabricated in a standard 0.35 µm CMOS process [10]. Each photosensor pixel was implemented with a P–N junction diode using a P substrate and an N+ active diffusion layer.

However, these proposed devices required ion implantation or diffusion processes to form the sensitive depletion region. This requirement is expensive and sometimes restrictive due to equipment availability. Moreover, their design follows the traditional structure with a vertical absorption trajectory relative to the incidence surface, resulting in the potential for reduced absorption [11]. This article introduces a novel test structure for ease of on-chip integration in microfluidic and gas sensing systems. It is based on a metal-oxide-semiconductor (MOS) architecture, offering photodetection capability without doping. The proposed device structure includes a trench that can be used as a reaction well and in which a catalyst can be deposited [11].

II. Test Structure

The test structure is based on the concept of the MOS capacitor with 500 nm aluminium contacts, 4.5 nm oxide and P-type silicon. Fig. 1 shows the cross-sectional diagram illustrating the structure profile. Silicon microfabrication techniques, such as photolithography, dry thermal oxidation, metal deposition, and wet and dry etching, were used for the realisation of the test structures. The 160 µm-wide trench, shown in Fig. 2, can be used as a microfluidic channel [12]. The MOS-based design provides several advantages, such as being easy to fabricate, useful as a photodetector, and not needing additional doping to ensure correct operation of its depletion region [13], [14].

III. Simulation

A. Device Structure

A 2D geometry of the device was used to simulate the electrical response. It was created using TCAD Sentaurus Structure, and an axis-aligned meshing was generated with Mesh [15]. The structure consisted of a p-type silicon ($3.0 \times 10^{14}/cm^3$) substrate, followed by a 4.5 nm silicon thermal oxide layer and 0.5 µm aluminium contacts. The software assumes a thickness of 1 µm, and the total current density is adjusted with a scaling area factor.

B. Simulation Models

Sentaurus Device was used to simulate the electrical and optical responses of the structure at room temperature.

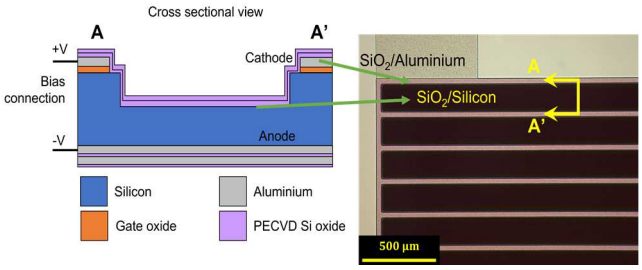


Fig. 1. (a) Schematic cross-section (not to scale) of the proposed test structure. (b) Photomicrograph of a device sample indicating some layers.

This project was supported by the Consejo Nacional De Ciencia y Tecnologia, Mexico Scholarship Number 495284, and by the EPSRC CDT in Intelligent Sensing and Measurement, Grant Number EP/L016753/1.

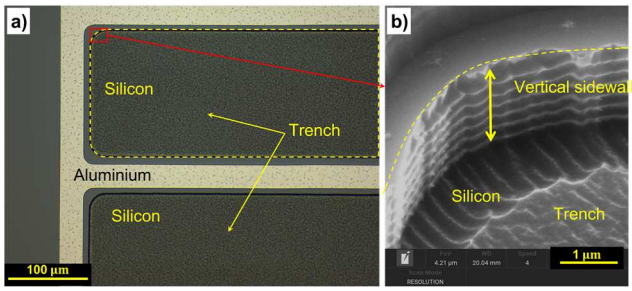


Fig. 2. (a) SEM photomicrograph of a feature after reactive ion etching (RIE). (b) A corner of a trench formed using RIE. Both samples were processed in a Plasmatherm Versaline IV.

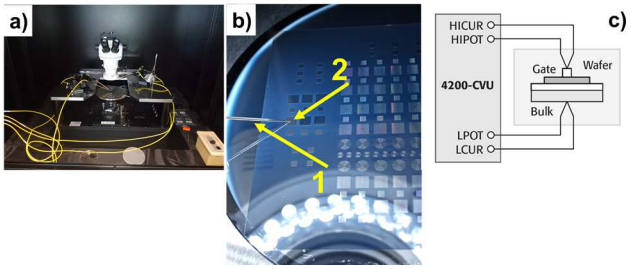


Fig. 3. (a) The EB-6 Station, (b) a probe needle (1) and a test structure (2) on a 3-inch wafer. (c) Diagram illustrating the basic electrical connection.

Doping-dependent band gap narrowing was included using the Jain-Roulston model [16], and carrier mobility was implemented using the unified Philips model, PhuMob. Carrier tunnelling was simulated using the direct tunnelling model for the silicon/oxide interface, based on the work developed by Schenk and Heiser [17].

The optical generation was computed using a monochromatic optical source, quantum yield as a function of the effective bandgap, and a wavelength-dependent complex refractive index [18]. The corresponding optical constants were used as given in the material parameter file in Sentaurus. Some simulation parameters are given in Table 1.

IV. Electrical Characterisation

The integration of characterisation structures on the silicon chip allows the verification of critical variables defining the device's electrical response. MOS structures are commonly used to evaluate the microfabrication process by measuring device parameters such as oxide thickness and quality, substrate doping, and others [19]. The test structures were electrically characterised with a Keithley Semiconductor Characterization System, model 4200-SCS, using voltage-current (V-I) and capacitance-voltage (C-V) measurements to validate some of their electrical and physical parameters. The structures were electrically connected with an Everbeing EB 6 Analytical Probe Station, as shown in Fig. 3. A 25 mV AC signal at 3 kHz was used to evaluate the capacitance.

The simulated and measured electrical responses are shown in Fig. 4. The measured capacitance-voltage values are plotted in Fig. 5. From this data, the background doping was calculated [20] and is shown in Fig. 6.

V. Optical Response

The MOS-based structure can be used as a photodetector [14]. When subjected to reverse bias, the oxide/silicon interface builds a region depleted of charge carriers [21]. Upon exposure to light, electron-hole pairs generated are driven to the electrodes by the action of the electric field [13]. These carriers then participate in the conduction and production of the photocurrent. TCAD simulation confirmed the structure's photodetection capability. A 1 μ^2 silicon sample was vertically irradiated with several power densities at 350 nm. The simulated photocurrent is shown in Fig. 7. Before using the test structure as a sensor, it was optically characterised using the setup shown in Fig. 8. Its linear response to 0.1 to 100 $\mu W/cm^2$ is shown in the inset in Fig. 9. The light source was a Thorlabs LED M470L3, 470 nm wavelength and 25 nm bandwidth. It was collimated with a collimating lens, and then calibrated with a Thorlabs PM100D power meter console and an S130VC photodiode power sensor. The photocurrent was measured with a Keithley 6485 picoammeter. The LEDs and sensors were powered with a Keysight E3647A DC power supply. Both instruments were connected to a computer via USB/GPIB interface using an

Table 1. Simulation Parameters and Models

Silicon at room temperature			
Parameter name	***electron, hole***	***Units***	***Ref.***
Eg_0, bandgap at 0 K	1.169	eV	[17]
α, material parameter	9.06×10^{-4}	eV/K	
β, material parameter	2.03×10^{3}	K	
Bandgap narrowing	Jain-Roulston	-	
m_{ox}, SiO_2 tunnelling mass	0.42, 0.5	m_0	
ε_{ox} Oxide dielectric constant	3.9	-	

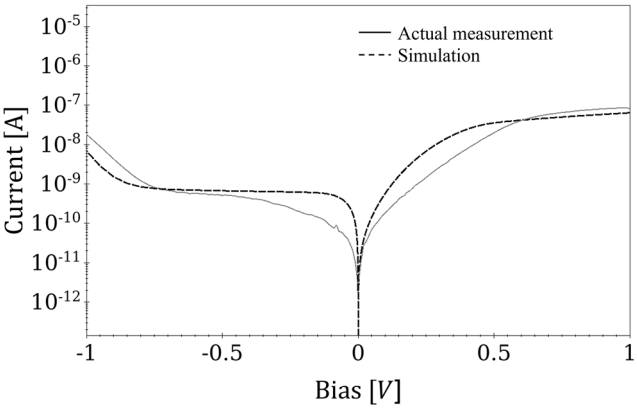


Fig. 4. Simulated and measured electrical responses of a MOS tunnelling structure.

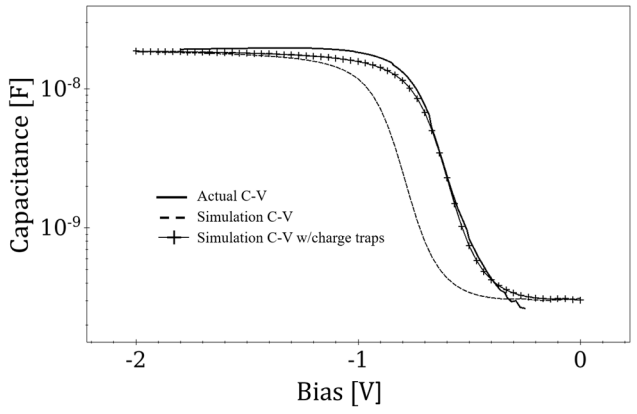


Fig. 5. Measured and simulated capacitance versus voltage.

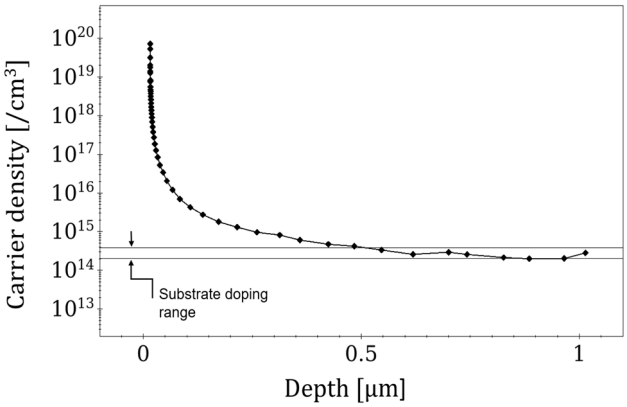


Fig. 6. Carrier concentration profile obtained with the C-V profiling technique.

Agilent 82357B cable connector. The optical response is shown in Fig. 9.

VI. Detection of Chemiluminescence

Chemiluminescence has been extensively used for chemical and biological analysis in gas and aqueous solutions [22], offering high sensitivity and very low detection limits [23]. Luminol is a CL precursor used due to its availability and cost [24]. The test structures presented here were used to measure CL in an aqueous solution. The first experiment compared the sensor response to CL against the response from a commercial PMT. The PR1405CE PMT was powered at −480 V with a Hamamatsu power supply model C3830.

Two solutions were prepared and mixed in a 1:1 proportion for the experiments. Solution A1 and A2 were made dissolving 50 mg of Luminol in 10 ml DI water and 50 ml of sodium hydroxide 0.25 and 0.5 molar, respectively. Solution B was prepared by dissolving 50 mg of potassium ferricyanide in 50 ml of DI water and 10 ml of hydrogen peroxide 0.16 molar. The sensor was placed in solution A1 and connected to the SCS. The PMT was placed beside the glass beaker as shown in Fig. 10. CL was produced by adding solution B. Another experiment was made using solutions A2 and B. The measured signals are shown in Fig. 11.

The second experiment consisted of measuring CL produced with different Luminol and catalyst concentrations using the setup shown in Fig. 12. Luminol was dissolved in 50 ml ammonium hydroxide 0.5 molar and then diluted in 10 ml

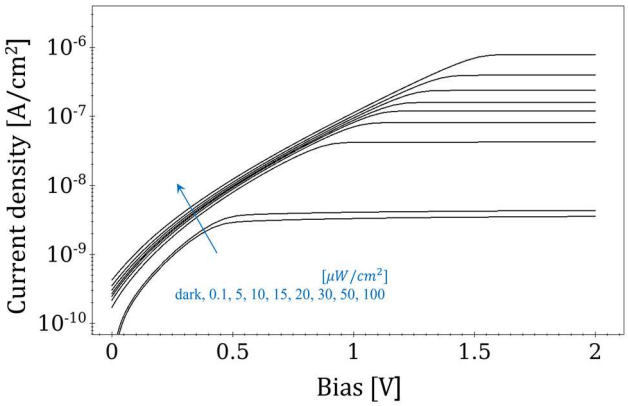


Fig. 7. Optical response of a MOS structure with λ = 350 nm.

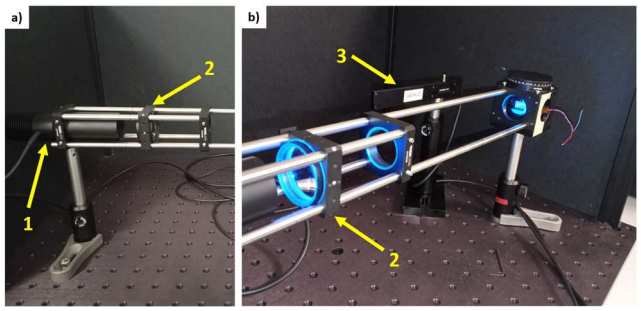


Fig. 8. Setup used for the device's optical characterisation. (a) The LED (1) and the collimating lens (2). (b) the S130VC photodiode.

of DI water to produce solution A. Solution B was obtained by dissolving potassium ferricyanide in 10 ml of hydrogen peroxide 0.65 molar and 50 ml of DI water. CL was produced after mixing 1:1 solutions A and B. The sensor was placed in solution A and connected to the probe station. The setup was placed inside a shield box HP 16058A to reduce electromagnetic interference. The solution B was manually added to the system with a syringe.

The functionality of the structures was tested in the Luminol-based chemiluminescent solution exhibiting 449 nm wavelength. The structures were capable of measuring different concentrations of Luminol and catalyst as shown in Fig. 13. The sensor's current remained practically unchanged when using the solution with 0 mg Luminol or catalyst. In both cases, the highest peak current occurred when using the highest concentration of both reactants. Its optical characterisation was conducted using a light source with the closest wavelength available in our laboratory at 470 nm. The sensor was calibrated, assuming the differences between the silicon optical constants were negligible. Under this caveat, the optical minimum detection and responsivity were approximated to 8 nW and 0.53 A/W, respectively.

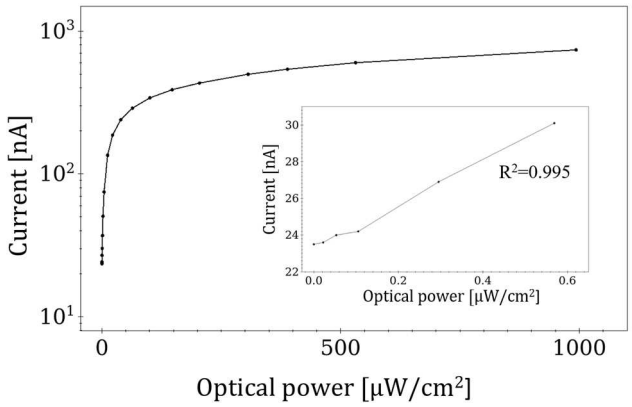


Fig. 9. Measured nominal optical response at 470 nm wavelength, biased at -0.3 V and different power densities. Inset: Optical response linearity at low power densities.

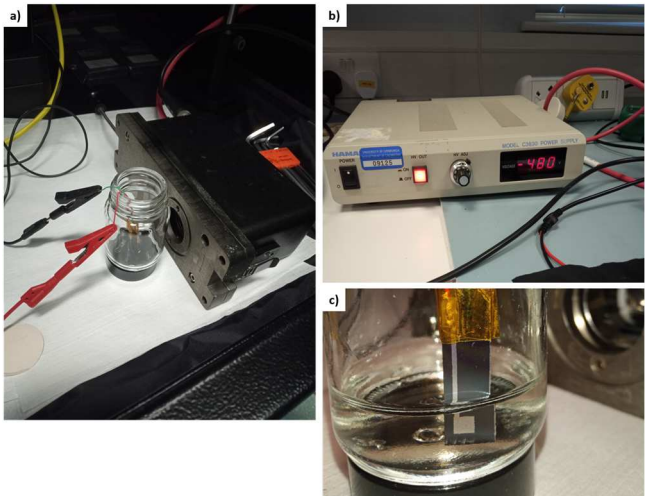


Fig. 10. Photographs of (a) the photomultiplier tube system setup inside the Everbeing shielding box, (b) the Hamamatsu C3830 power supply, and (c) sensor ID B-01 in solution A.

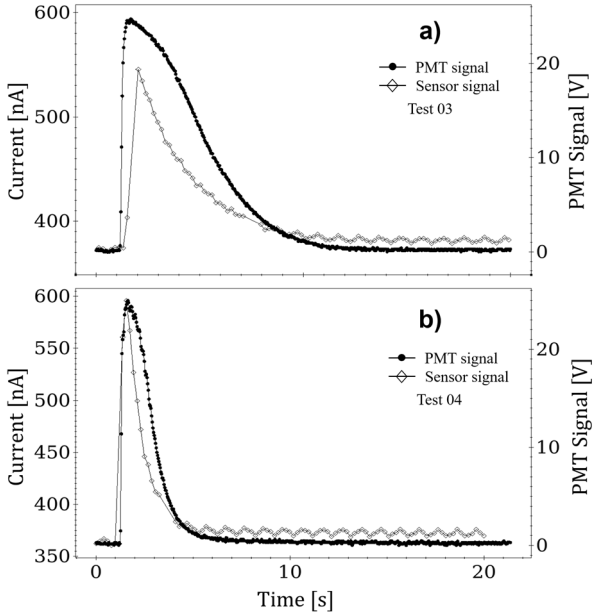


Fig. 11. Detection of Luminol CL with the PMT PR1405CE and the sensor ID 16. Chemical reactions using NaOH 0.25 molar (a) and 0.5 molar (b). In both cases, the sensor followed the CL signal. The sensor also indicated a less bright, but longer CL reaction when using the 0.25 molar base.

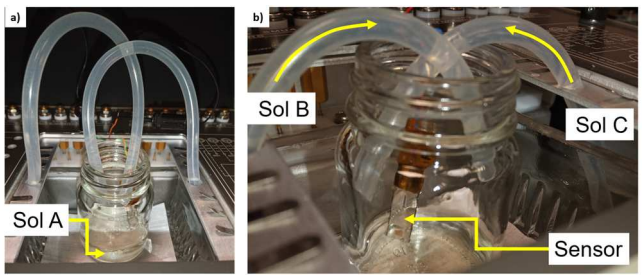


Fig. 12. Setup used for the experiment of the Luminol-based CL detection. Solution C was used in experiments not reported in this paper.

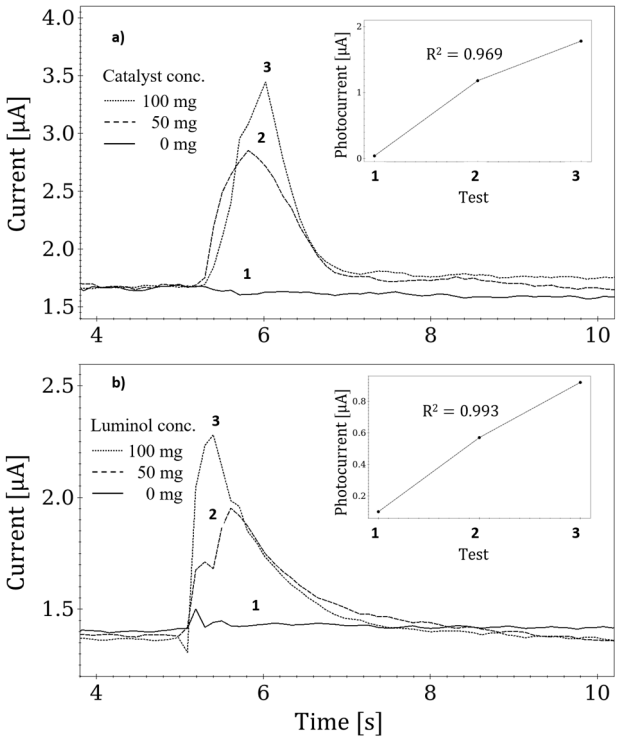


Fig. 13. Detection of chemiluminescence using different catalyst and Luminol concentrations, (a) sensor ID 4, and (b) sensor ID 5

The fabrication layers and the fabricated sensor are shown in Fig. 14.

VII. Summary

A novel test structure for chemiluminescence measurement in aqueous solutions has been proposed and characterised. Based on a MOS architecture, the structure aims to facilitate the on-chip integration of an optical measurement device for clinical diagnosis, food quality control, or environmental monitoring without requiring doping implantation or diffusion. Its optical responses at 350 nm and 470 nm were simulated, confirming its capability to measure two specific sources of chemiluminescence. The practicalities of its microfabrication were evaluated, suggesting that the feasibility of its realisation is uncompromised. The structure provides enhanced optical absorption, facilitated by the horizontal absorption paths created by the trenches' sidewalls. While this advantage has been primarily assessed through simulation, the initial characterisation reported here demonstrates that the structure competes favourably with current state-of-the-art sensors. It exhibits comparable linearity, rapid response, low power consumption, portability and sufficient sensitivity at a very low bias. The simulation and characterisation results and the proof-of-concept demonstration support the future development of the proposed test structures. However, it must first undergo optimisation work, such as improvement of the sidewall roughness and determination of the optimal depth of the trench to ensure maximum optical sensitivity.

Acknowledgment

The Authors wish to acknowledge Dr Andreas Tsiamis, Prof Anita Jones, Dr Andrew Alexander, Dr Ilka Schmueser and Prof Robert Henderson from the University of Edinburgh

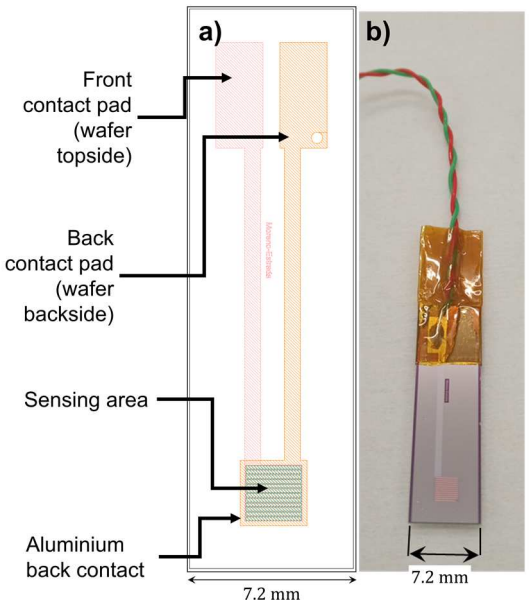


Fig. 14. (a) Klayout design of layers. Anode contact is shown as if the wafer were transparent for illustration purposes. (b) Finished microfabricated sensor.

for providing advice and testing equipment for the development of this work.

REFERENCES

[1] Y. Su, D. Deng, L. Zhang, H. Song, and Y. Lv, "Strategies in liquid-phase chemiluminescence and their applications in bioassay," *TrAC Trends in Analytical Chemistry*, vol. 82, pp. 394–411, Sep. 2016, doi: 10.1016/j.trac.2016.07.002.

[2] M. Iranifam, "Analytical applications of chemiluminescence methods for cancer detection and therapy," *TrAC Trends in Analytical Chemistry*, vol. 59, pp. 156–183, Jul. 2014, doi: 10.1016/j.trac.2014.03.010.

[3] J. F. Huertas-Pérez, D. Moreno-González, D. Airado-Rodríguez, F. J. Lara, and A. M. García-Campaña, "Advances in the application of chemiluminescence detection in liquid chromatography," *TrAC Trends in Analytical Chemistry*, vol. 75, pp. 35–48, Jan. 2016, doi: 10.1016/j.trac.2015.07.004.

[4] H. Gai, Y. Li, and E. S. Yeung, "Optical Detection Systems on Microfluidic Chips," in *Microfluidics: Technologies and Applications*, B. Lin, Ed., in Topics in Current Chemistry. , Berlin, Heidelberg: Springer, 2011, pp. 171–201. doi: 10.1007/128_2011_144.

[5] B. Kuswandi, Nuriman, J. Huskens, and W. Verboom, "Optical sensing systems for microfluidic devices: A review," *Analytica Chimica Acta*, vol. 601, no. 2, pp. 141–155, Oct. 2007, doi: 10.1016/j.aca.2007.08.046.

[6] C. Kang, F. Tang, Y. Liu, Y. Wu, and X. Wang, "A portable gas sensor based on cataluminescence," *Luminescence*, vol. 28, no. 3, pp. 313–317, 2013, doi: 10.1002/bio.2383.

[7] S. Xu *et al.*, "A novel chemiluminescence sensor for sensitive detection of cholesterol based on the peroxidase-like activity of copper nanoclusters," *Sci Rep*, vol. 6, no. 1, Art. no. 1, Dec. 2016, doi: 10.1038/srep39157.

[8] M. Zhang *et al.*, "Carbon dots-enhanced luminol chemiluminescence and its application to 2-methoxyestradiol determination," *Green Chemistry Letters and Reviews*, vol. 11, pp. 379–386, Sep. 2018, doi: 10.1080/17518253.2018.1511833.

[9] H. Kim, C.-G. Kang, and M. S. Kang, "Fabrication and Characterization of Silicon-Based Photodiode for Detection of Luminol Chemiluminescence in a Biosensor," Journal of the Korean Physical Society, vol. 42, pp. S336–S339, 2003.

[10] W.-J. Ho *et al.*, "Fabrication of a miniature CMOS-based optical biosensor," *Biosensors and Bioelectronics*, vol. 22, no. 12, pp. 3008–3013, Jun. 2007, doi: 10.1016/j.bios.2006.12.031.

[11] A. A. Moreno-Guerrero, C. Dunare, A. J. Walton, P. Lomax, J. G. Terry, and I. Underwood, "Detection of catallitically enhanced chemiluminescence with a novel photodetector for hydrogen sulfide sensing applications.," *Manuscript in preparation.*, 2024.

[12] S. K. Tiwari, S. Bhat, and K. K. Mahato, "Design and Fabrication of Low-cost Microfluidic Channel for Biomedical Application," *Sci Rep*, vol. 10, no. 1, Art. no. 1, Jun. 2020, doi: 10.1038/s41598-020-65995-x.

[13] C. W. Liu, W. T. Liu, M. H. Lee, W. S. Kuo, and B. C. Hsu, "A novel photodetector using MOS tunneling structures," *IEEE Electron Device Letters*, vol. 21, no. 6, pp. 307–309, Jun. 2000, doi: 10.1109/55.843159.

[14] J.-Y. Cheng, H.-T. Lu, and J.-G. Hwu, "Metal-oxide-semiconductor tunneling photodiodes with enhanced deep depletion at edge by high-k material," *Applied Physics Letters*, vol. 96, no. 23, p. 233506, Jun. 2010, doi: 10.1063/1.3449128.

[15] "TCAD." Accessed: Dec. 10, 2020. [Online]. Available: https://www.synopsys.com/silicon/tcad.html

[16] "Sentaurus Device User Guide." Synopsys, 2018.

[17] A. Schenk and G. Heiser, "Modeling and simulation of tunneling through ultra-thin gate dielectrics," *Journal of Applied Physics*, vol. 81, no. 12, pp. 7900–7908, Jun. 1997, doi: 10.1063/1.365364.

[18] M. Dresselhaus, G. Dresselhaus, S. B. Cronin, and A. Gomes Souza Filho, "Review of Fundamental Relations for Optical Phenomena," in *Solid State Properties*, in Graduate Texts in Physics., Berlin, Heidelberg: Springer Berlin Heidelberg, 2018, pp. 317–327. doi: 10.1007/978-3-662-55922-2_15.

[19] E. H. Nicollian and J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*. John Wiley & Sons, 1982.

[20] D. K. Schroder, "Carrier and Doping Density," in *Semiconductor Material and Device Characterization*, IEEE, 2006, pp. 61–125. doi: 10.1002/0471749095.ch2.

[21] S. M. Sze and K. K. Ng, "Metal-Insulator-Semiconductor Capacitors," in *Physics of Semiconductor Devices*, John Wiley & Sons, Ltd, 2006, pp. 197–240. doi: 10.1002/9780470068328.ch4.

[22] A. M. Garcia-Campana, W. R. G. Baeyens, and Z. Xingron, "Chemiluminescence-Based Analysis: An Introduction to Principles, Instrumentation, and Applications," in *Chemiluminescence in Analytical Chemistry*, 1st ed., CRC Press, 2001, p. 26.

[23] H. Y. Aboul-Enein, R.-I. Stefan, J. F. van Staden, X. R. Zhang, A. M. Garcia-Campana, and W. R. G. Baeyens, "Recent Developments and Applications of Chemiluminescence Sensors," *Critical Reviews in Analytical Chemistry*, vol. 30, no. 4, pp. 271–289, Jul. 2000, doi: 10.1080/10408340008984161.

[24] P. Khan *et al.*, "Luminol-Based Chemiluminescent Signals: Clinical and Non-clinical Application and Future Uses," *Appl Biochem Biotechnol*, vol. 173, no. 2, pp. 333–355, May 2014, doi: 10.1007/s12010-014-0850-1.

Droplet Impact Sensing with Low Noise Coplanar Reverse-Electrowetting Test Structures

Anotidaishe Moyo [a], Muhammad Wakil Shahzad [a], Stewart Smith [b], Jonathan Terry [b],
Yoshio Mita [c], Joseph Lewis [a], Yifan Li [a*]

[a] Department of Mechanical and Construction Engineering, Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne, NE1 8ST, UK
[b] Institute for Integrated Micro and Nano Systems, School of Engineering, The University of Edinburgh, Edinburgh, EH9 3FF, UK
[c] Department of Electrical Engineering and Information Systems, The University of Tokyo, Tokyo, Japan

***Abstract*—This paper examines the intricate dynamics of droplet impact on coplanar reverse electrowetting systems (CREW), specifically exploring how the electrode ratio and bias voltage interact to affect voltage generation. Employing a systematic experimental approach, we varied the electrode ratio and bias voltage to analyze their effects on the generated voltage during droplet impact. Our findings reveal significant dependencies between these parameters and the voltage output, shedding light on the underlying mechanisms governing energy harvesting efficiency in CREW systems. This research can inform better design strategies for improving energy conversion efficiency in various applications.**

I. Introduction

Within the field of energy harvesting, the primary emphasis has been on technologies such as triboelectric energy harvesters (TENG) [1]. However, the advent of the coplanar electrode configuration in reverse electrowetting (CREW) presents a fresh opportunity in this area. In contrast to conventional reverse electrowetting systems [2], the coplanar design of CREW provides notable benefits, especially in situations where traditional methods are two-plate configurations where the droplet is sandwiched between a conductive top plate and the bottom electrode plate [3]. The exposed top surface in CREW offers a promising pathway for raindrop energy harvesting, a field that holds great potential in renewable energy generation [4, 5].

The main purpose of this research is to use CREW test structures to investigate its potential as a highly effective rain droplet energy harvesting technique. Our goal is to examine the impact of the energy harvesting electrode ratio on energy generation. Additionally, our objective is to analyze the effect of voltage bias on energy harvesting, clarifying its contribution in amplifying voltage generation during droplet impact.

Most to date TENG type of rain drop energy harvesters employ multiple large droplets e.g. 100 µl [1] and uses costly characterization systems such as high impedance probes and nanocoulomb meter. Enabled by low noise set up, this research can employ relative low cost equipment and droplet sizes down to 5 µl, not only expands the range of energy harvesting methodologies but also paves the way for sustainable energy solutions in various applications, from microfluidics to renewable energy technologies. The CREW test structures shown in Figure 1 have been described previously in [4]. During the droplet impact characterisation, the droplet spreading and bouncing will need to be centered to the "flower pattern" to ensure the Energy harvesting Electrode (EHE) area ratio vs. counter electrode and gap areas remain the same. The study in

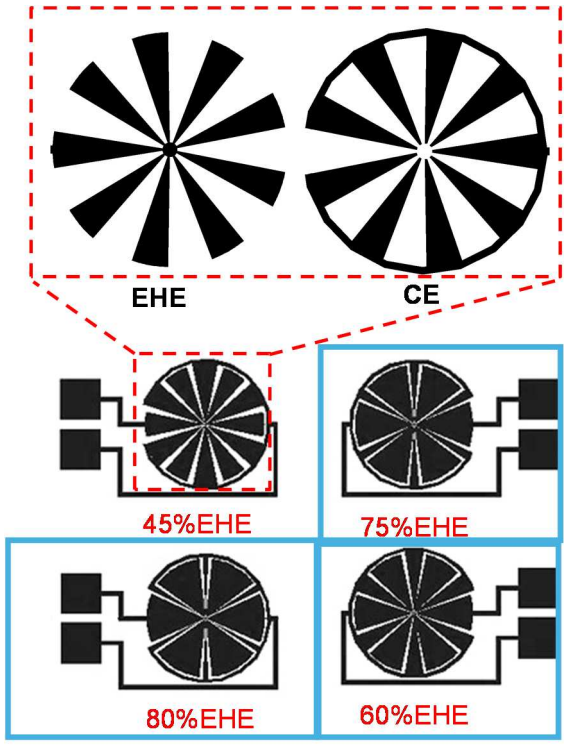


Fig. 1. Test structure design test structures with different energy harvesting electrode (EHE) area ratios (60% to 80% highlighted in blue), as well as different number of "petals" of 6 or 8.

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

this paper focuses on three different EHE area ratios: 60%, 75%, and 80%. Figure 2a shows the electrical circuit model of the set up. Figure 2b shows a typical "two-plates" reverse electrowetting setting in comparison to Figure 2c, a CREW configuration that suits for droplet impact study.

II. Experimental Setup and Methodology:

To examine the energy harvesting potential of coplanar reverse electrowetting (CREW) as a rain droplet energy harvester, a tailored experimental arrangement was created. The utilization of a 3D printed scaffolding structure was employed to provide support for a 22-gauge needle (0.718mm outer and 0.413mm inner diameters), which was then connected to a syringe containing DI (deionized) water. With a slow syringe motion, droplet sizes of 5 µl are constantly being produced. In order to maintain consistent impact conditions, the droplet height was consistently set at 80mm above the CREW setup, having a impact velocity of around 1.25m/s. The Weber number [1] is calculated by Equation 1 below

$$We = \frac{\rho v^2 L}{\sigma} = 0.0436 \tag{1}$$

Where ρ is the density of the fluid (kg/m³). v is the velocity of the droplet (m/s) calculated using a drop height of 80mm. L is the diameter of the droplet. σ is the surface tension of the fluid (N/m) approximately 0.072 N/m at room temperature.

The experiment involved testing three distinct ratios of energy harvesting electrode (EHE): 60% EHE, 75% EHE, and 80% EHE. Figure 1 shows of these electrode configurations. The purpose of these ratios is to demonstrate the differences in surface area coverage of the energy harvesting electrodes, allowing for an examination of their effects on energy generation efficiency.

The following equations denotes the parameters that govern energy generation in a REWOD setup:

$$E = \frac{5}{4} V^2 C_0 [1 - \tanh(\frac{1}{2}(1 - \log(\omega R C_0)))] \tag{2}$$

[2]

$$C_{cop} = \frac{A_d}{A_{total}} C_0 \tag{3}$$

Where V is the bias voltage, $\omega = 2\pi T^{-1}$ is the capacitance oscillation frequency, T is the characteristic period of one wetting-dewetting cycle, R is the load (i.e. input impedance of the oscilloscope), and $C_0$is the maximum REWOD capacitor when a droplet is at the maximum deformation (the maximum value of the changeable capacitor in Figure 2a).

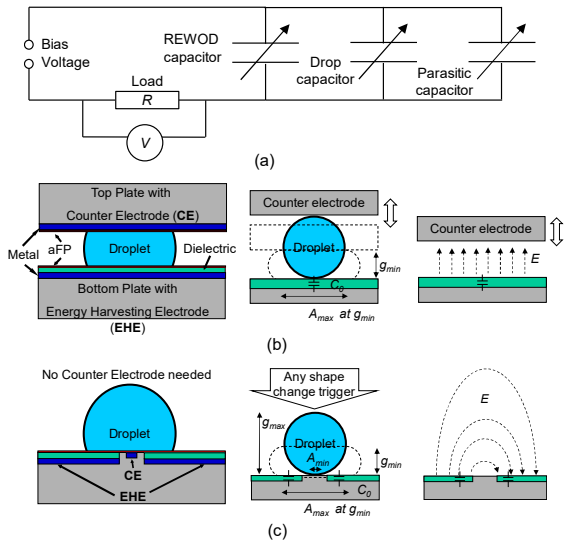


Figure 2. (a) Schematic of an REWOD energy harvesting circuitry; (b) Side view of common "two-plates" REWOD working mechanism (left), where counter electrode (CE) and energy harvesting electrode (EHE) are located on two parallel plates (middle), with electric field "shielded" in-between the two parallel plates; (c) Side view of c-planar REWOD configuration (left), where CE and EHE both situate on the bottom plate, leaving design flexibility on top structures (middle), the electric field is more exposed to parasitic elements

In EQ3, A_d/A_{total} is the energy harvesting electrode EHE area ratio to total areas (together with counter electrode CE + gap areas).

Droplets were dispensed onto the coplanar setup for each electrode ratio, with the voltage bias being systematically altered. The investigation involved three different voltage bias levels: 0V (no bias), 5V, and 10V. By utilizing this range of bias voltages, we were able to analyze how applied electric fields affect energy harvesting performance.

The voltage produced by the droplet impact on the CREW system was measured using an oscilloscope (Rigol 1054Z ™) during the experimentation. This enabled the recording and analysis of voltage outputs in real-time, corresponding to various electrode ratios and bias voltage conditions.

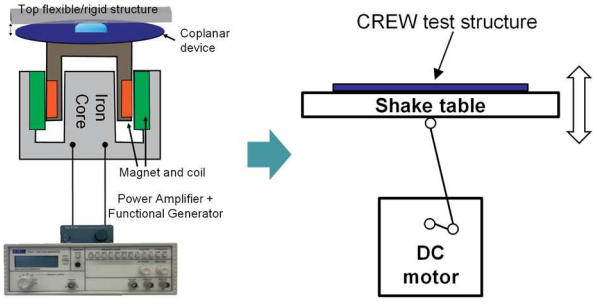


Figure 3. Schematic drawing showing the characterisation system set-up. DC motor and crank (right) used to replace the voice coil driven shake table for low-cost noise reduction.

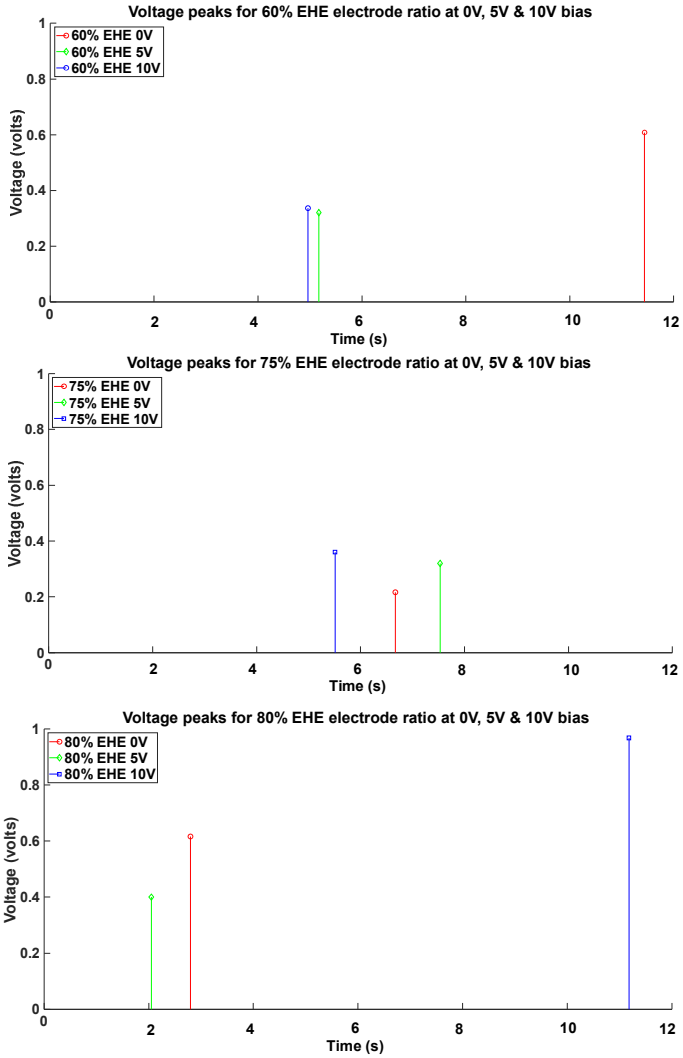


Figure 4. Peak Voltage Values for 60%, 75%, and 80% EHE Ratios at Various Bias Voltages (0V, 5V, and 10V) during Droplet Impact

III. Results and Discussion

Due to the CREW configuration more sensitive to parasitic elements and background noises, we have replaced the original characterisation system driven by RF power amplifier reported in [4] with a DC motor driven system as shown in Figure 3. The bias voltage was increased from 0V to 5V for the 75% EHE ratio, resulting in a peak voltage increase from 0.2160V to 0.3200V, indicating a significant 48% enhancement in voltage output. Increasing the bias voltage to 10V raised the peak voltage to 0.3600V, a modest increase of approximately 12.5%. Moderate bias voltages increase voltage generation. Higher biases may lead to diminishing returns due to increased charge and slower capacitance change (Figure 4).

Likewise, in the case of the 80% EHE ratio, an increase in bias voltage from 0V to 5V resulted in an initial decrease in peak voltage from 0.6160V to 0.4000V, indicating a significant reduction of around 35%. In spite of this, elevating the bias voltage to 10V caused a remarkable surge in peak voltage, reaching 0.9680V. There is a significant increase of about 142%. Bias voltage and voltage generation efficiency in CREW systems are intricately connected, as shown by this non-linear response.

In contrast, for the 60% EHE ratio utilizing a 6-blade electrode configuration, escalating the bias voltage from 0V to 5V resulted in a substantial decrease in peak voltage from 0.6080V to 0.3200V, marking a notable decrease of approximately 47%. Interestingly, further increasing the bias voltage to 10V led to a slight increase in peak voltage to 0.3360V, representing a marginal increase of approximately 5%. This divergent behavior suggests that the 60% EHE configuration exhibits a unique response to bias voltage variations compared to the other configurations.

Comparing the peak voltage output among the different electrode ratios, it's noteworthy that the 60% EHE configuration, despite having a 6-blade electrode design, showcased the highest peak voltage at 0V bias voltage. This suggests that the electrode blade number may have a significant influence on voltage generation efficiency during droplet impact.

Figure 5(top) visually shows droplet impact generated voltage signals on the oscilloscope. Three droplets caused three 40ms peaks on the CREW platform. Figure 5 (bottom) displays images captured with a high-speed camera, showcasing droplet impact, and bouncing dynamics.

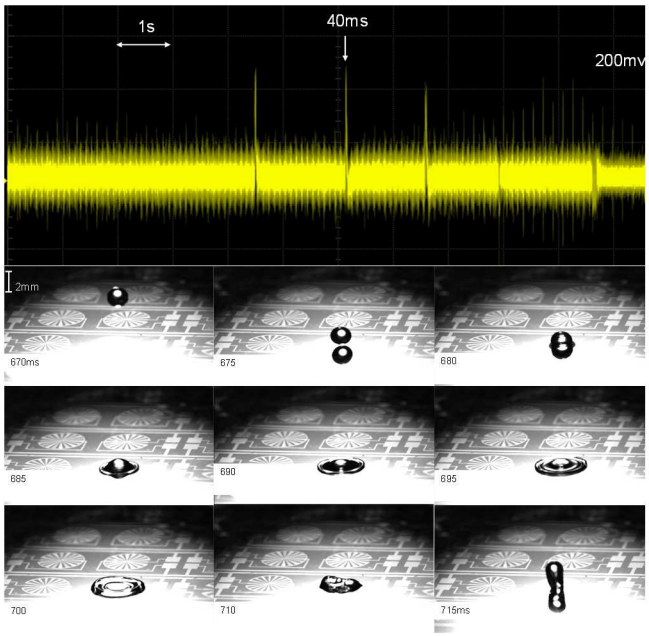


Figure 5: Oscilloscope waveform shows three distinct peaks from droplet impacts on CREW platform. Each peak in the waveform represents a 35 millisecond signal generated during droplet impact. High-speed camera captures droplet impact and bouncing stages lasting 35 ms from 680-715 ms.

Table 1 shows the peak voltage comparison of different test structures with design variations to comprehensively understand the dynamics of droplet impact and further improve the efficiency of CREW systems for sustainable energy generation.

Table 1 shows the Peak values for voltage output for each electrode ratio depicted in figure 4:

Electrode Ratio & Bias Voltage	Peak Voltage output (V)
60% EHE 0V	0.6080
60% EHE 5V	0.3200
60% EHE 10V	0.3360
75% EHE 0V	0.2160
75% EHE 5V	0.3200
75% EHE 10V	0.3600
80% EHE 0V	0.6160
80% EHE 5V	0.4000
80% EHE 10V	0.9680

IV. Conclusion

In this work, a low noise set up has enabled small droplet impact energy generation study using relatively low cost and simple equipment. The test structure characterisation of coplanar reverse electrowetting (CREW) systems has shed light on the influence of electrode ratio and bias voltage on voltage generation efficiency during droplet impact. Notably, while the 60% electrode ratio exhibited the highest droplet power generation, it's crucial to note that this observation may be attributed to the impact time being reduced more significantly with higher EHE ratio. This emphasizes the significance of electrode design in dictating power generation capabilities in CREW systems. Furthermore, our findings indicate a noticeable increase in power generation with increased electrode coverage for the 75% and 80% electrode ratios. This suggests that maximizing electrode surface area coverage, as seen in these configurations, plays a pivotal role in enhancing voltage generation efficiency during droplet impact. Our future research will focus on alternative flower petal electrode designs and their impact on power generation. We aim to explore higher voltage bias, tilted and sliding droplet scenarios, and varying droplet sizes. Despite the current low power generation per singular droplet, our study marks an encouraging step forward in demonstrating the potential viability of CREW raindrop energy harvesting. Through continued innovation and investigation, we anticipate significant advancements in the efficiency and practicality of CREW systems, contributing to the development of renewable energy technologies and addressing global energy challenges in the future.

Acknowledgment

We acknowledge Dr Camelia Dunare and Mr Stewart Ramsay, who fabricated the test structures in the cleanroom facilities of the Scottish Microelectronics Centre, at the University of Edinburgh. The authors would like to thank MEMSstar Ltd, who provided access to their aFP hydrophobic coating AURIX™ system. The authors would also like to thank the financial support from EPSRC ReNU Centre for Doctoral Training (CDT) fund (EP/S023836/1).

[1] W. Xu et al., "A droplet-based electricity generator with high instantaneous power density," Nature, vol. 578, no. 7795, pp. 392-396, 2020/02/01 2020, doi: 10.1038/s41586-020-1985-6.

[2] T. Krupenkin and J. A. Taylor, "Reverse electrowetting as a new approach to high-power energy harvesting," Nature Communications, vol. 2, no. 1, p. 448, 2011/08/23 2011, doi: 10.1038/ncomms1454.

[3] P. R. Adhikari, M. N. Islam, Y. Jiang, R. C. Reid, and I. Mahbub, "Reverse Electrowetting-on-Dielectric Energy Harvesting Using 3-D Printed Flexible Electrodes for Self-Powered Wearable Sensors," IEEE Sensors Letters, vol. 6, no. 5, pp. 1-4, 2022, doi: 10.1109/LSENS.2022.3170207.

[4] A. Moyo, M. W. Shahzad, J. Terry, Y. Mita, and Y. Li, "A facile coplanar reverse electrowetting-on-dielectric configuration for more flexible and integratable force/motion sensing applications," in 2022 IEEE Sensors, 30 Oct.-2 Nov. 2022 2022, pp. 1-5, doi: 10.1109/SENSORS52175.2022.9967283.

[5] A. Moyo, M. W. Shahzad, J. G. Terry, S. Smith, Y. Mita, and Y. Li, "Test Structures for Studying Coplanar Reverse- Electrowetting for Vibration Sensing and Energy Harvesting," in 2023 35th International Conference on Microelectronic Test Structure (ICMTS), 27-30 March 2023 2023, pp. 1-4, doi: 10.1109/ICMTS55420.2023.10094057.

Analysis methodology of Deep Trench Isolation Field-Effect Passivation techniques for Image Sensors through dedicated test structures

Samuel Tlemsani[1,2], Stéphane Ricq[1], Olivier Marcelot[2], Pierre Magnan[2]
[1] STMicroelectronics, Crolles, France
[2] Institut Supérieur de l'Aéronautique et de l'Espace (ISAE), Toulouse, France
Email address: samuel.tlemsani@isae-supaero.fr

***Abstract*— Passivation techniques are mandatory to reduce the CMOS Image Sensors dark current from interfaces. In this work a general analysis methodology is proposed to study the Deep Trench Isolation field-effect passivation thanks to dedicated test structures. Two deep trench isolations are characterized: an electrically active trench and an innovative passive trench using charged Al_2O_3/SiO_2 stacks. The results, validated by TCAD simulations, show that the best passivation is achieved here with the passive trench. Beyond, our approach is demonstrated to be a powerful tool for technologies and pixels developments with a fair passivation efficiency comparison through accumulation or inversion layer carrier concentrations. This work is also a first step toward dark current contribution modeling.**

I. Introduction

Deep Trench Isolations (DTI) have mainly emerged to improve the performance of CMOS Image Sensor (CIS) by targeting optical and electrical crosstalk reduction [1]. DTI has easily outperformed Shallow Trench Isolation (STI) on this point but has at first faced Dark Current (DC) issues by increasing the SiO_2/Si interface area. However, DC contribution from DTI interfaces can be reduced through passivation solutions [1][2].

There are different ways to passivate surfaces which can be sorted into two families. On one side the chemical passivation aims to reduce the interface trap density induced by the overall DTI process [2] thanks to hydrogen diffusion which fill silicon dangling bonds [3]. On the other side the field-effect passivation decreases the surface generation by increasing a charge carrier density at the interface thanks to an electric-field [4]. A common last option is to increase the carrier density by localized implantation [1].

The analysis and control of the passivation quality of DTI structures is crucial for CIS performances. Therefore, this study focuses on field effect passivation, on two DTI architectures, an electrically active and an innovative passive trench, represented on **Fig 1.** The DTI passivation is characterized from capacitance measurements performed on test structures, providing helpful insights for technologies and pixels developments decisions and DC modelling in general.

II. DTI Field-Effect Passivation

DTI interface states are at the origin of the generation of electron-hole pairs that contribute to the DC. These contributions can be expressed with the Shockley-Read-Hall (SRH) formalism at an interface thanks to the surface generation/recombination rate U_s ($cm^{-2}.s^{-1}$) [5]. Considering only the mid-band traps, U_s can be expressed as:

$$U_s = \frac{\sigma_n \sigma_p v_{th} N_{it} (n_s p_s - n_i^2)}{\sigma_n (n_s + n_i) + \sigma_p (p_s + n_i)} \quad (1)$$

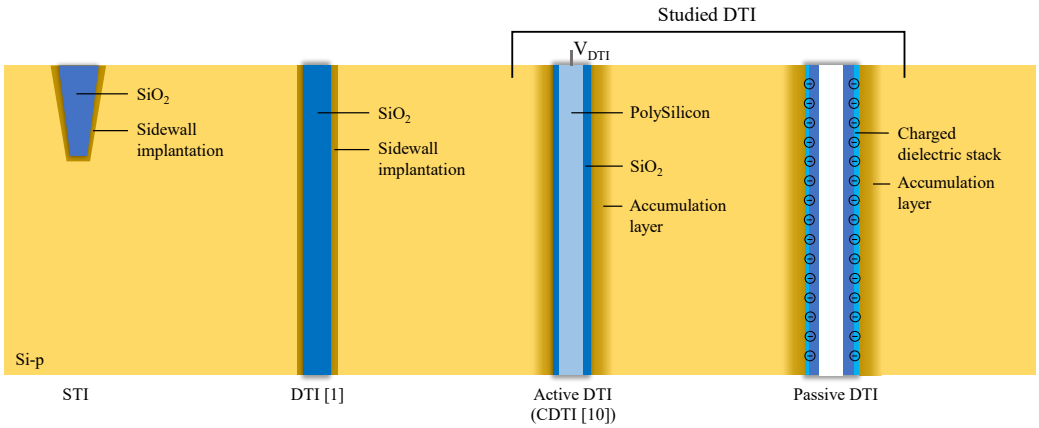


Fig. 1. STI and DTI Diagram. STI and DTI are commonly passivated by sidewall implantation. The active and the passive DTI studied in this work are passivated by an accumulation layer thanks to the field-effect passivation.

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

with N_{it} the interface trap density (cm^{-2}), $\sigma_{n,p}$ the surface capture cross sections for electron and hole (cm^2), v_{th} the thermal velocity (cm.s^{-1}), n_s and p_s the electron and hole densities (cm^{-3}) at the surface, and n_i the silicon intrinsic carrier density (cm^{-3}). When $U_s > 0$, interface traps recombine hole electron pairs and when $U_s < 0$, they generate pairs. An interesting way to study the DTI passivation is to investigate the surface generation velocity $S_g = -U_s/n_i$ (cm.s^{-1}) [6], defined when $n_s p_s << n_i^2$ which relates the dark current generation of an interface. **Fig. 2** shows with a computation from (1) that increasing the surface carrier concentration p_s decreases the surface generation velocity S_g, which is the objective of the field-effect passivation.

There are mainly two ways to achieve field-effect passivation. The first one consists in using charges included in dielectric stacks, here a DTI coupled with a Al_2O_3/SiO_2 stack known to be negatively charged [7][8] inducing a positive charge accumulation at the interface of the DTI with the p-silicon. **Fig.3** shows a computation illustrating the dependency of the surface potential ϕ_s and the surface carrier density p_s with effective oxide charges density N_{eff} (cm^{-2}) by using the following equations:

$$p_s = p_0 \, exp\left(\frac{-q\phi_s}{kT}\right) \qquad (2)$$

with p_s and p_0 the hole densities (cm^{-3}) respectively at the surface and in the bulk at the equilibrium, and ϕ_s the surface potential (V). From the charge neutrality, in the accumulation regime ($\phi_s < 0$) in a p-Silicon ($p_0 >> n_0$), it can be shown that [9]:

$$\phi_s \approx -2\frac{kT}{q} ln\left(\frac{-N_{eff} L_D q^2}{\sqrt{2}\epsilon_s kT}\right) \qquad (3)$$

with ϵ_s the silicon permittivity (F.cm^{-1}), L_D the extrinsic Debye length for holes (cm) and $N_{eff} < 0$.

The second way to achieve a field effect passivation is to use an active device based on a Capacitive Deep Trench Isolation (CDTI) architecture [10], in which a potential is applied to the core of the DTI, constituting a Metal Oxide Silicon (MOS) capacitor with capability to accumulate carriers at the silicon interface [9]. In this case the DTI is filled with doped polysilicon as conductor and the SiO_2 is used as the dielectric stack. By controlling the interface charge carrier density with the applied

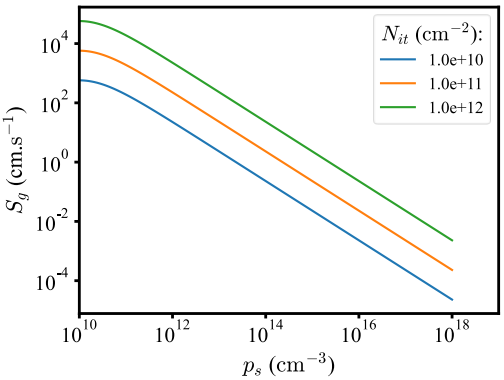


Fig. 2. Surface generation rate S_g as a function of the surface carrier concentration p_s, for different values of interface state density N_{it} and with $\sigma_n = \sigma_p$=1e-14 cm^2

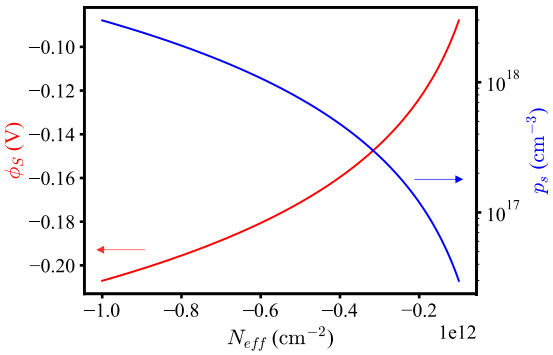


Fig. 3. Effective oxide charge density N_{eff} dependance of the surface potential ϕ_s and surface carrier density p_s for the passive DTI case.

voltage, the surface generation velocity can be reduced as shown in **Fig. 2**.

The characterization approach developed here on these two DTI architectures enables us to study and benchmark the field effect passivation efficiency whether it is driven by applied voltage or by charges in dielectric.

III. Test Structures And Experimental Settings

Device level works on test structures can provide useful insights on passivation and then can contribute efficiently to technology and pixel developments but also to process monitoring or to reliability studies. **Fig. 4b** presents the layout of the test structure designed to characterize the DTI with charged Al_2O_3/SiO_2 stack, where high contact nodes are arranged outside and the low inside a matrix of DTI square. This device has been studied as a Silicon Oxide Silicon (SOS) capacitor as shown on **Fig. 4a**. For the CDTI, **Fig. 5** shows a 2D cross-section of the characterized test structure. It's a device derived from the Photogate pixel [11] where the CDTI, the PWELL layer, the Vertical Gate (VEGA) and the sense node can be addressed. This CDTI has been treated as MOS capacitor with the CDTI contact node as the high contact, the sense node and the PWELL layer as the low contact nodes. The CDTI test structure reproduces the exact pixel array topology to be fully representative. It is not yet the case for passive DTI, but it is just a matter of test structure availability at that time.

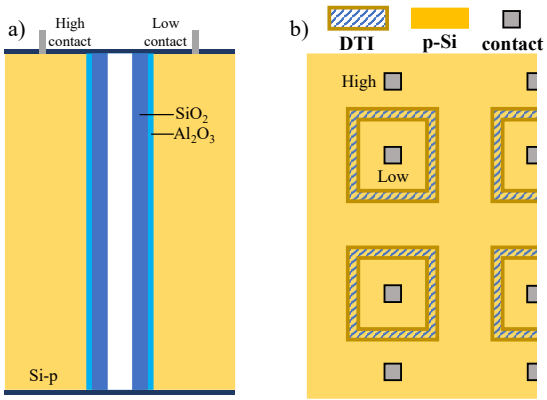


Fig. 4. a) 2D cross-section diagram of the DTI with Al_2O_3/SiO_2 stacks (Passive DTI). b) layout diagram of the test structure.

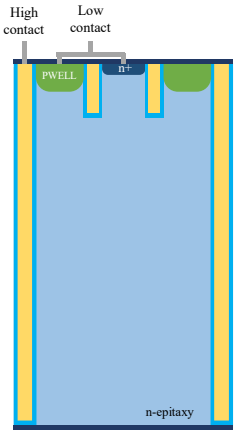


Fig. 5. 2D cross-section diagram of CDTI test structure Photogate pixel version (active DTI).

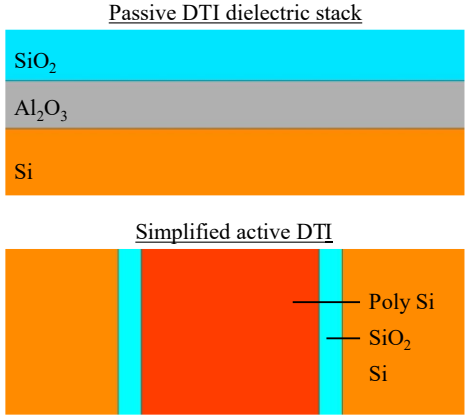


Fig. 6. 2D TCAD structures used as simplified experimental test structures.

The following characterizations have been performed with an impedance meter (Agilent E4980A), at wafer level and at ambient temperature. The objective is to extract from simple Capacitance Voltage (CV) measurements, parameters of interest for the two DTI architectures: the effective oxides charges density N_{eff} for the passive DTI with charged stack (and then compute p_s), and the voltage dependence of charge carrier densities n_s and p_s at the interface for the CDTI. These results are compared with simulations performed on the Synopsys Sentaurus technology computer-aided design (TCAD) software. Test structures and electrical characterizations are simulated with the extracted parameters from the measurements, in 2D and with the simplified layers as shown in **Fig.6**: a unique CDTI with majority and minority carrier available, and an Al_2O_3/SiO_2 liner in a silicon epitaxy.

IV. Parameters Extraction

A. *Passive DTI – Silicon Oxide Silicon capacitors*

To achieve the N_{eff} extraction from the measurement of the charged stacks, the DTI is used as a SOS capacitor. Indeed, the filling material has a very high conductivity, and we can consider in a first approximation the DTI as a capacitor with the two dielectric liner capacitors in series. **Fig. 7a** presents an experimental CV-measured on a SOS capacitor. One can observe a plateau from V_{high}= -12V to +12V, confirming that thanks to the negative charges in the dielectric, holes are accumulated in the silicon along interfaces. **Fig. 7b,c,d** illustrate with band-diagram the different regimes encountered: the accumulation, flat band and depletion. To compute N_{eff}, the Maserjian's function [12] is used to extract the flat band voltage V_{FB} from the measured CV (**Fig.8**) together with the relation:

$$V_{FB} = V_{FB_0} - \frac{qN_{eff}}{C_{ox}} \quad (4)$$

with V_{FB_0} the internal potential of the device (V), being null for a SOS capacitor since the high and low nodes are based on the same doping concentration in the silicon, and C_{ox} the oxide capacitance (F.cm^{-2}). Test structures have been characterized on full wafer (64 dies) and the extracted N_{eff} values are found in good agreement with the literature for this type of stack: [-3.0, -3.9] $\times 10^{12}$ cm^{-2} [8]. However, the presence of hysteresis phenomena can also be noticed (**Fig. 9**), which are linked to charge trapping on Al_2O_3 during the measurement [8]. This point is not further investigated here and contributes to the uncertainty surrounding the extractions. Nevertheless, the N_{eff} values obtained remain relevant and the methodology offers a powerful tool for passivation analysis and allows for the surface density p_s computation with (2) and (3). The extracted N_{eff} values are injected in the TCAD simulation of the Al_2O_3/SiO_2 stack, and it comes that the simulated density of accumulated charge carrier in silicon is similar to the one computed from measurement (see **Table. I**), thus validating the methodology.

B. *Active DTI – Metal Oxide Silicon capacitors*

The CDTI test structures are operated as a MOS capacitor. An example of experimental CV measurement is presented in **Fig. 10**. In this pixel with a N-epitaxy, passivation is achieved by the minority carriers provided by the PWELL layer. The effective oxide charge carrier concentration can be extracted in the same way as with SOS capacitors, with the difference that the minority carrier's density of the hole inversion layer has to be extracted according to the voltage applied to the core polysilicon. The Berglund's formula computes the surface potential for a given applied voltage V_g using the measured capacitance $C(V)$ and the oxide capacitance C_{ox} [13]:

TABLE I. Results Summary

Test structures	N_{epi} (cm^{-3})	V_{FB} (V)	N_{eff} (cm^{-2})	p_s (cm^{-3})	p_s TCAD (cm^{-3})
Passive-DTI	1.6×10^{16}	[12.0, 13.3]	[-3.0, -3.8] $\times 10^{12}$	[2.7, 4.3] $\times 10^{19}$	[2.4, 3.6] $\times 10^{19}$
CDTI	1.2×10^{16}	0.65 V	2.6×10^{11}	(1.3|2.0|4.3) $\times 10^{18}$ at (-1|-1.2|-1.8) V	(1.7|2.6|5.9) $\times 10^{18}$ at (-1|-1.2|-1.8) V

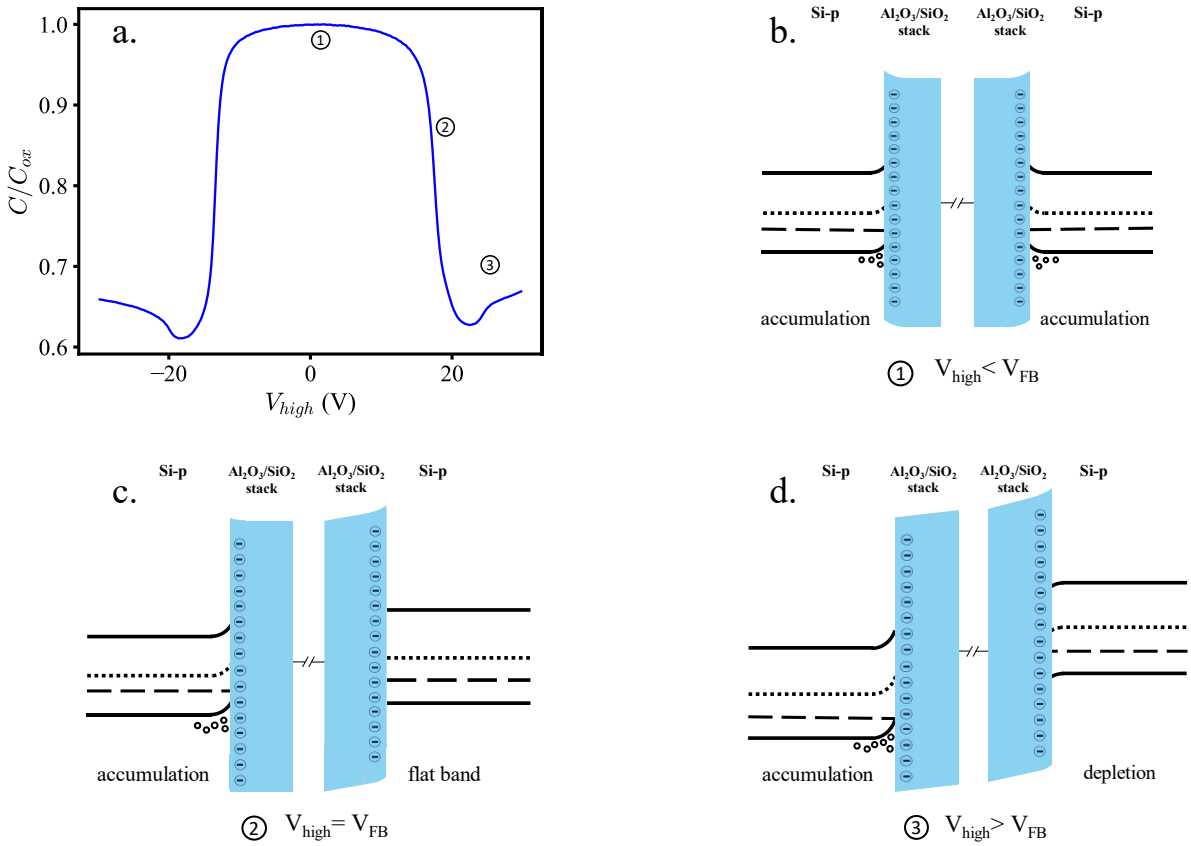


Fig. 7. a) Experimental CV-measure of the SOS capacitor, b) illustrative band-diagram of the accumulation regime, c) flat band regime, d) depletion regime

$$\phi_s(V_g) = \int_{V_{FB}}^{V_g} (1 - C(V)/C_{ox})dV \qquad (5)$$

However, in the CDTI case, the poly depletion effect [14] observed on **Fig. 10** in the inversion regime for V_{CDTI} = - 4V to -1V, leads to an error when using (5) since the term $1 - C(V)/C_{ox}$ no longer converges to 0, leading to the divergence of ϕ_s for the negative applied voltages. This error can be avoided by considering this effect in the computation. It can be shown by following the methodology of [13], that (5) becomes:

$$\phi_s(V_g) = \int_{V_{FB}}^{V_g} \left(\frac{1+C(V)/C_{poly}(V)}{1+C_{Si}(V)/C_{ox}} \right) dV \qquad (6)$$

with $C_{poly}(V_g)$ and $C_{Si}(V_g)$ respectively the polysilicon and the silicon depletion capacitance. To dissociate and approximate each contribution from the measured $C(V)$, the $1/C^2$ curve can be used (**Fig. 11**). Indeed, in the depletion regime, the $1/C^2$ curve is linear [12] thus allowing the identification of the C_{poly} contribution and then the deduction of C_{Si} contribution (C_{ox} being known with the measured capacitance in the strong

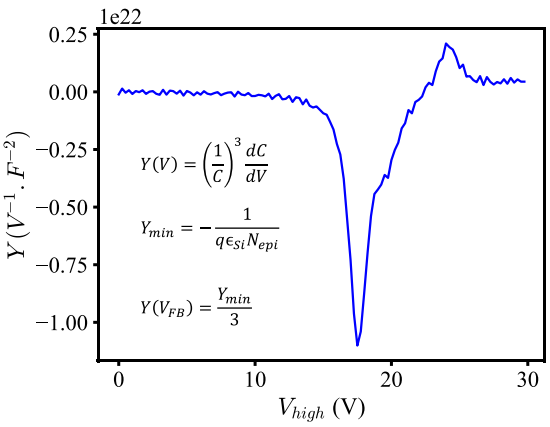


Fig. 8. Maserjian's Y function obtained from the experimental CV measurement, used to extract the epitaxy doping N_{epi} and the flat band voltage V_{FB}.

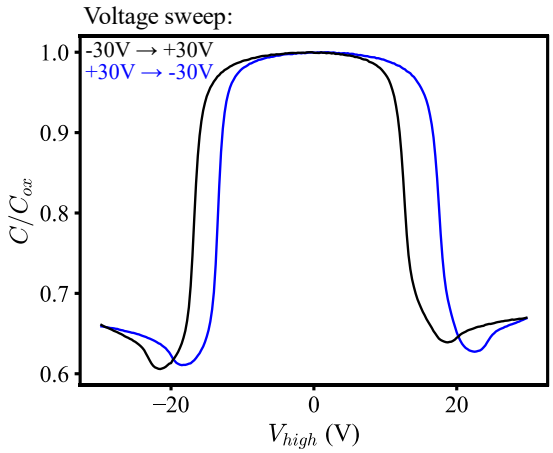


Fig. 9. Charge trapping on Al_2O_3 stack leading to hysteresis phenomena on experimental CV-measure.

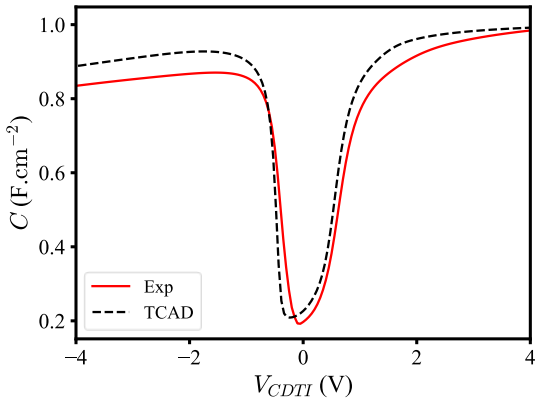


Fig. 10. Experimental and simulated CV curve for the CDTI test structure.

accumulation regime). The flat band voltage V_{FB} and the epitaxy doping N_D are extracted with the Maserjian's function. The poly silicon doping N_{poly} is obtained with the slope of the $1/C^2$ curve in the poly silicon depletion regime. These parameters are used to compute (6),(2) and to calibrate TCAD simulations to compare results and validate the methodology. **Fig. 12** presents the surface potential ϕ_s and the surface carrier density p_s as a function of the CDTI voltage obtained with TCAD simulation (solving Poisson's equation in the 2D structure) and those calculated with (6) and (2) based on the measured CV curve. Results are consistent in the depletion and inversion regime, but a gap is visible in the accumulation regime. The root cause of this deviation might be an additional parasitic capacitance in the measured device. This is still under investigation. The resulting TCAD CV is compared with the experimental one in **Fig. 10**, and it shows that the experimental capacitance in the accumulation regime decreases faster when reaching the depletion regime, to finally achieve to a constant offset in the inversion regime because of the suspected parasitic capacitance. This phenomenon is not considered in (6), leading to a gap in the ϕ_s-V curve. Thus, this gap can be observed in the same way on the p_s-V curves, obtained with TCAD simulation and with (2). The ϕ_s and p_s extraction methodology itself is not involved since we retrieve the same results between extraction from simulated

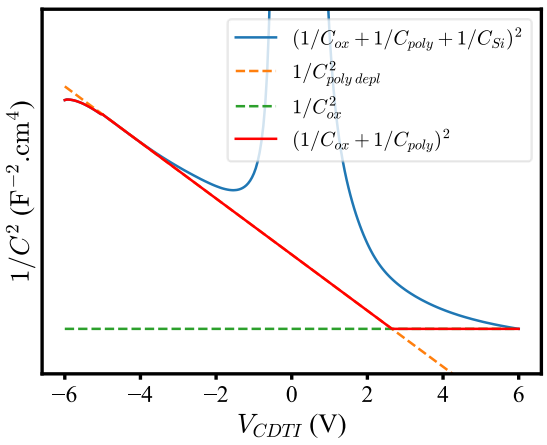


Fig. 11. $1/C^2$ curve allowing the estimation and dissociation of the Silicon and PolySilicon contribution on the measured CV curve for the CDTI.

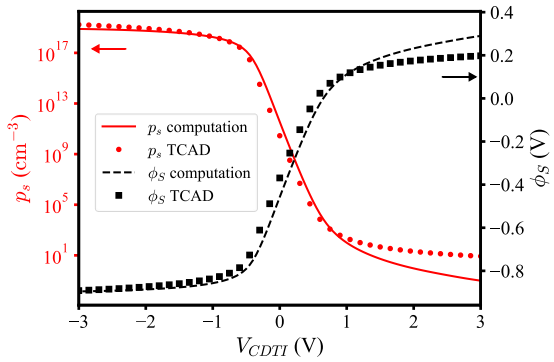


Fig. 12. Computed surface potential ϕ_s and surface carrier density p_s from the experimental data, and TCAD results for the CDTI.

CV or directly from TCAD simulation as shown in **Fig. 13.** Finally, no such shift is noted in the inversion regime, which is the operating range of interest, allowing a precise analysis of the FE passivation for the CDTI (**Table. I**).

V. Conclusion And Perspectives

Dedicated test structures characterized by CV measurement have been demonstrated to be a powerful analysis tool to study the DTI Field-Effect passivation, both on a passive DTI (driven with charged stacks Al_2O_3/SiO_2) and on an active CDTI (driven with the applied voltage). By extracting the flat-band voltage, the effective oxide carrier density can be estimated and then the surface hole carrier concentration is computed for the passive DTI. The surface potential voltage dependency measurement allows for the same computation for the active DTI in the depletion and inversion regime even if a suspected parasitic capacitance still introduces an error in accumulation regime, but TCAD simulations have allowed for the validation of the proposed methodology independently. **Table. I** summarizes for both DTI the extracted parameters obtained with experiment results, the computed surface carrier densities p_s with (2), and those obtained with TCAD simulation thanks to the Poisson's equation solving. It appears that with these test structures and these DTI architectures, the best FE passivation is achieved with the passive DTI as the charge density at the interface is ten times

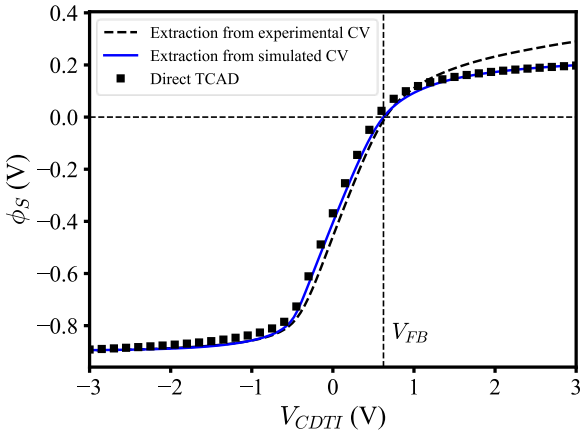


Fig. 13. Comparison between the computed surface potential from the simulated CV curve and the one obtained by solving Poisson's equation with TCAD for the CDTI.

higher. However, the carrier concentration extracted directly from device characterizations is not the only factor that drives dark current. Chemical passivation is also a key parameter as it can be seen with the SRH formalism, as the recombination/generation rate U_s depends on the interface state density N_{it}. It is not addressed here but the interface trap density and traps features could also be studied based on similar characterizations to provide all the required inputs for dark current modeling and simulation. Finally, we can note that depending on technology's architectures, other photodiodes interfaces such as front-side and back-side ones could be addressed with the same approach through dedicated test structures, leading to a better understanding of the dark current generation in these pixels.

REFERENCES

[1] A. Tournier et al., "Pixel-to-Pixel isolation by Deep Trench technology: Application to CMOS Image Sensor," Jun. 2011.

[2] J. L. Regolini, D. Benoit, and P. Morin, "Passivation issues in active pixel CMOS image sensors," Microelectronics Reliability, vol. 47, no. 4, pp. 739–742, Apr. 2007, doi: 10.1016/j.microrel.2007.01.067.

[3] D. Benoit, J. Regolini, and P. Morin, "Hydrogen desorption and diffusion in PECVD silicon nitride. Application to passivation of CMOS active pixel sensors," Microelectronic Engineering, vol. 84, no. 9, pp. 2169–2172, Sep. 2007, doi: 10.1016/j.mee.2007.04.071.

[4] P. Stampfer, G. Meinhardt, T. Grasser, and M. Waltl, "Simulating and Modeling the Influence of Deep Trench Interface Recombination on Si Photodiodes," in 2022 IEEE International Integrated Reliability Workshop (IIRW), Oct. 2022, pp. 1–5. doi: 10.1109/IIRW56459.2022.10032736.

[5] W. Shockley and W. T. Read, "Statistics of the Recombinations of Holes and Electrons," Phys. Rev., vol. 87, no. 5, pp. 835–842, Sep. 1952, doi: 10.1103/PhysRev.87.835.

[6] W. D. Eades and R. M. Swanson, "Calculation of surface generation and recombination velocities at the Si-SiO2 interface," Journal of Applied Physics, vol. 58, no. 11, pp. 4267–4276, Dec. 1985, doi: 10.1063/1.335562.

[7] G. Dingemans and W. M. M. Kessels, "Status and prospects of Al2O3-based surface passivation schemes for silicon solar cells," Journal of Vacuum Science & Technology A, vol. 30, no. 4, p. 040802, Jul. 2012, doi: 10.1116/1.4728205.

[8] P. M. Jordan, D. K. Simon, T. Mikolajick, and I. Dirnstorfer, "Trapped charge densities in Al2O3-based silicon surface passivation layers," Journal of Applied Physics, vol. 119, no. 21, p. 215306, Jun. 2016, doi: 10.1063/1.4953141.

[9] S. M. Sze and K. K. Ng, Physics of Semiconductor Devices, 3rd Edition. Hoboken, NJ: Wiley–Blackwell, 2006.

[10] N. Ahmed et al., "MOS Capacitor Deep Trench Isolation for CMOS image sensors," in 2014 IEEE International Electron Devices Meeting, Dec. 2014, p. 4.1.1-4.1.4. doi: 10.1109/IEDM.2014.7046979.

[11] F. Roy et al., "Fully Depleted, Trench-Pinned Photo Gate for CMOS Image Sensor Applications," Sensors, vol. 20, no. 3, Art. no. 3, Jan. 2020, doi: 10.3390/s20030727.

[12] G. Ghibaudo, S. Bruyere, T. Devoivre, B. DeSalvo, and E. Vincent, "Improved method for the oxide thickness extraction in MOS structures with ultrathin gate dielectrics," IEEE Transactions on Semiconductor Manufacturing, vol. 13, no. 2, pp. 152–158, May 2000, doi: 10.1109/66.843630.

[13] C. N. Berglund, "Surface states at steam-grown silicon-silicon dioxide interfaces," IEEE Transactions on Electron Devices, vol. ED-13, no. 10, pp. 701–705, Oct. 1966, doi: 10.1109/T-ED.1966.15827.

[14] P. Habaš and S. Selberherr, "On the effect of non-degenerate doping of polysilicon gate in thin oxide MOS-devices—Analytical modeling," Solid-State Electronics, vol. 33, no. 12, pp. 1539–1544, Dec. 1990, doi: 10.1016/0038-1101(90)90134-Z.

Parametric Optimization of RF MEMS Variable Capacitor with High Linearity for C-Band Applications

Shakila Shaheen[1,2], Peter Lomax[1,3], Tughrul Arslan[1,4]
[1]Scottish Microelectronics Centre, University of Edinburgh, EH9 3FF, Edinburgh, UK
[2]s.shaheen-2@sms.ed.ac.uk
[3]peter.lomax@ed.ac.uk
[4]Tughrul.Arslan@ed.ac.uk

***Abstract*—A novel RF Microelectromechanical System (MEMS) Variable capacitor is proposed to obtain a linear capacitance-voltage response and large tuning ratio. Rectangular electrostatic torsion actuators are connected to each end of the central capacitor plate by using actuator beams. When voltage is provided in the actuator section, the actuator beam assists in providing steady movement of the central plate in an upward direction. Structural optimization of the RF MEMS Varactor is performed from the perspective of capacitance-voltage response by using ANSYS. The research provides a notable dual contribution that includes both substantial tunability and high linearity. It demonstrates a high capacitance tuning ratio of 312% and a 99% linearity factor in the C-V response, indicating a high linearity factor. The optimized design is demonstrated on a 50-ohm coplanar transmission line to obtain S-parameters by using HFSS. Characteristics of RF performance are analyzed by comparing SiO2 with Si3N4 as the dielectric material in this design. At 4.2 GHz, the varactor with Si3N4 shows -50.2 dB isolation, while the varactor with SiO2 shows -48.2 dB isolation at 5.5 GHz. The proposed study shows that the optimized RF MEMS Varactor shows better performance using Si3N4 at lower frequencies, which makes it appropriate for C-Band applications.**

Keywords— RF MEMS, Full electrode length ratio, lower electrode length ratio, a rectangular electrostatic actuator, FEM, S-parameters, C-Band

I. INTRODUCTION

Due to uncomplicated design, and high-quality factors compared to alternate capacitor types, parallel-plate MEMS variable capacitors have collected significant attention in RF applications. These capacitors offer a variety of options for fine-tuning capacitance in RF circuits, along with other MEMS solid-state capacitors such as variable-gap capacitors, and switched capacitors. To attain minimal phase noise, these devices are extensively used in voltage-controlled oscillators (VCOs) and phase-locked loop (PLL) circuits [1]-[2]. Overall, parallel-plate MEMS variable capacitors play an exciting role in improving the efficiency of RF systems. Parallel plate MEMS variable capacitors show nonlinear capacitance-voltage (C-V) responses regardless of their benefits. This nonlinear behavior is most obvious in the part of the maximum control voltage. The conventional parallel MEMS varactor shows this nonlinear behavior because the air gap between its parallel plates decreases when exposed to an actuation voltage. The plates move more quickly toward one another as the actuation voltage rises resulting in a decrease in the gap and an increase in capacitance. A noteworthy drawback is seen in traditional parallel-plate MEMS variable capacitors due to nonlinear capacitance-voltage (C-V) characteristics [3].

A segmented plate linear capacitor with a noteworthy 99.7% LF and 68 % tuning ratio was introduced in [4]-[7]. In the literature, several analogs and digital MEMS tuneable varactors have been reported [8]-[13] to avoid these limitations of RF MEMS Varactors. Even with these improvements, it is still hard to get a varactor with a large tuning ratio since capacitance changes dramatically when the actuation voltage gets closer to the pull-in point. The pull-in shows the point when the plate moves about one-third of the distance toward the other plate, leading to a capacitance tuning ratio of 50% after applying an actuation voltage. The short tuning range of analog devices, which is less than 1.5, has, however, hampered their efficiency. This drawback results from the bridges in the varactors experiencing a pull-in effect [14].

To solve these limitations of variable capacitors, a simple design was proposed to achieve a varactor of higher tuning ratio [15]. Based on this, a novel design is proposed to avoid the pull-in point and to achieve a large capacitance tuning ratio using the rectangular electrostatic torsion actuator. This study focuses on the linear response of the capacitance and voltage. This paper presents a detailed structural description, modeling, and parametric optimization of the proposed design. The objective is to achieve a notable capacitance tuning ratio coupled with a high linearity factor in the capacitance-voltage response.

II. MEMS VARIABLE CAPACITOR DESIGN

The proposed design comprises an actuator and capacitor section. An electrostatic torsion actuator is used in this design for actuation purposes. An electrostatic rectangular torsion actuator with two similar springs, a moveable plate, and a fixed electrode is schematically shown in Figure 1(a). The horizontal gap between the nearest edge of the fixed electrode and the springs' center is represented by " e_1". The distances "e_2" and "e_3" are the horizontal gaps between the centers of the springs and the ends of the electrode and movable plate, respectively. Additionally, "g" represents the vertical separation between the plates, while "b" is the width of the plate and electrode. The length and width of each spring are l_s and w_s respectively.

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

The full layout of the MEMS varactor is shown by the 3-D view in Fig. 1(b). The electrostatic torsion actuator is linked to the capacitor plate at four points. By providing the voltage to the actuator section, the actuator plate moves downward and the gap between both sides decreases due to electrostatic force. The capacitor plate moves higher due to the actuator beam. As a result, the up-state capacitance decreases, allowing for greater capacitance tuning ratios. There is a linear capacitance-voltage response in the proposed design because the distance between the plates extends and capacitance reduces linearly with applied voltage. Here, as the distance increases, capacitance gradually decreases. An electrostatic torsion actuator is used to move the capacitor plate higher by supplying the voltage in the actuator part to obtain the linear response of capacitance and voltage.

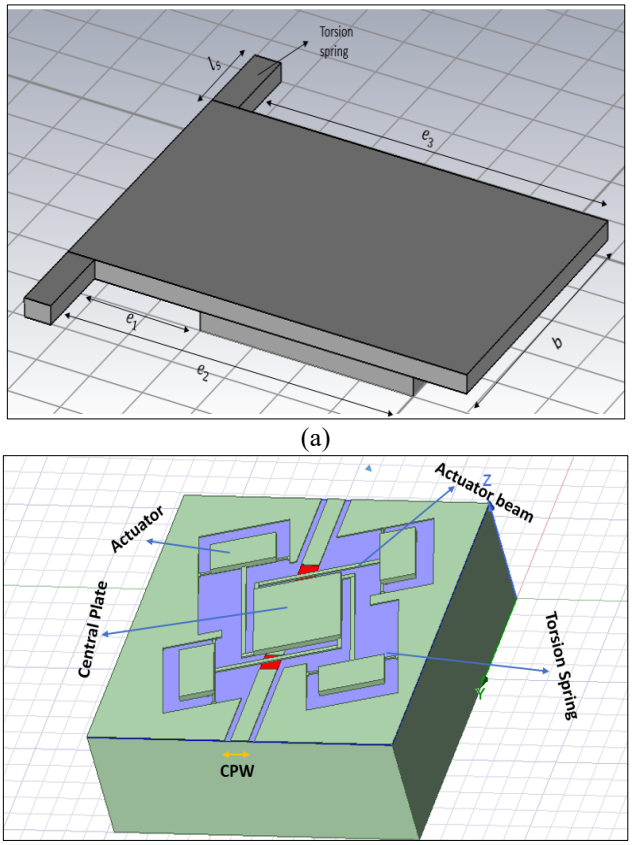


Fig.1. (a). Top view of the electrostatic torsion actuator. (b) 3-dimensional view of the proposed RF-MEMS variable capacitor

More charges gather on the plates as the voltage rises, which causes electrostatic force to increase and mechanical force to decrease relative to electrostatic force. Pull-in happens at this point of instability.

The tunability of the MEMS varactor is determined by the pull-in voltage of the electrostatic torsion actuator since the actuator plate snaps down further than this angle. The pull-in angle can be used to calculate the pull-in voltage [16].

$$V_{pull\ in} = \sqrt{\frac{2T_m g^3 \emptyset_{pull\ in}{}^3}{\varepsilon e_2{}^3 b \quad A}} \tag{1}$$

$$\mathrm{A} = \left[\frac{1}{1-\frac{e_2}{e_3}\emptyset_{pull\ in}} - \frac{1}{1-\frac{e_1}{e_3}\emptyset_{pull\ in}} + \ln\left(\frac{1-\frac{e_2}{e_3}\emptyset_{pull\ in}}{1-\frac{e_1}{e_3}\emptyset_{pull\ in}}\right)\right] \tag{2}$$

Where T_m is the torsion spring constant, g is the initial gap between the plates, $\frac{e_2}{e_3}$ is the electrode length ratio and $\emptyset_{pull\ in}$ is the fractional deflection, κ is the reduced electrode edge location.

$$Electrode\ length\ ratio = \alpha = \frac{e_2}{e_3} \tag{3}$$

$$\kappa = \frac{e_1}{e_3} \tag{4}$$

The relation of electrode length ratio and fractional deflection can be expressed as

$$\alpha\emptyset_{pull\ in} = 0.4404 \tag{5}$$

By applying the voltage, the movement of the capacitor plate in the upward direction can be expressed as,

$$g_1 = \alpha_{pull\ in} * l_a \tag{6}$$

Where l_a is the length of the actuator beam, $\alpha_{pull\ in}$ is the pull-in angle.

The capacitance between the central plates in the original state is given as:

$$C_{down\ state} = \frac{\varepsilon_o \varepsilon_r A}{g} \tag{7}$$

Where ε_o=8.854*10^-12 is the permittivity of free space, A is the area of the central plate, initial gap is represented by g, and ε_r is the relative dielectric constant.

The distance between the plates increases as the voltage is applied, and in this instance, capacitance results in

$$C_{up\ state} = \frac{\varepsilon A}{g+g_1} \tag{8}$$

Where g_1 is the increased gap between the plates after applying the voltage.

The capacitance tuning ratio can be established as [17]

$$C_r = \frac{C_{down\ state}}{C_{up\ state}} \tag{9}$$

III. Structural Optimization

The analysis of a 3D MEMS varactor model is presented by using a non-linear Finite Element Method (FEM) simulation in ANSYS. Here, a copper material with a shear modulus of 44.7GPa is employed. The plates initially have a 1.25um space between them. The impact of variations on the structural dimensions and layout of the Electrode length ratio, overlap area of the actuator plates, torsion and joint springs, and actuator beam length is observed by using ANSYS software.

A. Overlap area between the moveable plate and the stationary electrode

Fig. 2 demonstrates the displacement ratio of the actuator plate at different values of κ by using the full electrode length ratio. When κ =0 the electrode's closest edge is near the torsion spring and the displacement ratio is maximum because the reduction in the κ value corresponds to an increase in the overlap area between the actuation electrode and the bottom plate which leads toward higher tuning of capacitance. Therefore, the capacitor can achieve a larger capacitance tuning ratio at smaller k values. Hence best-performed varactor at $\kappa = 0\ is$ used in section B.

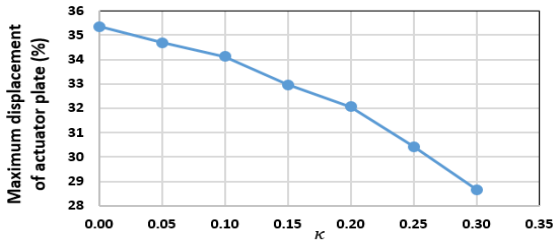


Fig.2. Variation in displacement for different κ values

B. Electrode Length Ratio Optimization

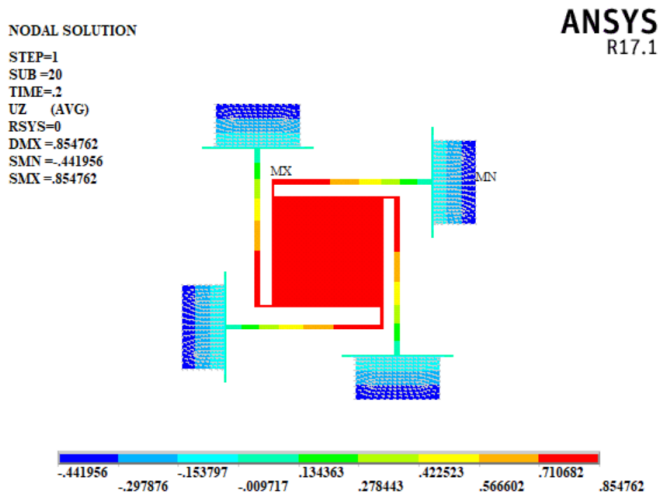
The electrostatic torsion actuators which determine the pull-in voltage and the motion of the capacitor plate, are connected to the central capacitor plate by a thin actuator beam' as shown in Fig. 1(b). The proposed MEMS varactor is observed by using different electrode length ratios. Calculated and experimental results of different electrode length ratios are compared and an optimized result is obtained.

TABLE 1: DIMENSIONS OF THE PROPOSED RF MEMS VARIABLE CAPACITOR

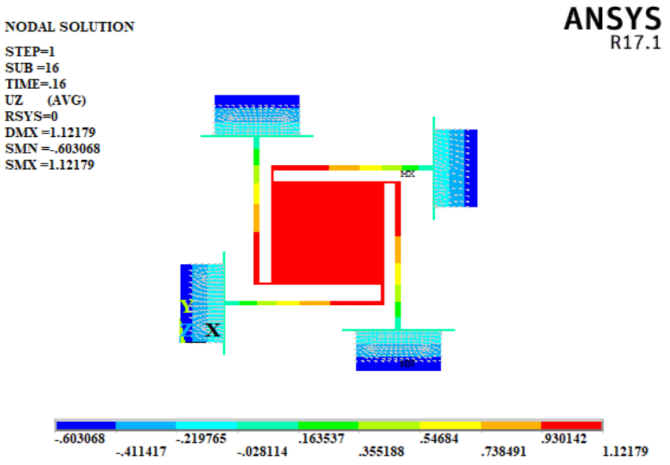
Parameters	Values (μm)
l_a	560
g	1.25
l_s	50
w_s	10
e_3	150
b	300
A	400*400

For MEMS Varactor with full electrode length ratio $\alpha = 1, \emptyset_{pull\ in} = 0.4404$ indicates that the pull-in point occurs after the moveable plate has traveled over 44% of its whole travel range. An actuator beam is presented to connect capacitor part to the actuator part. The actuator plate moves downward due to the electrostatic force. This electrostatic force is then transformed into an upward motion of the central plate by using the actuator beam. Thus, the capacitor plate is moved higher and distance between both sides increased. In this case, the maximum central plate displacement is 0.62μm after which pull-in occurs.

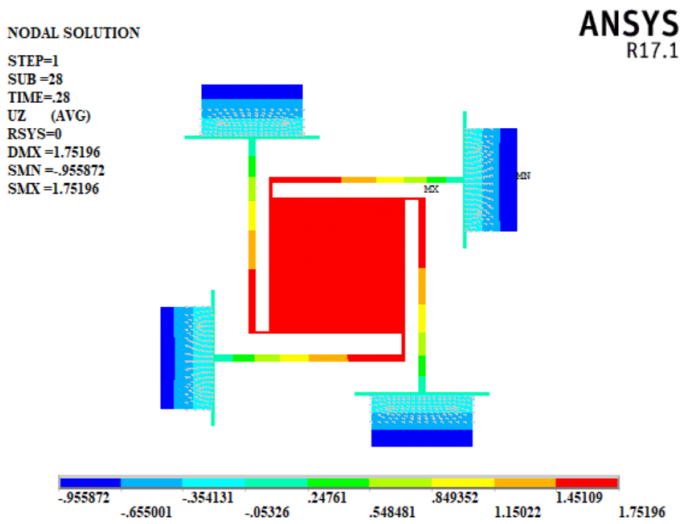
For partial electrode length ratio, $\alpha = 0.5, \emptyset_{pull\ in} = 0.8808$, which means the partial plate electrode ratio gives about 88 % travel range of the moveable plate in the actuator section. Here central plate displacement is 1.11 μm. For an electrode length ratio of 0.4404,$\emptyset_{pull\ in} = 1$. It is the case of the full travel range of moveable plates. Capacitor plate displacement in this case is 1.86μm. Experimental results for different electrode length ratios are shown in Fig 3.



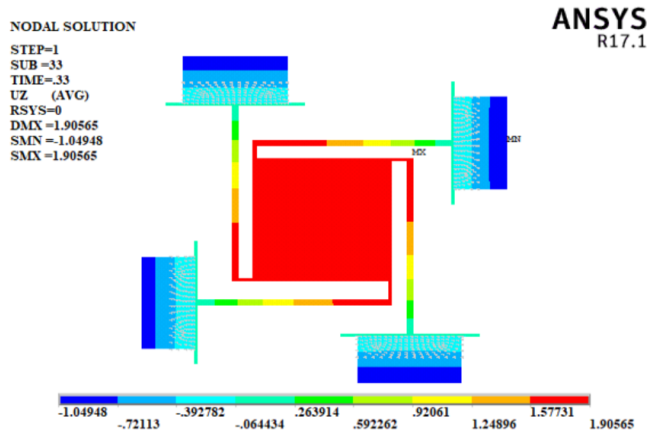
(a)



(b)



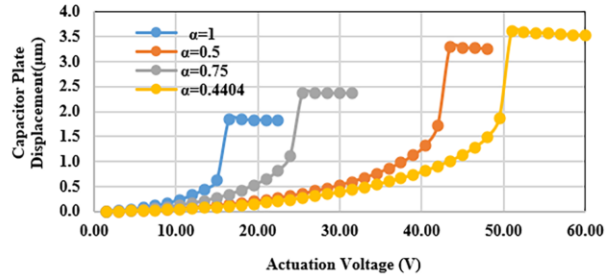
(c)



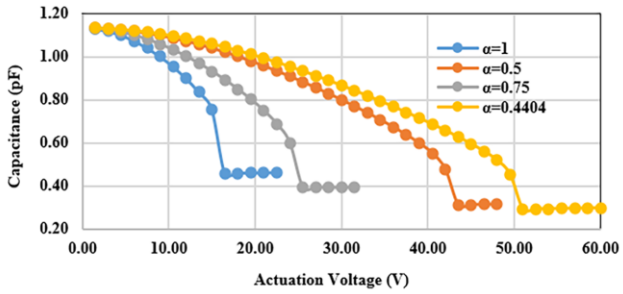
(d)

Fig.3. 3D FEM model of MEMS varactor: z-displacement (a) α =1 (b) α =0.75 (c) α = 0.5 (d) α = 0.4404

In Fig. 4(a). the central plate displacement at different electrode length ratios is presented. Among these, when the electrode length ratio is 0.4404, the capacitor plate displacement is higher. This is because it permits the actuator plate to have the maximum travel range, resulting in the greatest displacement. Furthermore, Fig. 4(b). shows that the up-capacitance is minimum when the electrode length ratio is 0.4404, so a large tuning ratio is obtained in this specific case. The highest tuning ratio is achieved by avoiding the pull-in point. In conclusion, case of α=0.4404 presented the best performance in terms of maximum displacement and minimum up-capacitance while avoiding the issues associated with the pull-in point. Variation in tuning ratio for different values of α is shown in Fig. 5.



(a)



(b)

Fig.4. (a). Capacitor plate displacement for α=1, 0.75, 0.5 0.4404 (b) Up-capacitance for α=1, 0.75, 0.5, 0.4404

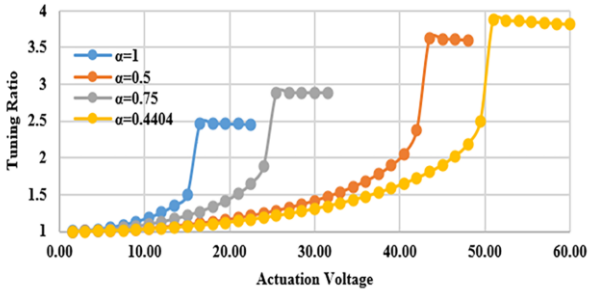


Fig.5. Variation in tuning ratio for different electrode length ratio

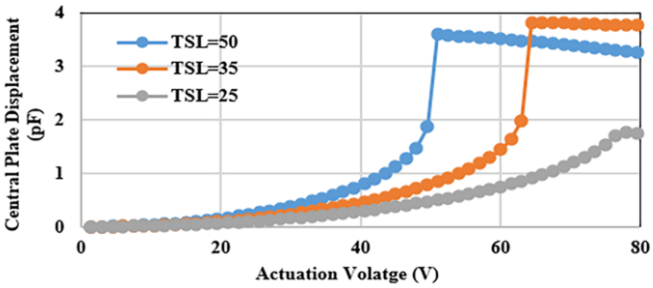
TABLE 2: COMPARISON OF DIFFERENT PARAMETERS OF PROPOSED RF MEMS VARIABLE CAPACITOR

α	κ	$V_{pull\ in}$	$\emptyset_{pull\ in}$ (%)		C_r(%)
			Theoretical	*Simulation*	
1	0	16	44	35	209
0.75	0	24	58	48	234
0.5	0	42	88	76	296
0.4404	0	49.5	100	83	312

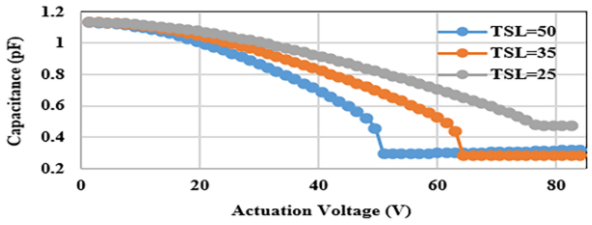
C. *Effect of Length and Width of Torsion Spring and Joint Spring*

Dimensions of the *Torsion spring and joint spring* have a deterministic effect on the varactor properties such as pull-in voltage, central plate displacement, and pull-in angle when voltage is applied in the actuator section. The impact of the length and width of the torsion spring and joint spring especially on the pull-in voltage and capacitor plate displacement of the varactor is shown in Fig. 6. The increase in pull-in voltage for a change in torsion spring length from 50μm to 25μm is depicted in Fig.6 (a) (b). Similarly, there is an increase in pull-in voltage when the torsion spring width increases. There is a slight difference in capacitance by varying the torsion spring lengths, so 50 um is considered best.

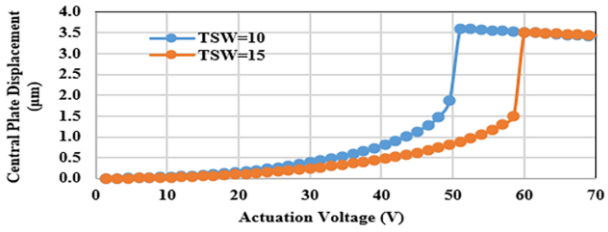
The effect of variation in joint spring length and width is shown in Fig.6 (c). & (d). and an optimized result is obtained



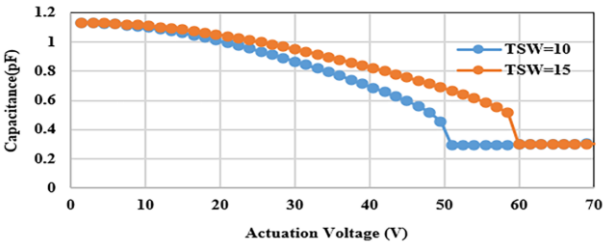
(a)



(b)



(c)

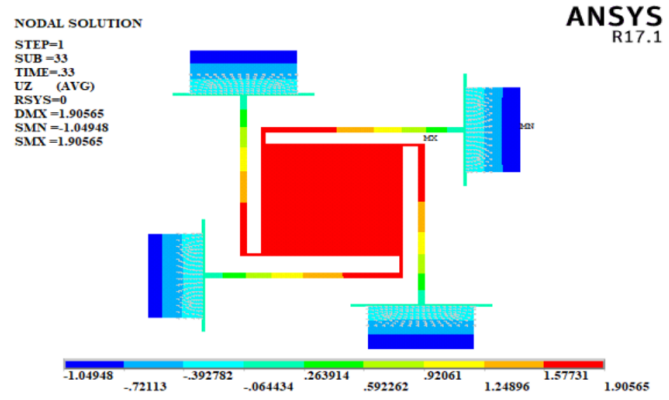


(d)

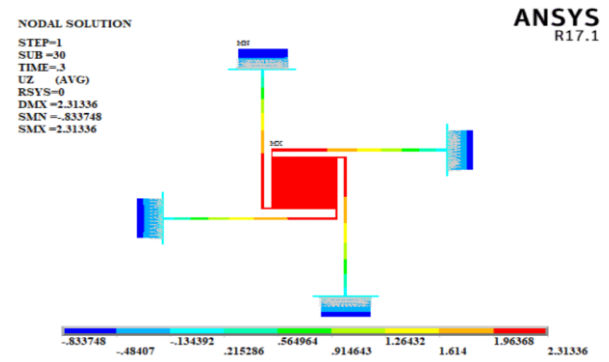
Fig.6 (a). Variation in central plate displacement for TSL=50, 35, 25 (b) Comparison of up-capacitance of TSL=50, 35, 25 (c) Comparison of central plate displacement for TSW=10, 15 (d) Comparison of up-capacitance of TSW=10, 15

D. Actuator Beam Length Optimization

The actuator section and the capacitor section are connected using an actuator beam. Performance parameters Of MEMS varactors are affected by the dimensions of the actuator beam length when voltage is provided in the actuator section. The impact of the actuator beam length is shown in the figure when it is changed from 560µm ($\frac{l_a}{e_3} = 3.7$) to 1050µm ($\frac{l_a}{e_3} = 7$). Better performance is obtained for an actuator beam length of 1050 µm in terms of pull-in voltage and central plate displacement as shown in Fig 7. However, to avoid the complexity of the structure actuator beam length of 560 µm has been used in this design.



(a)

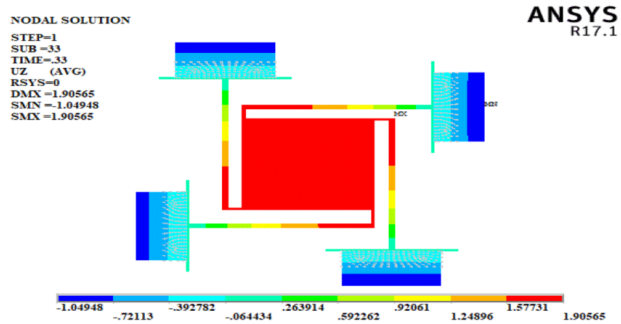


(b)

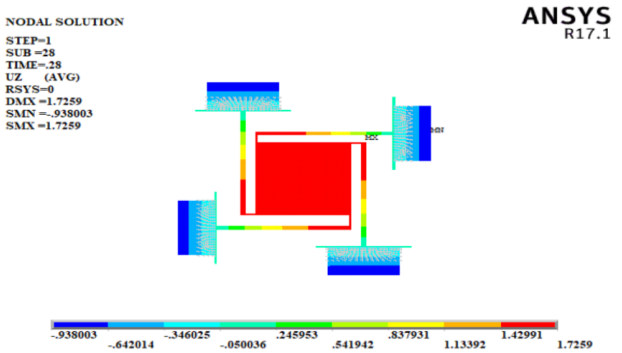
Fig. 7. ANSYS results showing z displacement (a) for l_a/e_3 =3.7 (b) for l_a/e_3 =7

E. Bridge Material and Linearity Factor Optimization

Copper is used as a material for bridge in the proposed design. The actuator plates and central plate are simulated using copper material with a thickness of 25µm, while connecting actuator beams, join springs, and torsion springs are simulated with a thickness of 6µm. Because gold has a low resistivity, it causes the central plate displacement to be less and the pull-in point to occur earlier than the case with copper. Therefore, using copper as the bridge material results in improved performance in terms of pull-in point, central plate displacement, and tuning ratio. A comparison of central plate displacement actuator plate displacement by using both gold and copper as bridge material is shown in Fig.8.



(a)



(b)

Fig.8. ANSYS results showing z displacement (a) for copper as bridge materials (b) 3D for gold as bridge material

The provided equation, which is defined in the following [12], is used to calculate the linearity factor. The range of the linearity factor is 0 to 1.

$$LF = \frac{n\sum_{i=1}^{n} C_i V_i - \sum_{i=1}^{n} C_i \sum_{i=1}^{n} V_i}{\sqrt{\left[n\sum_{i=1}^{n} C_i{}^2 - \left(\sum_{i=1}^{n} C_i\right)^2\right]\left[n\sum_{i=1}^{n} V_i{}^2 - \left(\sum_{i=1}^{n} V_i\right)^2\right]}} \quad (18)$$

Therefore, the linearity factor value is 100% for a perfectly straight line. A linearity factor of 99 % is achieved for optimized design with a high tuning ratio.

IV. S-Parameter Analysis of RF MEMS Varactor

The proposed study focuses on the RF MEMS varactors operating in a Coplanar Waveguide (CPW) configuration with a characteristic impedance of 50Ω and a frequency range of 1-10 GHz. The CPW comprises a signal line and two grounds, forming a transmission line considered for effective RF performance. The RF MEMS varactor is studied in both ON and OFF states. OFF state represents a state where there is no gap between the plates, targeting maximum isolation and 0dB return loss.

The study focuses on evaluating the characteristics of S-parameters of the optimized RF MEMS Varactor. The impact of variations in dielectric material on S-parameters is observed.

Figures 9a and 9b provide a comparative analysis of the ON state responses of the device. Meanwhile, Figures 9c and 9d illustrate the OFF-state responses using SiO2 and Si3N4 as dielectric materials. Notably, Si3N4, with its higher dielectric constant compared to SiO2, results in a shift of the optimum isolation value to a lower frequency. At 4.2 GHz, the varactor with Si3N4 shows -50.2 dB isolation, while the varactor with SiO2 shows -48.2 dB isolation at 5.5 GHz. This proposes that using Si3N4 significantly improved the isolation performance of optimized RF MEMS Varactor at a lower frequency.

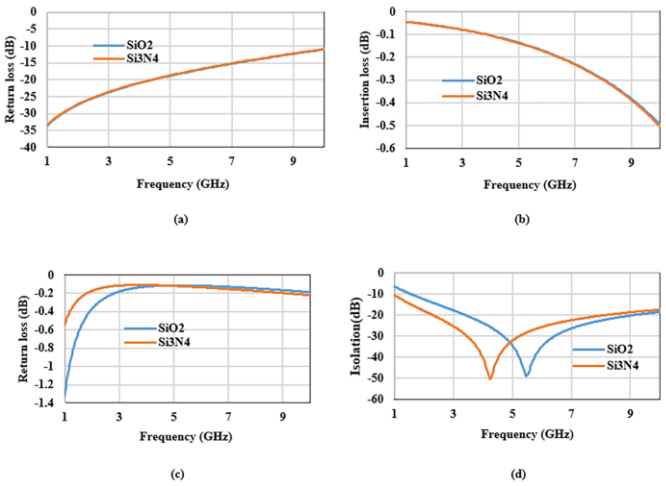


Fig.9 (a) Return loss in ON state (b) Insertion loss in ON state(c)Return loss in OFF state (d) Isolation in OFF state

V. Conclusion

In this study, parametric optimization of the proposed design has been performed to get a higher tuning ratio of 312% with a 99% linearity factor of capacitance-voltage response. This paper demonstrates that by using a lower electrode length ratio, a higher tuning ratio varactor can be attained by avoiding the pull-in effect. The optimized design shows better isolation performance at 4.2 GHz which leads to being used in the C-Band applications. In the next stages of our research, it is planned to proceed from the proposed design to its actual implementation through the fabrication stage which will be beneficial for additional testing and practical applications.

References

[1] A. Dec and K. Suyama, "Microwave MEMS-based voltage-controlled oscillators," IEEE Trans. Microw. Theory Tech., vol. 48, no. 11, pp. 1943– 1949, Nov. 2000.

[2] M. Yellepeddi and K. Mayaram, "Issues in the design and simulation of a MEMS VCO based phase-locked loop," in Proc. IEEE Int. Symp. Circuits Syst., 2007, pp. 1553–1556.

[3] G.M. Rebeiz, RF MEMS Theory, Design and Technology. Hoboken, NJ: Wiley, 2003.

[4] Shavezipur M, Khajepour A, Hashemi SM (2008a) Development of novel segmented-plate linearly tunable MEMS capacitors. J Micromech Microeng 18(3):035035

[5] Shavezipur M, Khajepour A, Hashemi SM (2008b) A novel linearly tunable butterfly-shape MEMS capacitor. Microelectron J 39(5):756–762

[6] Shavezipur M, Hashemi SM, Khajepour A, Nieva P (2008c) Development of a linearly tunable modified butterfly-shape MEMS capacitor. Int Mech Eng Congr Expos (ASME) 2008:495–500

[7] Shavezipur M, Nieva P, Khajepour A, Hashemi SM (2009) Development of parallel-plate-based MEMS tunable capacitors with linearized capacitance–voltage response and extended tuning range. J Micromech Microeng 20(2):025009

[8] Sharaf, A., Nasr, A., Elshurafa, A. M., & Serry, M. (2022). Design analysis and simulation of a digital RF MEMS varactor with high capacitive ratio. Microsystem Technologies, 28(8), 1831-1844.

[9] Ghalandarzadeh, M., & Afrang, S. (2021). A new wide tunability MEMS based variable capacitor using two separate electro-static vertical comb drive actuators. *International Journal of Engineering*, *34*(11), 2534-2544.

[10] Barzegar, S., Mirzajani, H., & Ghavifekr, H. B. (2015). A new linearly tunable RF MEMS varactor with latching mechanism for low voltage and low power reconfigurable networks. *Wireless Personal Communications*, *83*, 2249-2265.

[11] Jung, M., & Min, B. W. (2019). A widely tunable compact bandpass filter based on a switched varactor-tuned resonator. IEEE Access, 7, 95178-95185.

[12] Akhtar, A., Zaidi, S. M. T., Khan, F., Nadeem, J., & Shaheen, S. (2018, November). Design and simulation of MEMS based varactor with high tunability for digital communication systems. In *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 426-430). IEEE.

[13] Shahi, M. (2021). Development of tunable mems capacitor with dual-mode deformation to improve tunability and linearity (Doctoral dissertation, Southern Illinois University at Edwardsville).

[14] G. M. Rebeiz, RF Mems-Theory, Design and Technology, vol. 53, no. 9. John Wiley & Sons, 2003.

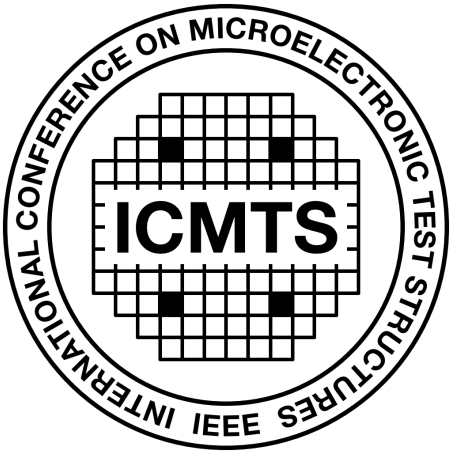
[15] Shaheen, S., Saleem, M. M., Zaidi, S. M. T., & Akhtar, A. (2018, November). Design, Simulation and Parametric Optimization of MEMS Based Varactor. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 1159-1164). IEEE.

[16] O. Degani, E. Socher, A. Lipson, T. Leitner, D. J. Setter, S. Kaldor, and Y. Nemirovsky, "Pull-in study of an electrostatic torsion microactuator," J. Microelectromech. Syst., vol. 7, no. 4, pp. 373–379, Dec. 1998.

[17] Rebeiz GM (2003) RF mems-theory, design and technology, vol 53. Wiley, New York. doi:10.1017/CBO9781107415324.004

SESSION 6

Wafer Measurements



17^{th} April 2024, 13:40–15:00

Session Co-Chairs: Francesco Driussi, *Università di Udine, Italy*
Kiyoshi Takeuchi, *University of Tokyo, Japan*

Efficient Characterization Methodology for Low-Frequency Noise Monitoring

Luca Pirro[1], Talha Chohan[1], Philipp Liebscher[1], Maximilian Juettner[1], Felix Holzmueller[1], Ruchil Jain[1], Yannick Raffel[2], Konrad Seidel[2], Ricardo Revello Olivo[2], Alban Zaka[1] and Jan Hoentschel[1]
[1] GlobalFoundries Fab1 LLC & Co.KG
01109 Dresden, Saxony, Germany
Email: luca.pirro@globalfoundries.com
Telephone: +49 3512774057
[2] Fraunhofer-IPMS
01109 Dresden, Saxony, Germany

***Abstract*—In this work, a novel methodology to characterize the Low-Frequency Noise LFN on large device statistics and suitable for production monitoring is proposed. The maximum drain current fluctuations over time are measured. The slope of the LFN distribution is modeled with physical equations related to the basic device properties. The approach is validated by studying the impact of transistor geometry (different gate width, number of fingers and gate length) as well as gate oxide thickness and characterization temperature. In conclusion, the proposed methodology is tested evaluating different process integration elements. The outcome is compared to classical LFN read-out for final confirmation.**

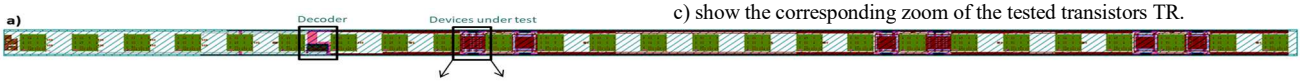
I. Introduction

Recent market trends show that superior analog performance is becoming a major prerequisite for advanced integrated circuits [1]. The continuous reduction of power consumption in conjunction with the device miniaturization makes analog designs more and more sensitive to current oscillations (noise) [2-3]. Thus, the low-frequency noise (LFN) becomes a key factor defining the transistor performance and consequently the product functionality [1]. In a production phase, it is essential to access the LFN data at transistor level. This is important for an early detection of line issue, and it is required to correlate back to the single integration elements, needed for process debugging. Additionally, due to the stochastic nature of low-frequency noise, a solid correlation with product data requires to measure LFN for numerous devices. At the same time, in production environment; shorter measurement time with the possibility of automated analysis is ideally desired. To meet all these needs, a novel algorithm to monitor the LFN is proposed. The used test structures and measurement methodology are presented in Section II. In Section III the approach is validated with physical equations for different study-cases (different device geometries, gate oxide thickness and measurement temperature). Eventually, the novel methodology is employed to investigate the impact of integration elements on noise signal, and the outcome is compared to classical LFN approach for final validation.

II. Test Structures and Measurement Setup

A. Test Structures

The low-frequency noise events happening in a MOSFET are random and independent from each other [2]. Thus, for accurate process monitoring and solid correlation with circuit data, the LFN signal must be evaluated over several thousand of devices fabricated on the same wafer. In this work we used the array test structure TS shown in Figs. 1 [4]. The TS is composed of 4 devices placed around a common centroid in a dense layout environment, emulating a circuit-like configuration (Figs. 2b-c). Using a 4 bit programmable decoder, it is possible to probe 6 devices at the same time [4], for a total of 96 transistors per test structure. The TS is optimized to fit on a standard production scribe-line. Implementing a wafer full map measurement, up to ~10k transistors per hour can be probed to access the corresponding maximum drain current oscillations $MinMax\_I_D$ for each device. As a result, the measurement time is shorter compared to a conventional LFN setup, where no parallelization is possible and the measurement time per device is estimated to be ~10s. Additionally, no data transformation is required (no Fast Fourier Transformation).



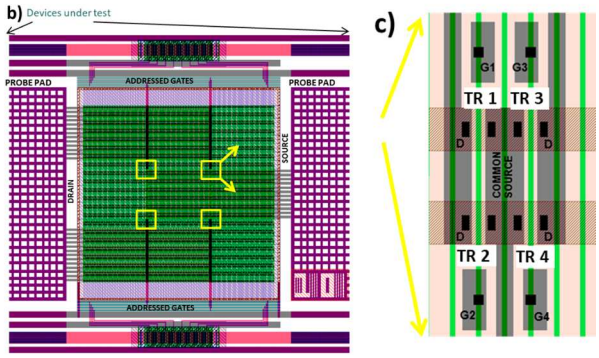


Fig. 1: a) Visual representation of the test structure TS used for noise characterization allowing to probe 6 devices in parallel for a total of 96 instances per TS. The TS fits with a standard frame scribe-line. Figs. b) and c) show the corresponding zoom of the tested transistors TR.

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

B. Measurement Setup

In a Field-Effect Transistor FET, the measured drain current I_D fluctuates over time. Several works investigate the root-cause and the corresponding physical origin of this effect [5-7]. From a circuit point of view, however, the origin of the LFN is a secondary factor in comparison to the magnitude of I_D fluctuations, which at the end affects the circuit functionality. For this reason, we propose to monitor for each transistor the maximum I_D oscillations over time (*MinMax_I*$_D$), as represented in the schematic of Fig. 2. The use of *MinMax_I*$_D$, as monitor parameter makes the characterization faster and avoids the needs to store the time domain trace for each die, reducing the hardware and software requirements. Fig. 2 shows an example of the reverse Cumulative Distribution Function (1-CDF) of *MinMax_I*$_D$/I_D in a semi-log scale. The tail of the distribution identifies the transistors with the highest (worst) noise. To monitor this tail with an automatized algorithm, 1-CDF distribution is fitted between 1% and 20% (see Fig. 2). This fitting range allows to study devices with strong noise signature and, at the same time, it shields the algorithm from abnormal measurement outliers, enabling reliable automatization. The flow-chart in Fig. 3 summarizes the characterization procedure, highlighting the main steps and parameters. In the next section, after introducing some basic equations to explain the measured trends, the impact of device geometry, gate oxide thickness and process integration elements are reported.

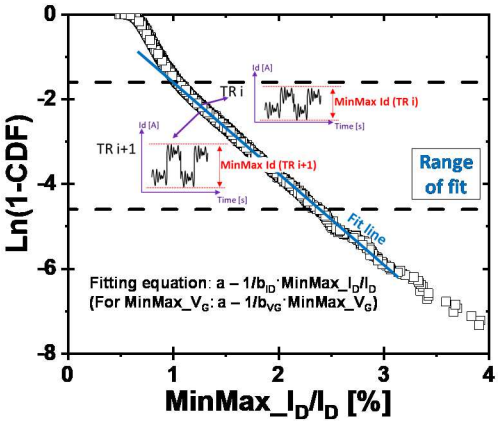


Fig. 2: Example of measured Ln(1-CDF) versus *MinMax_I*$_D$/I_D. The linear fit range between 1% and 20% of Ln(1-CDF) is highlighted. A schematic of the drain current versus time is represented for two consecutive transistors TR *i* and *i+1*, and the corresponding *MinMax_I*$_D$ over time is shown.

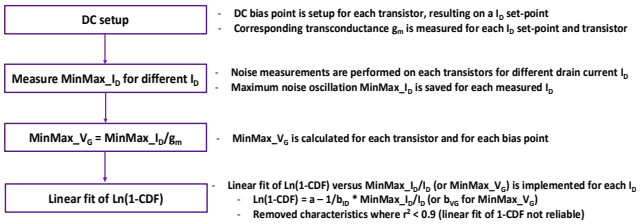


Fig. 3: Flow chart of the characterization methodology with associated explanations of key steps.

III. 3. Validation of Characterization Methodology

A. Basic Equations

The Random Telegraph Noise (RTN) measured on a transistor can be phenomenologically expressed as in (1), according to [8]:

$$\frac{\Delta I_D \approx MinMax\_I_D}{I_D} = \frac{\eta \cdot q}{W_T \cdot L \cdot C_{OX}} \frac{g_m}{I_D} \left(1 + \Omega' \frac{I_D}{g_m}\right) \quad (1)$$

where ΔI_D is the current jump between two RTN levels, approximated to *MinMax_I*$_D$ in this work, assuming enough devices are measured and focusing on the tail of the CDF distribution. η is a fitting parameter to account for the charge location and q is the unit charge. The device geometry is accounted in the transistor total width W_T, length L and oxide capacitance C_{OX}. Ω' is the mobility fluctuation parameter due to trap-related remote Coulomb scattering [8]. In subthreshold regime, Ω' can be neglected [5]. Thus, in first approximation and for a single device, the noise contribution with respect to the gate bias is independent of g_m/I_D:

$$MinMax\_V_G = \frac{MinMax\_I_D}{g_m} \propto \frac{\eta \cdot q}{W_T \cdot L \cdot C_{OX}} \quad (2).$$

The noise distribution over the whole wafer (CDF of *MinMax_V*$_G$ can be computed integrating (2) over the entire trap energy. At the same time, the linear fit of Ln(1-CDF) of *MinMax_V*$_G$ can be expressed as:

$$Fit\ Ln(1 - CDF) = a - \frac{1}{b_{VG}} MinMax\_V_G \quad (3)$$

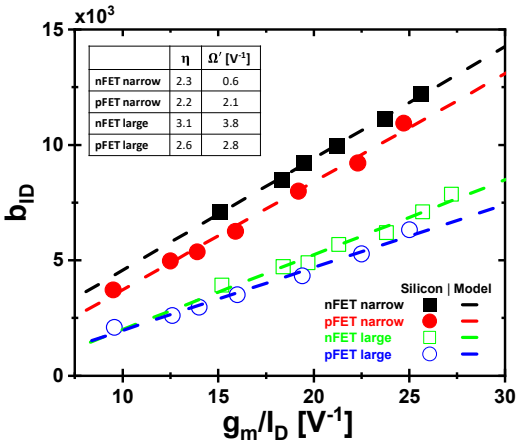
where the factor a does not relate to RTN events. For similarity with (2), the slope of Ln(1-CDF) distribution b_{VG} results:

$$b_{VG} \propto \frac{\eta \cdot q}{W_T \cdot L \cdot C_{OX}} \quad (4).$$

With the same approach it is possible to show that b_{ID} represents the slope extracted from Ln(1-CDF) of *MinMax_I*$_D$/I_D. In the next section, the derived equations will be compared to silicon trends for validation.

B. Silicon Validation

Fig. 4 shows the slope b_{ID} extracted from Ln(1-CDF) of *MinMax_I*$_D$/I_D as a function of transistor intrinsic gain g_m/I_D. Different device types are studied: nFET (square symbols) and pFET (circle symbols), narrow (plain symbols) and large (empty symbols) transistors. For all the investigated conditions, b_{ID} shows a linear dependency versus g_m/I_D, as predicted by (1). Additionally, the extracted η and Ω' (see table in Fig. 4) are consistent with the values reported in the literature [8], validating the initial assumption that ΔI_D can be approximated with *MinMax_I*$_D$.



	η	Ω' [V⁻¹]
nFET narrow	2.3	0.6
pFET narrow	2.2	2.1
nFET large	3.1	3.8
pFET large	2.6	2.8

Fig. 4: b_{ID} versus g_m/I_D for narrow (fully symbols) and large (empty symbols) devices. nFET (square symbols) and pFET (circle symbols) transistors are reported. Linear dependency of b_{ID} versus g_m/I_D is obtained as predicted by (1).

The omega factor Ω' relates to mobility Coulomb scattering. In first approximation, it can be neglected for transistors characterized in subthreshold regime [5]. Thus, b_{VG} slope of Ln(1-CDF) of *MinMax_*V_G results independent of g_m/I_D. To support this conclusion with measurement data, b_{VG} is plotted versus g_m/I_D in Figs. 5 for devices with different active gate width W in case of nFET (a) and pFET (b) structures. b_{VG} shows no evident dependency versus g_m/I_D for different transistor types and geometries, supporting the conclusion that no significant mobility Coulomb scattering is happening on the characterized devices (Ω' contribution is negligible). Additionally, one can note that b_{VG} scales proportional to $1/W$, as predicted by (4). The impact of number of fingers NF is also investigated (empty symbols in Figs. 5). The data follows the expected trends: b_{VG} level is defined by the total device width W_T (W_T=W·NF) as per (4).

To further extend the silicon validation of (4), the device length L impact on b_{VG} is reported in Fig. 6 for nFET (plain symbols) and pFET (empty symbols). The longer L is, the smaller b_{VG} is, in accordance to (4).

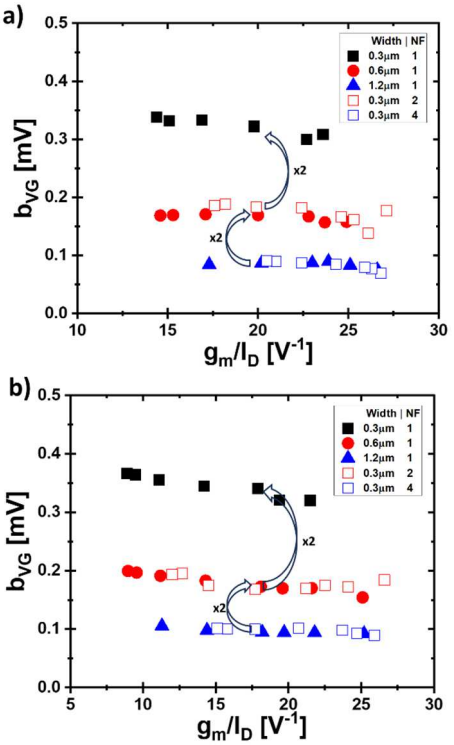


Fig. 5: Measured b_{VG} versus intrinsic gain g_m/I_D for nFET (a) and pFET (b) devices. Different transistor width W (plain symbols) and number of fingers NF (empty symbols) are reported. b_{VG} is independent to g_m/I_D and it scales directly to the total device width (W_T=W·NF) as per (4).

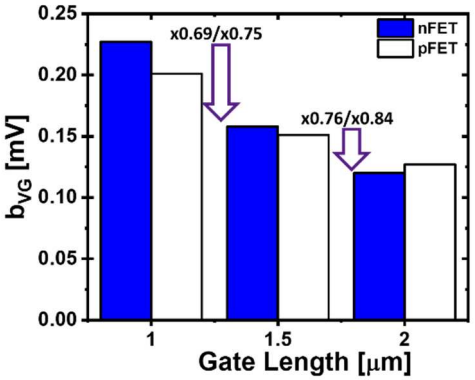


Fig. 6: b_{VG} measured at the same g_m/I_D for nFET (plain symbols) and pFET (empty symbols) devices with different gate length L. b_{VG} scales inversely proportional to transistor length as per (4).

Another main parameter defining the transistor performance is the physical oxide thickness. Its contribution is captured in (4) by the effective oxide capacitance C_{OX}. The corresponding modulation on the slope b_{VG} of noise distribution tail is shown in Fig. 7 for different device types (nFET in square symbols and pFET in circle symbols) and geometric areas (plain symbols for narrow FETs and empty symbols for large ones). Once again, a linear dependency of b_{VG} versus C_{OX} is obtained in accordance with (4).

Eventually, also the characterization temperature is varied in Fig. 8 for narrow (a) and wide (b) nFET devices. b_{VG} is reported as a function of the corresponding device intrinsic gain g_m/I_D. An increase of the characterization temperature (from ambient to 75°C) leads to a minor reduction of the Ln(1-CDF) $MinMax\_V_G$ slope b_{VG}. At the same time, a temperature reduction below ambient one clearly modifies the dependency of b_{VG} with respect to the intrinsic gain. This can be explained by the increased contribution of the parasitic device activated near to the STI, appearing when reducing the characterization temperature and for weak electric field [9]. (1) does not account for this edge transistor contribution. Further work is on-going to model the effect and expand (1)-(4) to capture it.

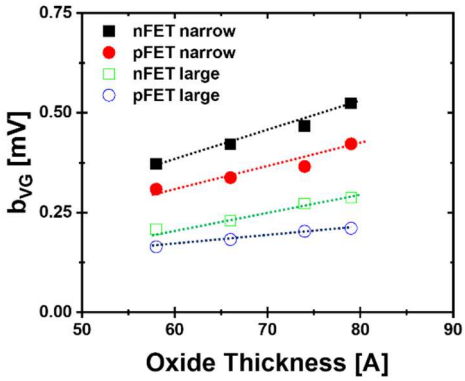


Fig. 7: b_{VG} versus oxide thickness for nFET (square symbols) and pFET (circle symbols) devices, for different transistor areas. b_{VG} scales proportionally to oxide thickness as predicted by (4).

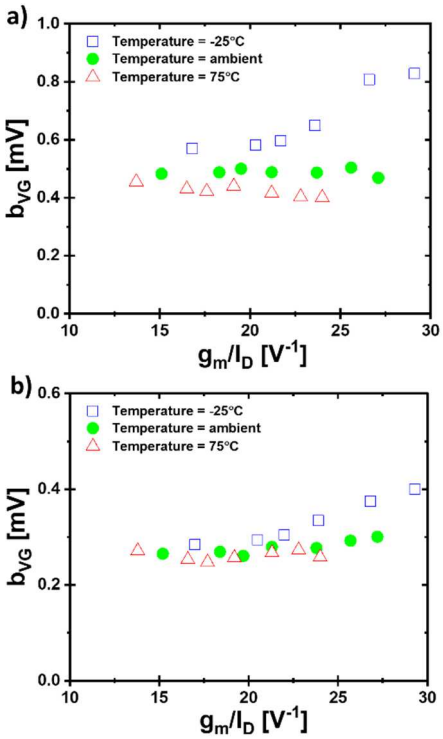


Fig. 8: b_{VG} as a function of intrinsic gain g_m/I_D for narrow (a) and wide (b) transistors measured for different temperatures. Increasing the characterization temperature from ambient to 75°C induces a minor reduction of b_{VG}, while for a characterization temperature lower than the ambient one, an additional dependency of b_{VG} versus g_m/I_D appears, which is not captured in (1) and (4).

In conclusion, Figs. 9 shows b_{VG} versus g_m/I_D measured for devices with different integration elements for nFET (a) and pFET (b). The introduction of HMKG layer in the gate oxide degrades b_{VG}, while negligible differences are seen between pFET employing silicon cSi or silicon germanium cSiGe channel, as already reported in inversion regime [10]. The insets in Figs. 9 show the corresponding gate voltage power spectrum density S_{VG} extracted at 100Hz. S_{VG} and b_{VG} show similar measured trends for the investigated process splits, confirming that the proposed approach can be employed for fast monitor of the low-frequency noise distribution of FETs.

IV. Conclusions

In this work, a novel methodology to monitor the low-frequency LFN noise is proposed. The maximum drain current fluctuations over time are measured. It is shown that the extracted parameters using proposed methodology relate to phenomenological random telegraph noise (RTN) equations, correlating well with the device properties: total transistor width and length, and gate oxide thickness. Different process integration elements have also been investigated. Consistent read-outs with common low-frequency noise approach are obtained, validating the methodology and demonstrating the possibility to monitor LFN with a fast and automated algorithm.

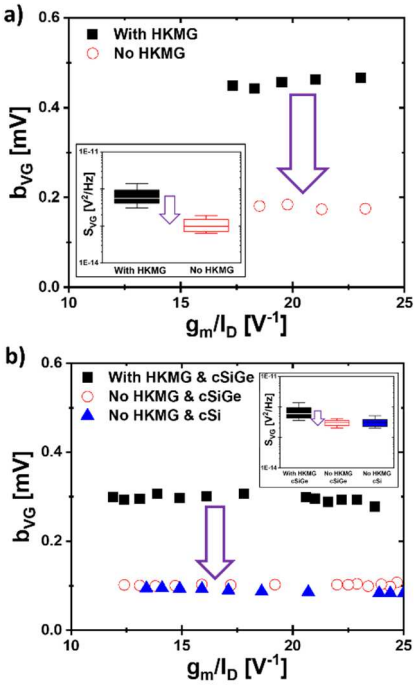


Fig. 9: Measured b_{VG} versus g_m/I_D for devices with different integration elements for nFET (a) and pFET (b). The corresponding gate power spectrum density S_{VG} extracted at 100Hz are reported in insets. Consistent split trends are measured for b_{VG} and S_{VG}, validating the characterization technique.

ACKNOWLEDGMENT

This work has received funding from the Ministry of Economy and Climate (BMWK) and the Free State of Saxony (SMWA) through the subsidy project IPCEI – EUROFOUNDRY.

REFERENCES

[1] Y. Oike, "Expanding Human Potential through Imaging and Sensing Technologies," IEDM, pp. 6-10, 2022.

[2] K. Takeuchi, T. Nagumo, S. Yokogawa, K. Imai, and Y. Hayashi, "Single-charge-based modeling of transistor characteristics fluctuations based on statistical measurement of RTN amplitude," VLSI, pp. 54-55, 2009.

[3] E.G. Ioannidis, S. Haendler, C.A. Dimitriadis, and G. Ghibaudo, "Characterization and modeling of low frequency noise in CMOS inverters," Solid State Electronics, vol. 81, pp. 151-156, 2013.

[4] A. Jayakumar, N. Chan, L. Pirro, O. Zimmerhackl, M. Otto, T. Kleissner, and J. Hoentschel, "Area-Efficient and Bias-Flexible Inline Monitoring Structure for Fast Characterization of RTN and Transistor Local Mismatch in Advanced Technologies," ICMTS, sec. 3.1, 2020.

[5] G. Ghibaudo, O. Roux, Ch. Nguyen-Duc, F. Balestra, and J. Brini, "Improved Analysis of Low Frequency Noise in Field-Effect MOS Transistors," Physica Status Solidi, vol. 124 Is. 2, pp. 571-581, 1991.

[6] M. von Haartmann and M Östling, "Low-frequency noise in advanced MOS devices", Dordrecht: Springer 2007.

[7] A. J. Scholten, L.F. Tiemeijer, R. van Langevelde, R.J. Havens, A.T.A. Zegers-van Duijnhoven, and V.C. Venezia, "Noise modeling for RF CMOS circuit simulation," IEEE Transactions on Electron Devices, vol. 50. Is. 3, pp. 618-632, 2003.

[8] T. A. Oproglidis, T. A. Karatsori, C. G. Theodorou, A. Tsormpatzoglou, S. Barraud, G. Ghibaudo, and C. A. Dimitriadis, "Impact of Hot Carrier Aging on the 1/f and Random Telegraph Noise of Short-Channel Triple-Gate Junctionless MOSFETs," IEEE Transactions on Device and Materials Reliability, vol. 21, Is. 3, pp 348-353, 2021.

[9] L. Pirro *et al.*, "Impact of Electrical Defects located at Transistor Periphery on Analog and RTN Device Performance," IRPS, p. 59 1-5, 2022.

[10] Ch. L. N. Pavan, R. Divakaruni, A. Chakravorty, and D. R. Nair, "Characterization and Analysis of Random Telegraph Noise in Scaled SiGe Channel HKMG pMOSFETs," IEEE Transactions on Electron Devices, vol. 69. No. 2, pp. 456-461, 2022.

Rapid MOSFET threshold voltage testing for high throughput semiconductor process monitoring

Michael H. Herman, Trenton T. Nguyen, Ken Wong, Jeff Johnson, Ben Morris[1],
[1] Authors are with Parametric Test Group, Advantest America, San Jose, CA 95134 United States.
Principal author contact: michael.herman@advantest.com

***Abstract*— We describe a method for rapid MOSFET threshold voltage (Vt) measurement. Multiple spot Ids measurements are compared to stored reference data. Each spot measurement yields an independent Vt estimate, and these enable quality metric calculation. A Vt and quality metric can be measured within 7 msec, using two spot measurements. The method permits parallel MOS testing.**

Keywords –Test time reduction, MOS devices, parameter extraction, threshold voltage extraction

I. INTRODUCTION

The threshold voltage (Vt) metric represents the gate voltage at which MOS transistor conduction transitions from weak to strong inversion. The Vt is among several key transistor characteristics required for proper circuit simulation and designed operation. Consequently, the Vt is one of the key specifications of manufactured MOS transistors, and a required element of wafer acceptance test (WAT) measurements for commercial integrated circuit (IC) semiconductor wafers. As there are many transistor types used in IC design, each requiring its own Vt tests across an IC wafer, Vt testing is one of the most time-consuming and costly steps in wafer production and process monitoring.

Many forms of Vt measurement have been developed to accurately determine the SPICE model parameter (VT0 or VT), the threshold voltage expected if the transistor were operated at zero Vds drain-source bias [2, 3, 25]. A commonly used method is the first-derivative (FD) method "extrapolated" Vt (Vt_e or $V_{G,Si}$), found as the Vgs axis intercept of Ids-Vgs curve tangent line at its point of maximum transconductance Gm[1]. The Vt_e measurement is performed using a sweep of Vgs gate voltage in fine intervals at a constant Vds drain voltage. To obtain high accuracy of the linear extrapolation requires low Ids noise [21]. Data fitting such as linear regression of local Ids points is often used to reduce noise and Vt_e variance. In general, high accuracy requires long tester settling and hold times, and internal averaging of multiple measurements, all of which increase test time. Therefore, proposed advanced VT methods such as 2nd and higher derivative methods, which further exacerbate noise effects, are not employed in mainstream manufacturing production due to test time and associated cost concerns.

Various methods have been previously developed to reduce Vt test time. For extrapolated Vt tests, one way is to calculate Gm at each Vgs step and terminate the Vgs sweep after the Gm peak is determined. On our 2N7002 devices, using hardware acceleration, the terminated Vgs method reduced the test time to 27 to 37 msec. Terminated Vgs sweeps are not amenable to simultaneous parallel testing, as the maximum Gm point may vary. Another method to reduce test time is to perform the complete Vgs sweep in parallel on multiple MOSFETs. For parallel testing, MOSFETs must be similar in maximum Vgs level, else Vgs must be limited to the weakest transistor. In both Vgs sweep methods, suitable tester settings must be used to reduce Ids noise to a level allowing accurate Vt calculation.

In IC wafer production monitoring, an alternative fast method of tracking the Vt uses a constant Ids current level [9]. Often the current level specified for the constant current Vt (Vt_cc) is scaled by the drawn W/L ratio in planar transistor processes. Test equipment logic zeroes in on the Vgs that produces the required Ids, frequently by a binary search. Like the terminated maximum Gm method, the binary search method cannot be used for parallel Vt testing.

For rapid characterization or efficient process control, a method of faster Vt measurement with parallel DUT testing ability is desirable. Generally, increasing test speed raises measurement noise, and degrades measurement accuracy. This creates a tradeoff for production monitoring of Vt values. To

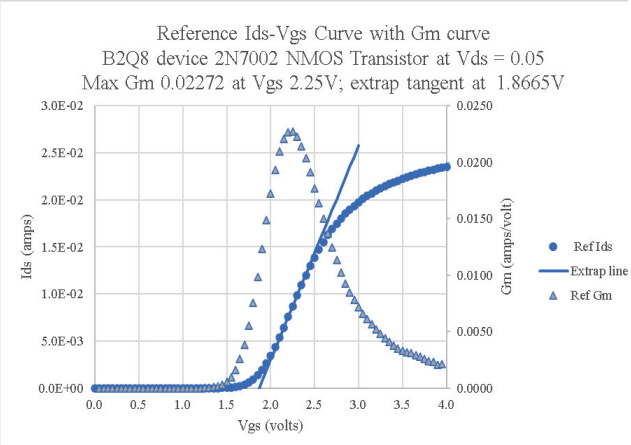


Fig. 1: Reference transistor Ids – Vgs curve showing Vgs of maximum Gm at 2.25V, and tangent intercept, Vt_e, on the Vgs axis at 1.8665 V.

[1] The term "extrapolated Vt" is labeled by Schroder both as "Vt_e" [1, p.222] and "$V_{G,Si}$" [1, pp.223-4]. This is related to the SPICE modeling parameter VT or VT0 by [1, eq. (4.106)], which we show as eq. (1).

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

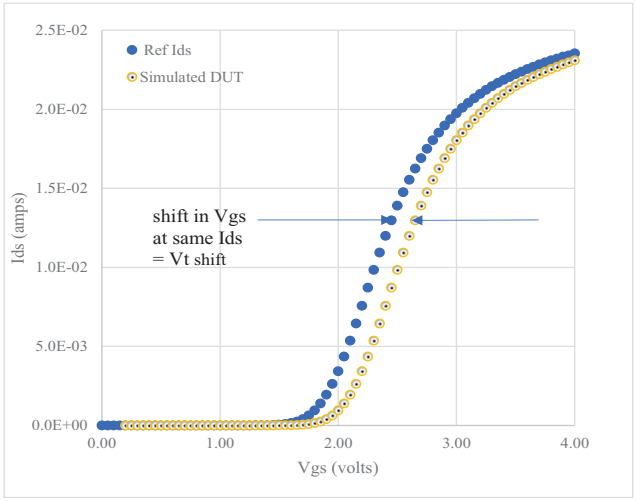


Fig. 2: Reference and ideal DUT transistor Ids – Vgs curves showing the lowered Ids expected if DUT VT is raised by 0.2V over the Reference transistor. The lateral Vgs shift is identical for every Ids value, and can be ideally determined using any single Ids point of the DUT and compared to the Reference dataset.

overcome these difficulties, this paper describes a new approach which offers the accuracy of low noise testing and the high speed from making very few DUT measurements. The high accuracy results from doing the time-consuming low noise tests only on a Reference device, then reducing other DUT test times by testing only few "spot" Ids current measurements. These data are combined to find the Vt by applying the Vgs shift between DUT and Reference. We discuss limitations of the method and the development of a Quality metric which can quantitatively diagnose the rapid Vt method effectiveness.

II. Experimental

For demonstration purposes, we tested the Ids-Vgs characteristics of two package groups of commercial 2N7002 NMOS transistors. Our tester is an Advantest 93000 tester fitted with the additional Advantest Parametric SMU8 system. That add-on subsystem provides up to 112 SMUs which can run in parallel. Each SMU can return a current or voltage measurement in 80 µsec. Any number N of those measurements can be internally averaged. Moreover, multiples or fractions of preset settling and hold times can be specified by the test engineer to improve accuracy and reduce data noise.

Various factors affect overall test time. Test time is influenced by the choices of tester settings of hold time, settling time, the time multiplier, and number of internal tester measurements to average. The test time per point is also affected by the current to be measured due to the ammeter circuit noise characteristics and response time. During the Vgs sweep on these 2N7002 NMOS devices, Ids changes from ~0.3 pA until limited by a protective clamp level set to 50 mA, some 11 orders of magnitude in current. The tester requires longer time for low Ids currents; the tester is faster in measuring larger currents. For example, on our tester with equal settings, using a binary search seeking an Ids of 40 nA takes about 11 msec, but about 9 msec seeking about 4 mA.

In this investigation, we specified conditions to achieve low Ids noise. We compared conditions of the time multiplier factor, built-in "Short" "Medium" and "Long" hold and settling times, and N=5 measurements[2]. Noise was evaluated by comparing the raw Ids data to the data smoothed by logarithmic averaging. The difference between Vt_e measurements with or without logarithmic smoothing differed by 5 mV, and the difference between Short and Long settings differed by under 0.9 mV. These small differences demonstrated that the unsmoothed data was sufficient to calculate accurate Vt_e values for these devices using Short time settings with a multiplier of 0.5.

III. Test Method

A. *Overview of the method*

In present-day testing, there are two primary methods to determine Vt. In the "constant current" method, the engineer specifies a fixed current and seeks the Vgs which causes that Ids. That Vgs is found by either a linear or binary search. At each Vgs test step, the current is measured, and the test system takes the next logical step until the current or the Vg is determined within specified tolerances. To converge within 10mV of the optimal Vgs point, about 9 test steps are required, but allowing 50mV tolerance can take 7 steps. The only data retained is the constant current Vt estimate, Vt_cc. Larger tolerance of 50mV took about 9 msec, while the 10mV criterion took about 11 msec.

The second popular method is to extrapolate the Ids curve at the point of maximum Gm. The intercept point on the Vgs axis is the extrapolated Vt, Vt_e. In this method, raw Ids-Vgs data is generally discarded after parameters like Vt are calculated. Each new transistor DUT is treated as essentially unknown, so the full Ids-Vgs range is measured or sampled for every DUT, depending upon the chosen Vt method.

In contrast, for the technique described here, we only generate a complete Ids-Vgs dataset for one representative DUT deemed an appropriate Reference device. Once that data is acquired it is stored, and subsequently read from a stored Reference dataset. The measured Reference data file includes its Vt, its optimal Vgs test point, and other metrics. Multiple Vt methods can be used, and recorded in that Reference file.

Once Reference files are available, subsequent similar DUT devices are only tested at one or a few spot Vgs bias points. The

Table 1: Measured and Calculated Terms

Measurement Description	Name of Reference Measurement	Name of DUT Measurement or Calculation
Vgs of Reference maximum Gm	Vgs_Ref_maxGm	
Ids at Vgs_Ref_maxGm	Ids_Ref_maxGm	Ids_DUT_VmaxGmRef
Vt_extrap using max Gm	Vt_Ref_maxGm	Vt_DUT_calcmaxGm
Ids at Vt_extrap	Ids_Ref_Vtextrap	
Vgs of the Reference device at Ids_DUT_RefmaxGm	Vgs_Ref_Ids_DUT	

[2] The actual durations of the named "Short" "Medium" or "Long" depend upon the current range being measured. Lower current ranges require longer measurement time due to larger RC in transimpedance current to voltage conversion.

DUT measurements are compared to the Reference. Shifts in Vgs between the DUT and Reference are applied to the known Vt of the Reference, giving the Vt of the DUT. Our tests show that if the DUT and Reference devices have similar Ids-Vgs curves, the Vt determined from the Vgs shift is accurate. We describe below a method to quantify Ids-Vgs curve similarity, allowing a logical decision whether to utilize the rapid Vt value, or to retest the DUT under other Vt test conditions.

B. Details of the method

The rapid Vt method is performed in the following manner. The test program determines if a Reference file exists from earlier recorded Ids-Vgs of a transistor typical of the same process and layout. If a Reference dataset is absent, the DUT itself is employed as a new Reference device. The new Reference is measured over the full Vgs range at the required Vds bias for Vt testing. The test conditions are tailored for each transistor type, size and orientation.

There are requirements for Reference data quality. Each Reference dataset must contain an accurate low noise Ids vs Vgs sweep over the full Vgs range. Different transistors will have noise commensurate with their gate area and process. For the commercial 2N7002 NMOS we tested, we found no difference in Ids noise or calculated Vt values over a wide range of test settings. We determined that Short settling and hold times, multiplied by the factor 0.25, and setting the tester's internal averages to 5 gave repeated Vt values with a small standard deviation.

C. Reproducibility compared to Full Vgs Sweep

We established that the chosen test conditions were sufficient by re-testing the Reference device itself as an ideally similar DUT. Testing the Reference device initially at 5 different settling times resulted in a Reference Vt of 2.462 V with a standard deviation of 1.1 mV. A subsequent rapid-Vt test of this same "DUT" using 2 spot points showed a Vt average of 2.490 V. The discrepancy between DUT and Reference (the same device) was 28 mV. We note that these Vt values were the result of using 50 mV Vgs steps for the Reference sweep. A 2-point rapid Vt test took 7.4 msec and its subsequent analysis took 0.4 msec.

D. An example of the method

To illustrate the rapid Vt method, let us examine actual data. In producing our Reference datasets and resulting "extrapolated Vt" of commercial 2N7002 NMOS devices, we measured Ids over a full Vgs sweep at 50 mV steps from 0 to 4 V, at each chosen Vds bias. On our tester, the time to measure Ids of those 201 Vgs steps varied between 430 and 615 msec at the Short setting, with 0.5 time multiplier. Reference curve acquisition time is not a strong concern as subsequent DUT measurements are made at high speed. For a nominally consistent production batch of transistors, the Reference would generally be acquired once, stored on the tester computer system, and re-used. Once data is acquired, Vt calculation time for the Reference or other DUTs takes about 0.4 msec.

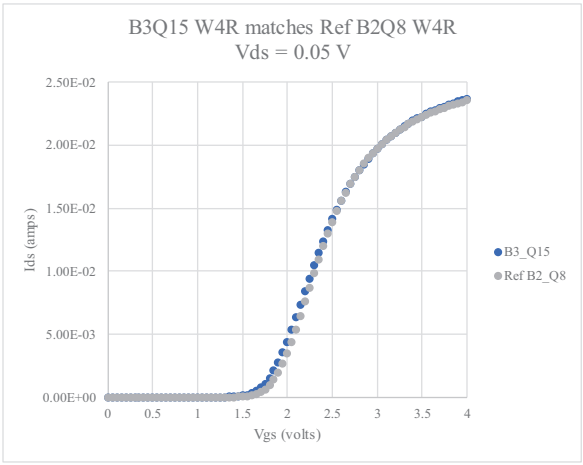


Fig. 3: Reference and DUT transistor having same top mark W4R. The Ids vs Vgs curves match closely at this Vds = 0.05V.

A Reference dataset is shown in Fig. 1 from a 2N7002 transistor, at Vds = 0.05 V, with Vgs from 0 to 4 V in 50 mV steps. The transconductance curve Gm = d(Ids)/d(Vgs) is also shown, with its extrapolated Ids tangent line at the Gm peak. On this Reference, maximum Gm occurred at Vgs = 2.25 V, with maximum Gm = 0.02272 mho at Ids = 8.7152 mA[3].

Often adjustments are applied to the extrapolated Vt to more closely estimate the Spice constant V_T. Schroder gives the formula as [1, eq. (4.106)]:

$$V_T = V_{G,Si} - V_{ds}/2 \qquad (1)$$

Hu [26, Ch 6] adds the "body charge factor" α:

$$V_T = V_{G,Si} - \alpha\, V_{ds}/2 \qquad (2)$$

Hu suggests that α has a typical value of 1.2 in IC technologies and this can be measured [25].

The Vgs axis intercept point $V_{G,Si}$ is given by:

$$V_{G,Si} = Vgs_{peakGm} - Ids_{peakGm}/Gm_{peak} \qquad (3)$$

Following (1), $V_{G,Si}$ = 1.8665 V, Vds = 0.05 V, VT = 1.8415 V.

To find the Vt of a DUT we choose a Vgs_1 bias and measure the DUT Ids current. For example, Fig. 2 illustrates the Ids behavior of a DUT identical to the Reference but having 0.2V higher Vt. A DUT current exceeding the Reference curve at a given Vgs_1 implies the DUT's Vt is below the Reference's, and vice-versa. The DUT current is then found by inspection and interpolation from the Reference Ids data, to determine the Reference Vgs_2 giving the same current. We calculate the DUT Vt using adjusting the known Reference Vt by the difference between Vgs_1 and Vgs_2.

[3] The region surrounding the Gm curve peak points can be fitted to a parabolic curve to further perfect the optimal Vgs point and the accuracy of the extrapolated intercept point.

E. *Calculating the Vgs shift*

For rapid Vt measurements of similar DUTs, we first determine the most advantageous Vgs test points. The Vgs giving highest sensitivity is found from the Reference at its point of maximum Gm.

Therefore, after we measure the Reference data, we calculate and store these key test metrics with the entire Reference Ids Vgs data in the Reference file:

1. Vds and Vbb voltages
2. Maximum Gm
3. Vgs at maximum Gm
4. Ids at maximum Gm
5. Vt values of interest and related data
    a. Vt_extrapolated ($V_{G,Si}$)
    b. Ids at Vt_extrapolated
    c. Vt (eqn. 1)
    d. Ids at Vt (eqn. 1)

The DUT Ids curve is assumed identical in shape, but shifted in Vgs, from the Reference device. With that assumption, we measure the DUT at a Vgs of high sensitivity of Ids to Vgs, which we know from the Reference point as its Vgs of maximum Gm, which we label Vgs_Ref_maxGm. The DUT Ids is then measured at Vgs_Ref_maxGm in a spot measurement using low noise test conditions. We label this value Ids_DUT_VmaxGmRef. From the Reference dataset, we determine by interpolation the Reference device Vgs for which its Ids would match the DUT Ids current, Ids_DUT_VmaxGmRef. We label this Vgs_Ref_Ids_DUT. The difference between these Vgs values is:

$$Vg\_shift = Vgs\_Ref\_maxGm - Vgs\_Ref\_Ids\_DUT \quad (4)$$

Finally, the Vt of the DUT is:

$$Vt\_DUT\_calcmaxGm = Vt\_Ref\_maxGm + Vg\_shift \quad (5)$$

For our Fig. 1 example, the Reference device has a Vt = 1.8415 V, a maximum Gm (Vgs_Ref_maxGm) = 2.25 V, and at that Vgs, Ids_= 8.7152 mA. At Vgs_Ref_maxGm = 2.25 V, the DUT Ids = 4.3518 mA, a lower current. The Reference data shows that lower current 4.3518 mA at Vgs = 2.05 V. We label the value 2.05 V as Vgs_Ref_Ids_DUT. The Reference required a lower Vgs to achieve the DUT Ids. Following (4) Vg_shift = 2.25V – 2.05 V = 0.2 V. From (5), Vt_DUT_calcmaxGm = 1.8415 V + Vg_shift = 2.0415 V. The DUT Vt has been determined using a single spot Ids measurement by re-using the complete Vgs-Ids Reference device dataset. A single spot measurement can allow calculation of the Vt in about 5 msec, about half the time required for the binary search of a constant current Vt method.

IV. Determining Transistor Similarity

For accurate Vt determination, the rapid Vt method requires close similarity between the DUT and Reference devices. The L, W, and other physical layout and process parameters of the Reference and DUT must be nominally identical. We will show that the similarity between the DUT and Reference device can be quantified for each DUT by using multiple spot measurements, at Vgs points surrounding the Reference's point of Gm_{max}.

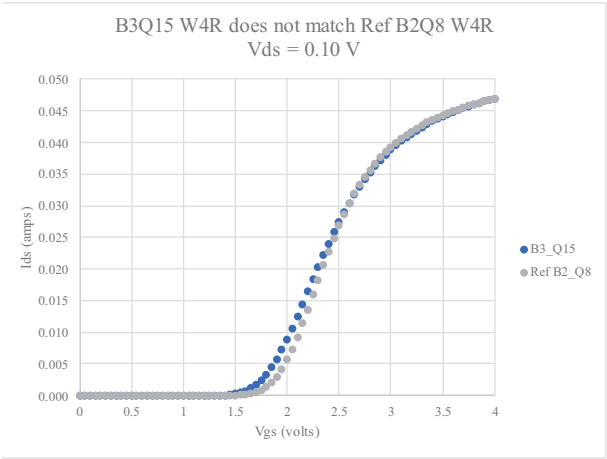


Fig. 4: Reference and DUT transistor having same top mark W4R. The Ids vs Vgs curves appear mismatched at this Vds of 0.10 V.

A. *Measuring the Validity of Spot Vt Estimates*

The applicability of the rapid Vt spot method depends on the similarity of Ids-Vgs curve shapes between the DUT and Reference datasets. We seek a way to quantify that similarity, so automated logic can decide whether to use or reject the rapid Vt value. One possible Goodness of Fit metric is the variance of a Vt estimates derived from separate Vgs intervals. The variance will depend on the chosen spacing of Vgs test points. Consequently, to use variance as a discriminator requires a specific Vgs step size around the Reference Gm_{max} point. To determine a variance also will generally require 3 or more spot measurements. The use of the variance adds constraints on Vgs spacing, and adds a minimum of 3 spot measurements to the test engineering effort.

A better Goodness of Fit metric, independent of Vgs step size, utilizes the slope of the Vt estimates at each of multiple Vgs bias points. If the slope of the Vt vs Vgs data is small, the DUT and Reference curves are nearly identical in shape around the Gm_{max} point. If the slope is large, the DUT curve shape does not match the Reference. A slope can be constructed in as few as 2 points. In the case of a large slope metric, the test system can logically decide that the DUT requires a new Reference. Consequently, the test system could run a full sweep of the DUT

TABLE 2: MEASURED SLOPE METRIC AND TOP PACKAGE MARKINGS

DUT#	Slope metric d(Vt_e)/d(Vgs)	Top Mark
Reference B2_Q8	0	W4R
B4_Q8	-0.0734	W4R
B4_Q15	0.0075	W4R
B4_Q9	0.3541	702
B4_Q10	0.1944	702
B4_Q11	0.1755	702
B4_Q12	0.1883	702
B4_Q13	0.1726	702

Note that the different package top marks are distinguished by their Slope metric. A metric < 0.1 indicates good matching to the Reference device; poorly matching DUTs have a metric > 0.1.

and store that data as an independent Reference curve for other similar DUTs.

Slope metrics from 2N7002 NMOS measurements are presented in Table 2. The values from 2 measurements matches data from 3 to 7 spot measurement points, all surrounding the VgsmaxGm of the Reference device, using 0.1V spaced Vgs steps. As 2-point spot measurements are fastest and suffice to provide both the Vt_e and a valuable Slope as Quality metric, we find the 2 point spot rapid Vt method sufficient. Each additional Vgs point increases test time by about 1 msec. we find the 2 point spot rapid Vt method sufficient. Each additional Vgs point increases test time by about 1 msec.

V. Test time Comparisons

In our system using the built-in Short hold and settling time settings, the average test time shows about 3 to 4 msec baseline, and each spot measurement adds nearly 0.9 msec per point. The required conditions for this new method can be assumed fulfilled in on-wafer, or in-lot wafer testing. Table 3 illustrates the total test time required for between 1 and 7 spot Vt measurements. A single spot measurement can suffice for Vt calculation in as low as 5 msec, but does not provide the slope goodness of fit metric. The rapid Vt method with 2 to 7 measurements provides slope metrics and can speed up Vt testing significantly compared to other Vt methods. Testing 2 to 7 spot measurements, permitting the slope to determine validity, takes 6 to 10 msec. These test times show the remarkable progess made over the decades in transistor test times, which were reported to take on the order of a second in 1988 [16].

Measurement time dominates the overall Vt test. Subsequent calculation of the Vgs shift adds about 0.4 msec. Time to read an existing Reference file is around 1 msec; the Reference file data can be installed in SRAM memory to have faster access. These short access times allow for multiple Reference files to be used, so the test system can determine the Reference dataset providing the best Goodness of Fit. For production IC wafer lots, it may be possible to use a single Reference for each wafer, or possibly one Reference per wafer lot for some transistor types.

VI. Drain Voltage Effects on Similariy

We observe diminished similarity between Reference and DUT devices as drain voltage increases. Compare Fig. 3 at Vds of 0.05 V, to Fig. 4 at Vds of 0.10 V. Fig. 3 shows the two transistor curves appear identical at Vds of 0.05 V, and the rapid Vt method produced a low error between Vt_e values calculated from a full Vgs sweep, and low slope using 2 or more spot measurements. As drain voltage increased to 0.10 volts, shown in Fig. 4, the Ids curves show increasingly different shapes and higher slope metric.

If a discrimination level were imposed on the slope metric, the rapid Vt method could be programmed to create a new Reference curve. Changes in Ids-Vgs shape might be expected due to DIBL or channel length modulation effects for short channel devices. Fig. 5 shows the Slope metric vs Vt_e difference between the Full Sweep and the 2 point rapid Vt_e values, on the same 2N7002 transistors, resulting from increased Vds values. The graph demonstrates the similarity metric, and the Vt_e difference calculated, both worsen as Vds increases.

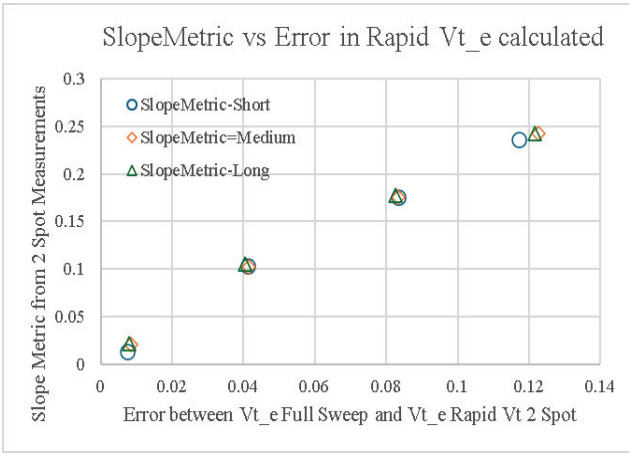


Fig. 5: Slope as Goodness of Fit Metric worsens as Vds increases, along with the difference between the Full Vgs Sweep and the 2 Spot Measurement Vt e values.

As our full sweep utilized 50 mV Vgs steps, we accepted an error of up to 50 mV in Vt. The Slope level of 0.1 distinguished that level of Vt quality. That is, if the Slope were 0.1, the error would be about 40 mV, within the 50 mV Vgs measurement step size. Table 2 shows that 2N7002 devices having 2 different top markings, were separated by their Slope metric of similarity to the Reference device marking. If the part had the same top mark, its Slope metric was below 0.1. If the part had a different top mark, its Slope metric was above 0.1. These two device groups would require separate Reference datasets.

VII. Further Advancements

The spot Vt method runs quickly. Using 2 spot DUT measurements, with a single Reference produces an accurate Vt and a Goodness of Fit metric, closely matching the Vt of a full Ids-Vgs sweep, in about 7 msec. As importantly, the spot rapid Vt method can run in parallel on multiple MOS devices even sharing a common Gate node, capable of further speeding production wafer testing.

Additional insight of the MOSFET process can result from rapid Vt data collection. For example, Vt vs Vbb back-bias data permits calculation of the channel doping [1, Sec. 2.3.2, p. 81]. Alternatively, the rapid Vt method may facilitate wafer level reliability tests (WLR) by rapid assessment of Vt shifts from transistor stresses such as hot carrier degradation [6] or NBTI [20]. These possibilities remain to be investigated.

TABLE 3: 2N7002 NMOS Vt measurement and calculation time using Advantest Parametric Test System

Vds (V)	Vt Method total measure and calculation time (msec) FS = Full Sweep 0.05V steps, TS = Terminated Sweep, CC = Constant Current, SP = Spot									
	FS	TS	CC	SP 1pt	SP 2pt	SP 3pt	SP 4pt	SP 5pt	SP 6pt	SP 7pt
0.05	540	32	11	5.7	6.7	7.1	7.6	8.2	8.6	9.2
0.10	539	33	11	5.1	7.8	8.5	9.1	9.5	10.4	10.7
0.15	539	32	11	4.9	6.1	6.3	6.9	9.4	10.2	10.9
0.20	540	33	11	5.2	5.9	6.6	6.9	7.9	8.1	10.7
Quality metric?	No Quality metric				Quality metric capable					

VIII. SUMMARY

A rapid method of determining transistor Vt was described. The method compares a few spot Ids measurements on the DUT to more extensive Ids-Vgs data stored as the Reference from similarly manufactured transistors. The method accuracy was demonstrated by the small difference between Vt measurements made by the rapid Vt method to those of the full Vgs sweep. The rapid Vt method requires previously stored low noise full Ids vs Vgs sweep dataset from a Reference transistor. In our tests, 201 Vgs points took about 580 msec, versus about 7 msec for a rapid Vt using 2 spot values.

From the Reference dataset, any Vt method of engineering interest can be calculated [1, 4, 10]. These include Vt values from a Constant Current, first derivative extrapolated from maximum Gm, 2nd derivative [6], 3rd derivative [14], or a ratio method [d(gm/Id)/dVg] [13, 17, 23]. After the Reference dataset and its Vt values are stored, new DUTs are measured at 2 or more values of Vgs using fast spot measurements, taking about (5 + Npoints) msec on our system. We evaluated Ids using Npoints = 2 chosen at Vgs = Vgs_maxGm ± (ΔVgs/2), using ΔVgs = 0.1V, to determine Vt values in about 7 msec on our Advantest Parametric Test system.

While a single Vgs spot measurement can provide a Vt estimate in about 6 msec, it is beneficial to add multiple measurements, permitting the Slope of estimated Vt vs Vgs to form a valuable Goodness of Fit metric. For the transistors we evaluated, a Slope Goodness of Fit metric value below 0.1 ensured reasonable applicability of the rapid Vt method. As seen in Fig. 5 the error between the exact full Vgs sweep Vt and the rapid Vt values were under 45 mV, comparable to the 50 mV Vgs steps in the Reference data sweep.

In IC production, transistors on each wafer receive the same process and lithography, which should ensure close similarity in Ids-Vgs characteristics. The rapid Vt method using 2 or more spot Ids measurements should therefore be amenable for MOS Vt testing on IC wafer production.

IX. REFERENCES

[1] D. K. Schroder, *Semiconductor Material and Device Characterization*, 3rd edition, Hoboken, New Jersey, USA: Wiley-Interscience/IEEE Press, 2006.

[2] D. Foty, *MOSFET Modeling with Spice: Principles and Practice*, Upper Saddle River, NJ, USA: Prentice Hall, 1997.

[3] G. Massobrio and P. Antognetti, *Semiconductor Device Modeling with SPICE*, 2nd edition, New York, NY, USA: McGraw-Hill, 1993.

[4] A. Ortiz-Conde, F.J. Garcia Sanchez, J. J. Liou, A. Cerdiera, M. Estrada, and Y. Yue, "A Review of Recent MOSFET Threshold Volts Extraction Methods," *Microelectr. Rel*. vol. 42, pp. 583-596, Apr.-May 2002.

[5] S. C. Sun and J. D. Plummer, "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces," *IEEE Trans. Electr. Dev*. Vol. ED-27, pp. 1497-1508, Aug. 1980.

[6] H. S. Wong, M. H. White, T. J. Krutsick, and R. V. Booth, "Modeling of Transconductance Degradation and Extraction of Threshold Voltage in Thin Oxide MOSFETs," *Solid State Electron*. vol. 30, pp. 953-968, Sep. 1987.

[7] S. W. Tarasewicz and C. A. T. Salama, "Threshold Voltage Characteristics of Ion-implanted Depletion MOSFETs," *Solid State Electron*. vol. 31, pp. 1441-1446, 1988.

[8] G. Ghibaudo, "New method for the extraction of MOSFET Parameters," *Electron. Lett*. vol. 24, pp. 543-545, Apr. 1988.

[9] H. G. Lee, S. Y. Oh, and G. Fuller, "A Simple and Accurate Method to Measure the Threshold Voltage of an Enhancement Mode MOSFET," *IEEE Trans. Electr. Dev*. vol. ED-29, pp. 346-348, Feb. 1982.

[10] K. Terada, K. Nishiyama, and K-I Hatanaka, "Comparison of MOSFET-Threshold Voltage Extraction Methods," *Solid State Electron*. vol. 45, pp. 35-40, 2001.

[11] F. M. Klaassen and W. Hes, "On the Temperature Coefficient of the MOSFET Threshold Voltage," *Solid State Electron.* vol. 29, pp. 787-789, Aug. 1986.

[12] C. Hao, B. Cabon-Till, S. Cristoloveanu, and G. Ghibaudo, "Experimental determination of Short Channel MOSFET Parameters," *Solid-State Electron.,* vol. 28, no. 10, pp. 1025-1030, 1985.

[13] T. Rudenko, M. K. Md Arshad, J.-P. Raskin, A. Nazarov, D. Flandre, V. Kilchystska, "On the gm/Id-based approaches for threshold voltage in advanced MOSFETs and their application to ultra-thin body SOI MOSFETs," *Solid-State Electron*. vol. 97, pp. 52-58, 2014.

[14] G. Espiñera, D. Nagy, A. García-Loureiro, N. Seoane, G. Indalecio, "Impact of threshold voltage extraction methods on semiconductor device variability" *Solid State Electron.*, vol. 159, pp. 165-170, Sep. 2019.

[15] G. Moore, J. -H. Liao, S. McDade and B. Verzi, "Accelerating 14nm device learning and yield ramp using parallel test structures as part of a new inline parametric test strategy," *Proc. 2015 Intern. Conf. on Microel. Test Struct.*, Tempe, AZ, USA, pp. 44-49, 2015, doi: 10.1109/ICMTS.2015.7106106.

[16] H.P.Tuinhout, S. Swaving, J. J.M. Joosten "A Fully Analytical MOSFET Model Parameter Extraction Approach", *1988 IEEE Proceedings on Microelectronic Test Structures,* vol. 1, no. 1, pp. 79-84, Feb. 1988.

[17] K. Ahmed, "Extracting Mobility and Threshold Voltage from Noisy Current-Voltage Data," [Online]. Available: TechArchiv. 24 Jan 2020 — CC-BY 4.0 — https://doi.org/10.36227/techrxiv.21905007.v1

[18] T. Rudenko, V. Kilchytska, M.K.Md Arshad, J.-P. Raskin, A. Nazarov, and D. Flandre, "Influence of Drain Voltage on MOSFET Threshold Determination by Transconductance Change and gm/Id methods," IEEExplore.

[19] N. Makris, M. Bucher, "On MOSFET Threshold Voltage Extraction Over the Full Range of Drain Voltage Based on Gm/ID," [Online]. Available: ARXIV.org/abs/2106.00747.

[20] B. Djezzar, et al "Recovery investigation of NBTI -induced traps in n -MOSFET devices," *Microel. Rel.* vol. 110, pp. 113703, 2020. ISSN 0026-2714, https://doi.org/10.1016/j.microrel.2020.113703

[21] C. Stankus, M. Ahmed, "Comparing smoothing techniques for extracting MOSFET threshold voltage," *Solid State Electr*. vol. 164 (2020) pp. 107744, 2020, Elsevier, https://doi.org/10.1016/j.sse.2019.107744.

[22] Z. X. Yan, M. J. Deen, "Physically-based method for measuring the threshold voltage of MOSFETs," *IEE Proceedings-G*, vol. 138, no. 3, pp. 351-357, June 1991.

[23] D. Flandre, V. Kilchytska, and T. Rudenko, "gm/Id Method for Threshold Voltage Extraction Applicable in Advanced MOSFETs With Nonlinear Behavior Above Threshold," *IEEE El. Dev. Lett*, vol. 31, no. 9, pp. 930-932, 2010.

[24] F. J. Garcia Sanchez, A. Ortiz-Conde, J. A. Salcedo, J. J. Liou and Y. Yue, "Extraction of the bulk-charge effect parameter in MOSFETs," *2000 22nd Intern. Conf. on Microelectr. Proc.* (Cat. No.00TH8400), Nis, Yugoslavia, vol.1, pp. 133-136, 2000, doi: 10.1109/ICMEL.2000.840539

[25] C. Hu, "Modern Semiconductor Devices for Integrated Circuits" [Online]. Available: https://www.chu.berkeley.edu/modern-semiconductor-devices-for-integrated-circuits-chenming-calvin-hu-2010/

[26] See for example, BSIM 4.1 manual. [Online]. Available: http://www2.eecs.berkeley.edu/Pubs/TechRpts/2000/ERL-00-48.pdf

Statistical investigation of SnOx RRAM memories for switching characteristics

Alin Panca[1], Alexantrou Serb[1], Spyros Stathopoulos[1], Themis Prodromakis[1]
[1]Institute for Integrated Micro and Nano Systems, University of Edinburgh, UK, EH9 3JU
Contact: Alin.Panca@ed.ac.uk

***Abstract*—Resistive Random Access Memory (RRAM) has seen significant developments in the past years. An important improvement is to reduce the device-to-device variability that limits RRAM implementation. A proposed approach is to measure a high number of devices to statistically identify trends and evidences behind the cause. Here, we present a method to characterise the RRAM switching behaviour and analyse a large dataset of SnOx RRAM, in a statistical fashion for further optimisation and modelling.**

I. Introduction

THE progression of RRAM [1] in commercial applications and circuit design is dependent on reliable models to account for their variability on large scale. Automated tests of RRAM devices would require specialized functions and algorithms to extract their features on a roll-to-roll process to make them more commercially-friendly. General-purpose testing procedures are established through instrumentation as Keithley [2] and Agilent [3]. Moreover, automated RRAM characterisation routines have been proposed through specialised tools as ArC Instruments [4][5][6], acquiring large datasets. A supporting element is an automated testing to additionally provide statistical significance to wafer-scale characterisation, that could direct to further fabrication optimisation and more variability control. In this work, we present a software module that performs fast chained data analysis (filtering, data extraction, fitting, goodness-of-fit and plotting) on a large dataset resulted from an automated RRAM characterisation [4] and subsequently, could be integrated in the ArC software.

II. Test Structures and Measurement Set-up

Our testing system involves a Cascade Summit 12000 semi-automatic probe station connected to the ArC1 through a specialised module, to allow automated wafer-level characterisation with minimum intervention, fig. 1. The system operates a 32×32 crossbar array probecard and can navigate independently to any wafer location and perform intensive RRAM characterisations until the dataset is complete. The user selects the wafer area to be characterised, and the ArC1 testing protocols, such as retention, electroforming, switching etc. The automated ArC set-up is easily transferable to other semi-automatic probe stations, allowing more flexibility in using the ArC1. Then, the resulted data can be loaded to the standalone post-characterisation module, that processes it into histograms and complementary statistics, where the user inputs several parameters as follows (fig. 2).

The module presented is a data analysis tool, with a top-down sequential structure starting with browsing and loading the dataset file resulted from the automated process in fig. 1. Then the file undergoes: i) filtering, ii) parameter extraction, iii) clustering, iv) fitting, v) plotting.

i) The file is filtered by wafer location in *quadrants* e.g. Top-Right (TR), but this can be simply extended to any pattern presets. Then, the device can be filtered by *subdie* number, containing variations of the device structures. Next, a *function* filter extracts the desired protocol data, as dataset can have a string of characterisation protocols. The available options in the interface are set visible according to the data available in the file. Additional filters are used, particularly if the data is made of a chain of repeated function that need to be identified.

ii) Next, the *extraction* step, decides and computes one parameter to be explored/device, from the filtered dataset e.g. formed voltage when the resistance drops by 25% from the initial resistance. This step can include any computation delivering a final parameter/device. Then, the extracted data is split into clusters that identify distinct distributions, whereas the points where the distributions split, are calculated using K-Means algorithm or Gaussian Mixture Model (GMM) - machine learning methods detailed [7]. Mostly, the difference between the models is: K-Means has a hard bin separation between distributions - useful for non-overlapping clusters, whereas GMM considers a region of overlapping bins belonging to more distributions - sometimes offering a more accurate clustering. As a result, clustering prepares the dataset for optimal individual distribution fitting.

The optimal number of clusters is identified using both Silhouette score and K-Means [7]. The clusters are generated using the K-Means algorithm, which inputs a predefined number of k clusters (starting with $k=1$). It works by randomly picking a single centroid in the distribution, and calculates the sum squared error (SSE) for each data point. The centroid is then moved until reaching the maximum number of centroids and SSE reaches its minimum, fact that identifies separation points between each cluster. Then, the Silhouette algorithm gives a score of how similar a data point belongs to its own cluster (0 to 1) or a neighboring one (0 to -1). By averaging each data point Silhouette score in each cluster, we obtain an individual score for each k. By repeating the K-Means with

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

an incremental k, we pick the one with the highest Silhouette score.

iii) The clustered data undergoes a parametric fitting process, on roughly 90 known distributions, where the best fitting distribution is decided on one of the following statistical technique where appropriate: SSE, Kolmogorov-Smirnov (KS), Bayesian information criterion (BIC) or Akaike information criterion (AIC). The number of distribution fit trials could be reduced, or targeted on a specific one, to improve the processing time.

iv) Plotting can be in the form of 1D histograms, or 3D graph to represent the Gaussian parameters - mean (μ), deviation (σ), amplitude (A) for more than one device size variation (subdie). The algorithm exports the stats of each device size in terms of cluster, Gaussian parameters, best fitted distribution, distribution parameters, and goodness-of-fit (KS Test).

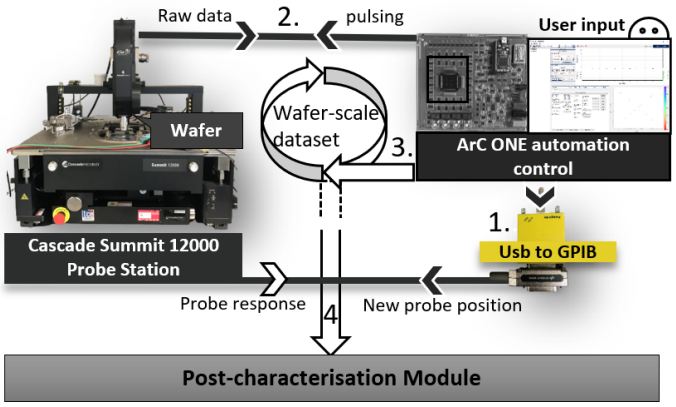


Fig. 1. Automated RRAM characterisation using ArC1 instrumentation, Cascade Summit 12000 probe station and custom-made modules to achieve large-scale measurements and analysis [4]. Routine starts with the user input parameters, then ArC1 communicates movement commands to Cascade (1), ArC1 receives data during voltage pulsing routines (2), and builds the dataset (3). Last, the dataset is sent to the analysis module (4).

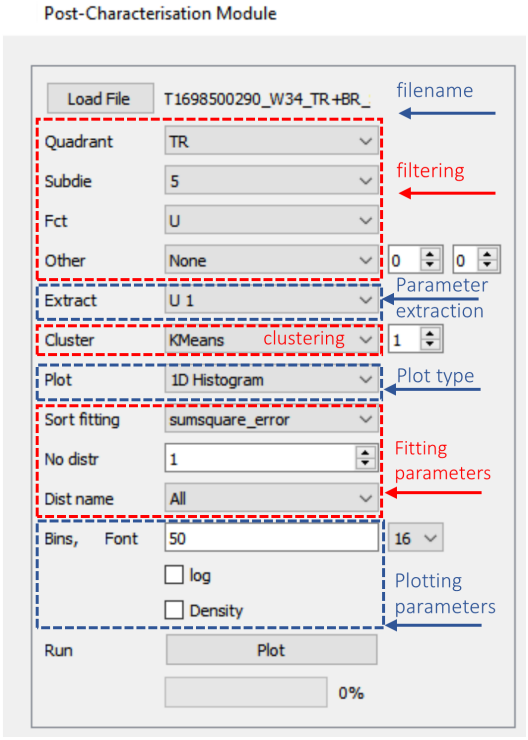


Fig. 2. This module operates in a sequence starting with data filtering by wafer location, then continues with identifying the target RRAM functions. After, it extracts a particular parameter e.g. switching voltage according to various conditions e.g. rate of change $>10\%$. Then, statistical analysis is done to identify data clusters and perform distribution fitting. Lastly, a histogram plot is shown, identifying cluster stats, operational device counts and fitting results.

III. Demonstration of Results

In this section we present a test measurement on two wafer quadrants with electroforming-free single-array RRAM stacks: **1.** *W (40 nm)/SnOx (30 nm)/Pt (15 nm)* and **2.** *W (40 nm)/SnOx (30 nm)/W (40 nm)* (leftmost material is the bottom electrode), fig. 3. The fabrication process starts on a Si wafer, dry oxidated with 200 nm SiO_2. The SnOx layer is RF sputtered with gas flow Ar/O_2: 20/1 sccm, power 70 W, and chamber pressure 1 mTorr at room temperature (RT). W and Pt electrodes are RF-sputtered at RT. There are 22 dies/quadrant, each with 9 subdies ranging from 1 to 60 μm^2 of 32 devices/subdie.

There are two characterisation protocols used in this demonstration: Resistance Read Uniformity (RRU) and Resistance Switching Uniformity (RSU). Briefly, RRU performs an initial resistance read at low voltage for a defined number of times, whereas RSU runs a switching dynamics algorithm [6] where an alternating series of pulses switches devices between high- and low-resistive states. The module is tested with a sequence of RRU and RSU as follows: 1^{st} *RRU* - 1^{st} *RSU* - 2^{nd} *RRU* - 2^{nd} *RSU* - 3^{rd} *RRU*, on approximately 700 devices per stack. Then, each sequence step result is fed to the post-characterisation module for statistical calculations. The clustering used was K-Means, and best fitting was based on SSE from 89 model distributions. The processing can be run immediately after the measurement is completed, taking several seconds to visualise and export stats for further interpretation. Table I shows the fitted distribution stats of each RRU from the characterisation sequence, for the most significant cluster identified, in both available stacks **1** and **2**. Then, tables II-III show the stats of each RSU run for stacks **1** and **2**, respectively, split in one or three clusters.

The KS test [8] verifies if the fitted distribution to each data cluster does not reject the null-hypothesis, proving that it belongs to a specific parametric distribution e.g. Cauchy, whenever the resulting p value is greater than 0.05. Generally, the KS method verifies if a given distribution is Gaussian, however, in our approach, we pair the KS method with reference datasets from other known distributions. Additionally, table IV presents exported stats and fitting on the initial resistance

reading per all device sizes and clusters, in stack 1 only, that passed the goodness-of-fit KS test.

1) Resistance Read Uniformity: This function averages 50 initial resistance reads/device and plots a distribution of the log resistance. This has direct utility in evaluating the general operational uniformity over the wafer and spot fabrication defects as shorted or damaged devices. A successful fitting can provide a unit of comparison and predictability between fabrication optimisation trials. For example, the module identifies three data clusters in fig. 4 with this possible interpretation: cluster $\mu_0 = 3$ shows possible shorted devices, μ_1= 5.37 includes active devices, μ_2= 9 includes not connected (e.g. electrode or contact failure) or non-switching devices.

TABLE I
EXTRACTED UNIFORMITY RESISTANCE STATISTICS ON TWO STACKS: 1) W/SNOX/PT AND 2) W/SNOX/W DEVICES, FOR SIZE 20 μm, THREE TIMES, IN BETWEEN TWO SWITCH SEEKER RUNS.

Stack	RRU run	μ	σ	A	Distribution	KS Test
1	1^{st}	5.38	0.43	0.67	Cauchy	**0.18**
1	2^{nd}	5.40	0.42	0.65	Cauchy	0.004
1	3^{rd}	5.37	0.42	0.65	Cauchy	**0.07**
2	1^{st}	5.54	0.38	0.65	Cauchy	0.004
2	2^{nd}	5.35	0.40	0.66	Cauchy	0.006
2	3^{rd}	5.31	0.37	0.70	Cauchy	0.007
Average:		5.48	0.87	0.83		0.04

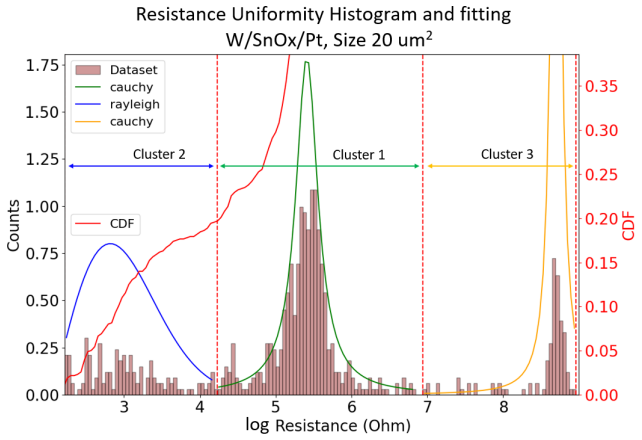


Fig. 4. It presents the distribution of resistance log exponents of the 3^{rd} RRU run, of stack 1) W/SnOx/Pt, device size of 20 μm^2, including best parametric fitting for each automatically identified cluster. The *Cauchy* cluster passes the KS test with p value 0.07

2) Resistance Switching Uniformity: The switch seeker takes a ramp voltages of ± 1 to 5 V, of 0.1 V steps, in 20 ns pulses, and 50 cycles/device. The post-characterisation processor extracts the averaged minimum absolute voltage, where in at least 20 cycles, a resistance change $>10\%$ is detected, fig. 6 for a visual data interpretation from the ArC software, after a test measurement. This function identifies the optimum switching voltage and volatility behaviour, useful for designing and calibrating peripheral circuits [4]. Using the post-characterisation tool, we process the dataset into clustered histograms as in figure 5 and we perform distribution fitting by cluster. Then, we extracted the stats in tables II-III. Notice that few fittings are slightly offset (graphically only) due to the effort to include all clusters in one plot, and further plotting improvements will be addressed in the future.

TABLE II
EXTRACTED UNIFORMITY SWITCHING STATISTICS ON STACK 1. W/SNOX/PT DEVICES, FOR SIZE 20 μm ON THE TWO POSITIVE RSU RUNS. EACH RUN IDENTIFIES STATS FOR DATA SPLIT IN 1 OR 3 CLUSTERS.

RSU run	Cluster	μ	σ	A	Distribution	KS Test
1^{st}	1/1	1.44	0.39	1.00	gamma	1E-11
1^{st}	1/3	1.18	0.09	0.59	invgauss	2E-4
1^{st}	2/3	1.62	0.15	0.30	beta	**0.94**
1^{st}	3/3	2.29	0.22	0.11	exponnorm	**0.35**
2^{nd}	1/1	1.51	0.50	1.00	gamma	8E-04
2^{nd}	1/3	1.20	0.11	0.62	norminvgauss	3E-03
2^{nd}	2/3	1.76	0.20	0.28	burr12	**0.85**
2^{nd}	3/3	2.70	0.40	0.10	nakagami	**0.99**

TABLE III
EXTRACTED UNIFORMITY SWITCHING STATISTICS ON STACK 2. W/SNOX/W DEVICES, FOR SIZE 20 μm ON THE TWO POSITIVE RSU RUNS. EACH RUN IDENTIFIES STATS FOR DATA SPLIT IN 1 OR 3 CLUSTERS.

RSU run	Cluster	μ	σ	A	Distribution	KS Test
1^{st}	1/1	1.57	0.47	1.00	gamma	1.00E-43
1^{st}	1/3	1.16	0.08	0.43	gibrat	5.23E-06
1^{st}	2/3	1.61	0.15	0.35	johnsonsb	**0.97**
1^{st}	3/3	2.30	0.30	0.22	dweibull	6.94E-03
2^{nd}	1/1	1.55	0.46	1.00	powerlaw	1.00E-23
2^{nd}	1/3	1.19	0.11	0.50	alpha	9.39E-06
2^{nd}	2/3	1.73	0.16	0.36	nakagami	**0.77**
2^{nd}	3/3	2.41	0.28	0.14	chi	**0.88**

TABLE IV
EXTRACTED UNIFORMITY RESISTANCE STATISTICS ON W/SNOX/PT DEVICES FOR SIZES AND CLUSTERS THAT PASSES THE KS P VALUE$>$0.05.

Size (μm^2)	Cluster	μ	σ	A	Distribution	KS Test
1	1/3	6.05	0.40	0.64	normal inverse	**0.24**
2	1/3	5.86	0.38	0.63	normal inverse	**0.29**
5	1/3	5.65	0.47	0.60	normal inverse	**0.08**
10	1/3	5.47	0.37	0.62	johnsonsu	**0.30**
20	1/3	5.37	0.43	0.67	cauchy	**0.18**
30	1/3	5.10	0.66	0.57	genhyperb	**0.20**
40	1/3	5.10	0.57	0.58	genhyperb	**0.14**
Avg:		5.51	0.46	0.61		
1	2/3	8.60	0.38	0.35	normal inverse	**0.24**
2	2/3	8.58	0.44	0.36	genhyperb	**0.29**
Avg:		8.59	0.41	0.35		

IV. DISCUSSION

The experiment presented above aims to observe any statistical discrepancies in between RRU and RSU runs as

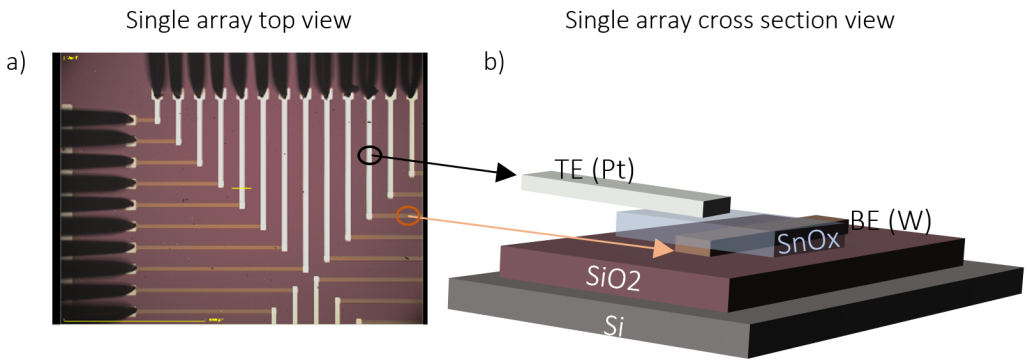


Fig. 3. Illustrates device vertical structure of individual W/SnOx/Pt devices, in one of the 23 available subdies, used in the measurement in: a) top view - microscope iamge; b) cross section view.

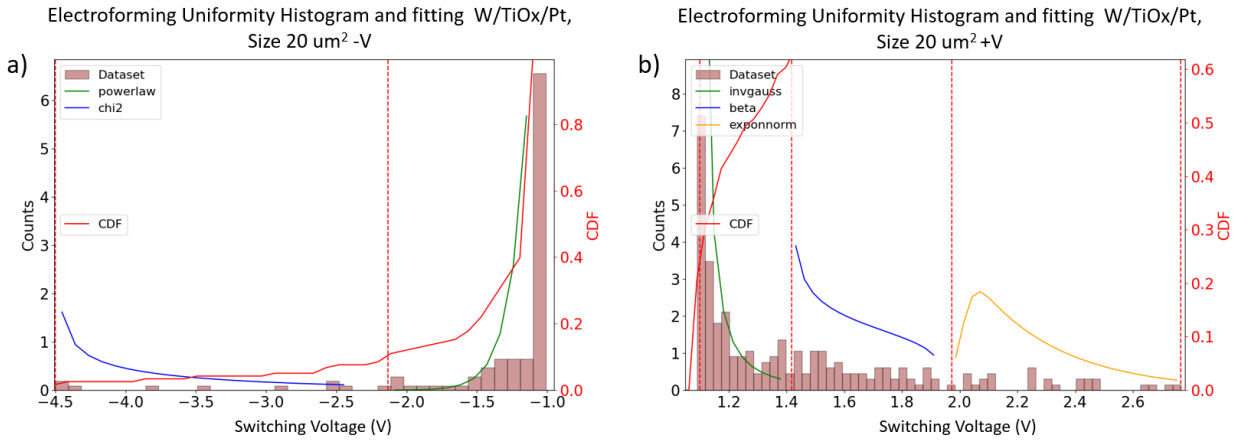


Fig. 5. Histogram showing the distribution and best fitting of average: a) positive- and b) negative-switching "stimulation voltages recorded when the device-under-test exhibits a 10% resistive state displacement from the start of each stimulation ramp" ("threshold voltages"). Only considered a valid average if the 10% threshold is passed in at least 20/50 ramps, thus recording at least 20x such threshold voltages.

well as demonstrating the operation of an automated post-characterisation tool that generates appropriate clustering and fitting over a significant number of devices. In table I we observe no significant variation between the individual characterisation runs of each stack. For example, stack 1 mean varies at maximum 17.5 $k\Omega$, whereas deviation remains almost identical among all three runs. Similarly, this is observable in stack 2. The Cauchy distribution identified was consistent in all runs, however the KS test was passed only in stack 1.

In tables II -III, if the data is not clustered (Cluster 1/1), then the distribution fitting could not be related to any parametric distributions, fact reflected in the very low KS p values (1E-11 and 1E-43). However, when the dataset is clustered, the components begin to resemble distributions, as cluster 2. Overall, there is a 5% increase observed on μ and σ, from 1^{st} to 2^{nd} run in the corresponding clusters. However, data found in the 3^{rd} cluster may not be sufficient for fitting, therefore the KS test result could be superficial. Similarly, in stack 2, table III, the same μ and σ shift are observed as in stack 1, however, the distributions identified vary. Also, there is higher tendency for devices to switch on positive bias as in figure 5b compared to negative bias switch, 5a, as the device counts is significantly higher in the former figure. Lastly, observing different device sizes, in stack 1, table IV, the most significant cluster in each size (highest A) produces a good fit with varied distributions. Moreover, from the mean, it results that the higher the device size, the lower the resistance mean, indicating a higher change of conductive filaments to be formed when the device area is higher.

V. Conclusions

This work brings a supplement to the automated RRAM testing capability that aims to increase the testing yield and automate the process development for providing a new approach on device variability and fabrication optimisation. This has a direct impact on developing models for an efficient IC design implementation and could expand to an integrated fabrication routine for quality tests industry batch production. Further improvements needs to be addressed as an automated way to link particular distribution trends to causes and give indication of fabrication parameter adjustment. We expect that the post-fabrication routine to benefit from a larger set of RRAM testings on various structures and device stacks with a direct implementation of such system in ArC instrumentation.

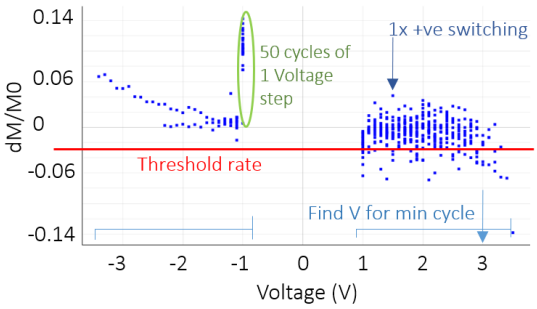


Fig. 6. RRAM switch seeker 2D graph, with memristance rate of change over a range of voltages for a repetition of 50 cycles. This plot visualize the switch seeker function result of one W/SnOx/Pt device.

REFERENCES

[1] R. Waser and M. Aono, "Nanoionics-based resistive switching memories," *Nature Materials*, vol. 6, no. 11, pp. 833–840, Nov. 2007.

[2] *Keithley Instruments (2009), 4200.* [Online]. Available: https://download.tek.com/manual/4200-900-01K$_{Feb2017_{User.pdf}}$

[3] *Agilent Technologies (2022), B1500A.* [Online]. Available: https://www.keysight.com/gb/en/assets/7018-01289/data-sheets/5989-2785.pdf

[4] A. G. Panca, A. Serb, S. Stathopoulos, S. K. Garlapati, and T. Prodromakis, "Automated rram measurements using a semi-automated probe station and arc one interface," in *2023 35th International Conference on Microelectronic Test Structure (ICMTS)*, 2023, pp. 1–4.

[5] R. Berdan, A. Serb, A. Khiat, A. Regoutz, C. Papavassiliou, and T. Prodromakis, "A μ-controller-based system for interfacing selectorless rram crossbar arrays," *IEEE Transactions on Electron Devices*, vol. 62, no. 7, pp. 2190–2196, 2015.

[6] S. Stathopoulos, L. Michalas, A. Khiat, A. Serb, and T. Prodromakis, "An Electrical Characterisation Methodology for Benchmarking Memristive Device Technologies," *Scientific Reports*, vol. 9, no. 1, p. 19412, Dec. 2019, number: 1 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-019-55322-4

[7] H. Belyadi and A. Haghighat, "Chapter 4 - unsupervised machine learning: clustering algorithms," pp. 125–168, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128219294000020

[8] J. Gammon, "Handbook of Methods of Applied Statistics," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 19, no. 1, pp. 75–75, 1969, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2307/2986705. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.2307/2986705

Making Accurate and Consistent Wafer Measurements with Next Generation Guarded True-Kelvin MEMS DC Probes

Choon Beng Sia, Yoichi Funatoko, Isao Kunioka, Masanori Watanabe, Peter Andrews, Ken Dawson, Masahiro Sameshima, Takao Saeki, John Yang, Jesse Li, Xiaoyu Li, Siming Guo, Li Fan, Wei Meng Lim, Eric Wilcox, Anthony Lord, Sarah Lastella, Nobuhiro Kawamata, Jens Klattenhoff, Jarek Kister, Matt Losey and Mike Slessor

FormFactor Inc.
30 Marsiling Industrial Estate Rd 8, #05-02, Singapore 739193
Email: choonbeng.sia@formfactor.com
Telephone: (65) 8168 7090

***Abstract*—Gate length down-scaling of silicon-based transistor results in very small on-state drain-source resistance, making it challenging for test engineers to perform precise and repeatable wafer measurements. Size reduction of aluminum-capped copper test pads to save on lithography, prototyping and production costs implies that it is very difficult to re-probe the same device with low contact resistance. Novel true-Kelvin MEMS analytical DC probes, new test and modelling strategies are proposed in this paper to address these emerging test challenges.**

***Keywords*—Device modeling, DC, CV, flicker noise, probe, MEMS probe, true-Kelvin probe, quasi-Kelvin probe, Kelvin, wafer measurements, contact resistance.**

I. Introduction

The semiconductor industry continues to pursue the relentless downscaling and development of new architectures for silicon-based transistor down to 2 nm and beyond. On-state currents of such advanced transistors are increasing, while off-state currents are kept very low to reduce power consumption. Smaller test pads to reduce lithography and prototyping costs, and the use of copper backend metallization have increased the difficulties for probes to have low and stable contact resistance as there are inadequate fresh pad metal available for deeper probe scrubs or re-probing.

These issues aggravate especially at elevated temperatures when pad aluminium cap layer has been probed and removed, and the exposed underlying copper oxidizes, hindering the ability to establish consistently good electrical contacts for every probe touchdown. The popular quasi-Kelvin analytical DC probes, widely used in the industry, are not able to cope with these challenges. In this paper, a novel coaxial guarded, true-Kelvin MEMS analytical DC probe with replaceable tips, is introduced to achieve precise and consistent device modelling wafer measurements. Despite recent low contact force and low contact resistance improvements on MEMS probe cards [1-4], they are excluded in the scope and discussions of this work because their probe tip layout is usually fixed, unable to support test structures with variable pad pitches.

II. Design Of Engineering Analytical DC Probes with True-Kelvin MEMS Tips

For more than 20 years, the quasi-Kelvin, coaxial guarded, analytical engineering probe shown in Fig. 1, has been a very popular choice for engineers performing device wafer tests. This engineering probing solution usually consists of a ceramic blade with tungsten probe tip, coaxial guarded probe holder, probe arm, cables, Kelvin triaxial connectors and a XYZ positioner to allow accurate probing on the test pads. As channel resistance of advanced node device reduces, parasitic resistance of the quasi-Kelvin probe holder, shown in Fig. 2(a), is no longer negligible. This parasitic resistance also

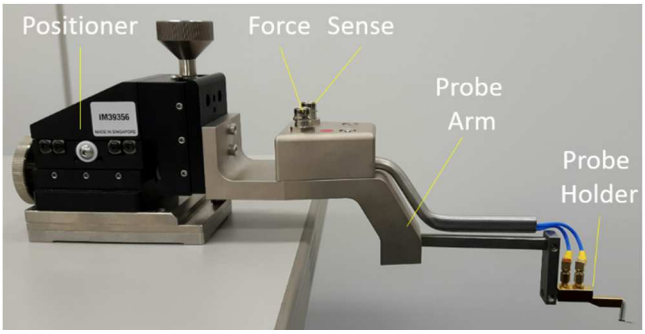
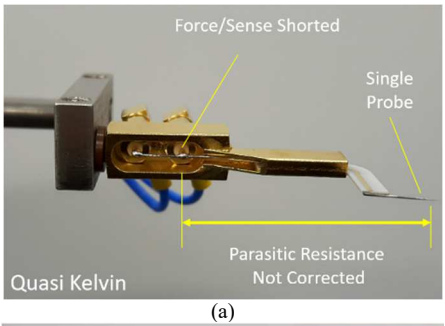
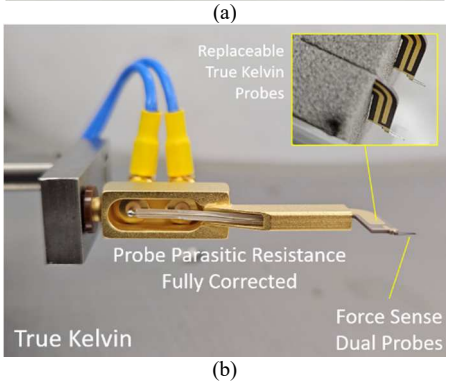


Fig. 1. DC probe supporting guarded and Kelvin wafer measurements.



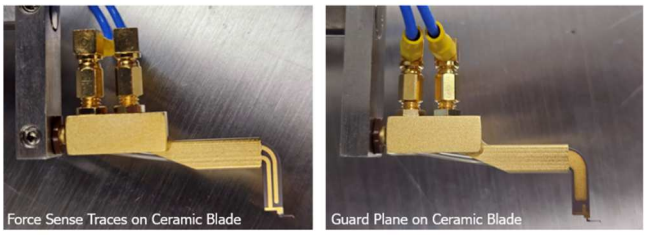
(a)



(b)

Fig. 2. Comparing cantilever quasi-Kelvin QK (a) and MEMS true-Kelvin TK (b) guarded probe holder with replaceable probe tips.

979-8-3503-2987-2/24/$31.00 © 2024 IEEE



(a) (b)

Fig. 3. Force-Sense traces and guard plane on the replaceable ceramic blade of MEMS true-Kelvin probe tips.

increases with temperature, affecting the accuracy and repeatability of device's measured IV curves. To overcome these challenges, a novel true-Kelvin MEMS DC probe as shown in Fig. 2(b) and Fig. 3 is used. The innovative design consists of a new coaxial guarded true-Kelvin probe holder, extending force and sense connections to the removable ceramic probe blade having 2 separate MEMS tips. This facilitates true-Kelvin probe contacts on the device test pads, eliminating all parasitic resistances and inconsistencies in probe contact resistances for every probe touchdown on the test pads.

III. Experimental Setup
Quasi-Kelvin Probes vs True-Kelvin Probes

300mm probe station with automatic wafer loader and Keysight B1500 semiconductor parameter analyser are used to perform device measurements as depicted in Fig. 4. B1500 is utilized as it supports guarded low leakage, force-sense Kelvin measurements. Probe scrubs versus Z over-travel of the quasi-Kelvin cantilever probes and true-Kelvin MEMS probe are presented in Fig. 5. The probe scrub ratio of true-Kelvin MEMS probe is about 8% compared to quasi-Kelvin cantilever probes of 40%. The quasi-Kelvin cantilever probes require a probe scrub of at least 30µm or Z over-travel of about 60 to 70µm for good probe contact resistance.

Fig. 4. 300mm fully automatic probe system with cantilever quasi-Kelvin QK and MEMS true-Kelvin TK (b) probes for wafer-level device characterization.

On the contrary, true-Kelvin MEMS probe requires only 20µm Z over-travel for accurate and repeatable contact. This implies that it is able to support pad sizes smaller than 30×30µm as the probe scrub is only about 7 to 8µm. Fig. 6 shows the total resistance of 2 quasi-Kelvin cantilever probes on an aluminium-capped copper test pad at 25°C and 150°C for 100 contact cycles. This work was previously published in 2017 ICMTS [5]. The total parasitic resistance is observed to be about 1 and 5 Ω for 25°C and 150°C respectively on the first probe contact – such high test setup resistance limits the

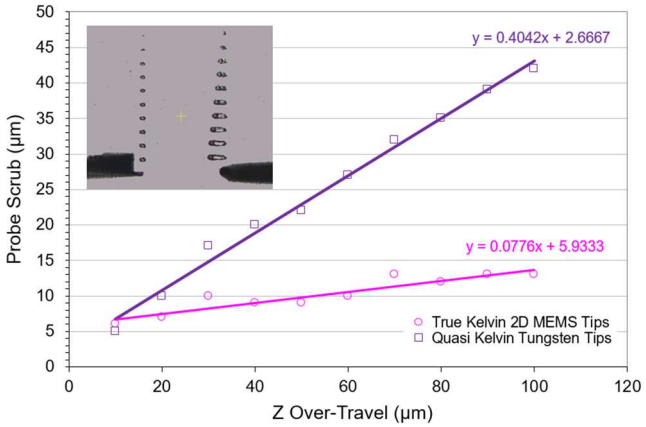


Fig. 5. Probe scrub versus Z Over-Travel for cantilever quasi-Kelvin (QK) and MEMS true-Kelvin (TK) probes.

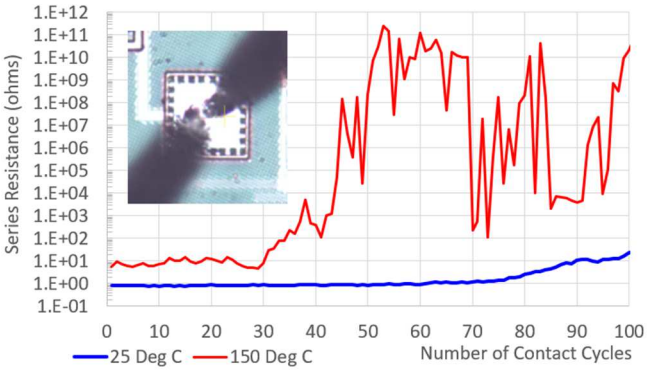


Fig. 6. Series resistance of 2 cantilever quasi-Kelvin (QK) probes on aluminium-capped copper test pad over 100 contact cycles.

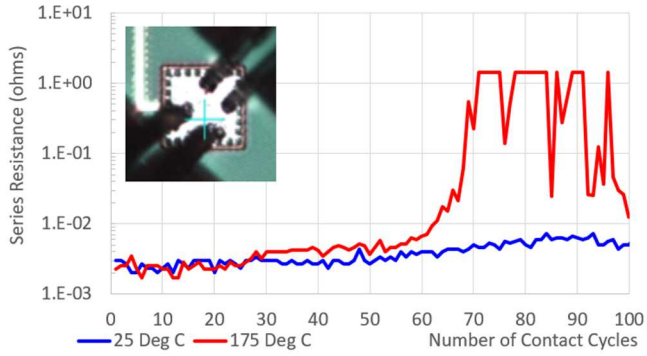


Fig. 7. Series resistance of 2 MEMS true-Kelvin (TK) probes on aluminium-capped copper test pad over 100 contact cycles.

ability for test engineers to accurately characterize devices at high temperatures. Particularly, at 150°C, from 30th probe contact cycle onwards, the total parasitic resistance increases drastically and eventually became open circuit. This is due to repeated deep probing on test pad causing the aluminium cap layer to be removed, exposing the underlying copper metal layers which oxidize, eventually resulting in an open circuit.

Such experimental result reveals the difficulties and challenges to have consistency and accuracy when measuring device characteristics with these aluminium-capped copper test pads. When the same experiment is repeated with the proposed true-Kelvin MEMS probes, the total parasitic resistance shown in Fig. 7 remains low, less than 10 mΩ at 25°C. The total parasitic resistances at 25°C and 175°C are comparable, less than 5 mΩ for the first 50 contact cycles, with no large gaps observed. Since true-Kelvin MEMS probe

requires Z over-travel of only 20µm, less pad damage has been inflicted and therefore little underlying copper oxidation and no open circuit has been observed. Fig. 8 shows the parasitic resistance of 2 true-Kelvin probes on gold pads. It is observed that at 25°C and 175°C, the average total parasitic resistances are consistently low at 1.4 and 2.2 mΩ respectively over 100 probe contact cycles since there is no metal oxidation observed during these tests. Fig. 9 shows the leakage current for the guarded true-Kelvin probe when 10V is applied, with probe tips in the air at 250µm above the wafer. The probe leakage current is less than ±5 and ±10 fA at temperatures of 25 and 175°C respectively. Such excellent probe leakage performance is required to characterize a single transistor with typical off-state current in the pico-ampere range.

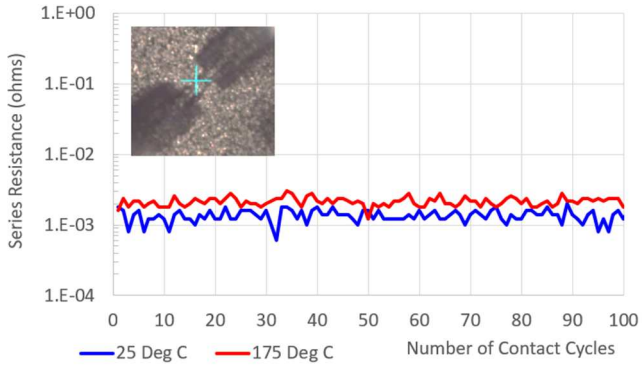


Fig. 8. Series resistance of 2 MEMS true-Kelvin (TK) probes on gold pad over 100 contact cycles.

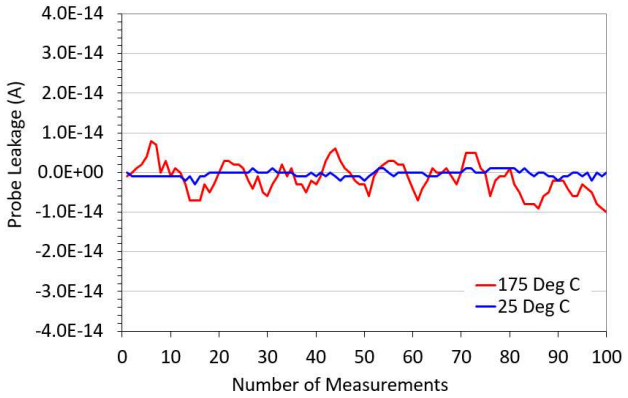


Fig. 9. Probe leakage current for MEMS true-Kelvin (TK) probe when 10 V is applied, probe is 250 µm above the wafer surface at 25 and 175°C.

IV. Test Structures and Device Test Results

NMOSFETs with size of width/length=100/0.06µm are used for device tests to compare performance of cantilever quasi-Kelvin and MEMS true-Kelvin probes. The NMOSFET test structures consist of both standard 4-pad and true-Kelvin 6-pad layout and the test pads are 50×50µm in size.

A. Current-Voltage (IV) Measurements

Characterizing the NMOSFETs with quasi-Kelvin cantilever probes, inconsistent drain current, Ids and drain-source resistance, Rds have been obtained when Vds is swept from 0 to 1.2V while the gate is biased at 1.2V over 100 probe contact cycles, as shown in Fig. 10 (a). This is due to the parasitic resistance of quasi-Kelvin probe body and the unstable probe-pad contact resistance. On the contrary, true-Kelvin MEMS probes provided extremely good repeatability

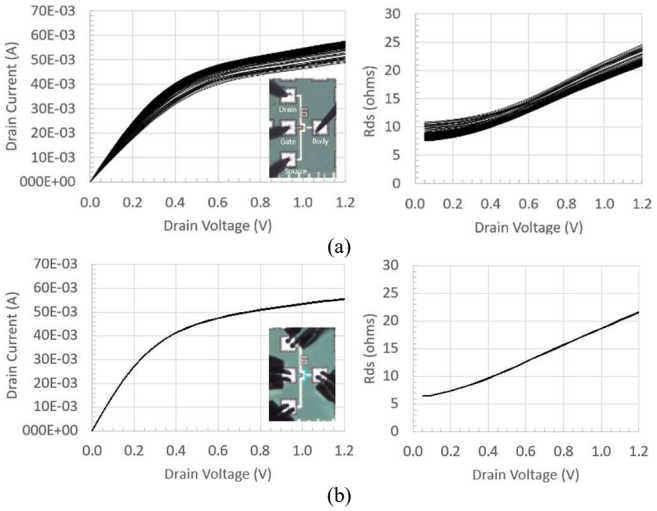


Fig. 10. Id and Rds for 100/0.06µm NMOSFET measured repeatedly over 100 contact cycles with cantilever quasi-Kelvin probes (a) and MEMS true-Kelvin probes (b) at 25°C in a 4-pad test layout.

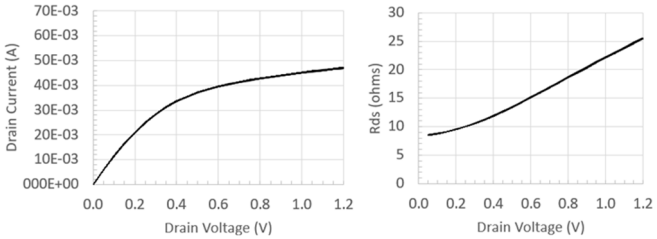


Fig. 11. Id and Rds for 100/0.06µm NMOSFET measured over 100 contact cycles with MEMS true-Kelvin probes at 175°C.

in the same test. Fig. 10 (b) and Fig. 11 show identical Ids and Rds versus Vds plots over 100 probe contact cycles at 25°C and 175°C correspondingly, demonstrating accuracy, repeatability and real-time parasitic resistance corrections by B1500 for every measurement. The true-Kelvin MEMS probe requires only 20µm Z over-travel, demonstrating ability to handle small test pads. It also requires minimal probe tip maintenance – only cleaning that is necessary to prevent force and sense tips from shorting together.

Quasi-Kelvin probe with one single MEMS probe tip instead of cantilever tungsten tip is also fabricated in this work. Together with true-Kelvin MEMS probes, 6-pad true-Kelvin NMOSFET test structures are characterized to establish the correlations between measurement accuracy, true-Kelvin MEMS probes versus quasi-Kelvin MEMS probes and 4 versus 6-pad layout configurations. Fig. 12 (a) shows Test A with 4 quasi-Kelvin MEMS probes on 4 test pads of the NMOSFET device. Fig. 12 (b) depicts Test B with NMOSFET characterize by 2 quasi-Kelvin MEMS and 2 true-Kelvin MEMS on drain and source terminals. Fig. 12 (c) shows 6 quasi-Kelvin MEMS probes on a 6-pad true Kelvin device test structure as Test C. Ids versus Vds plots for the 3 test configurations are compared in Fig. 12 (d).

Test C with a 6-pad true Kelvin test structure yielded the highest current and accuracy because parasitic resistances associated with the drain source test leads are fully corrected during the measurements. Test A resulted in the lowest Ids. Compared to Test C, Test B is recommended instead because true-Kelvin probes are able to correct for probe contact resistances in a 4-pad layout with 33% reduction in test structure size. Test B setup in a 4-pad layout is also preferred for capacitance-voltage as well as 1/f noise measurements

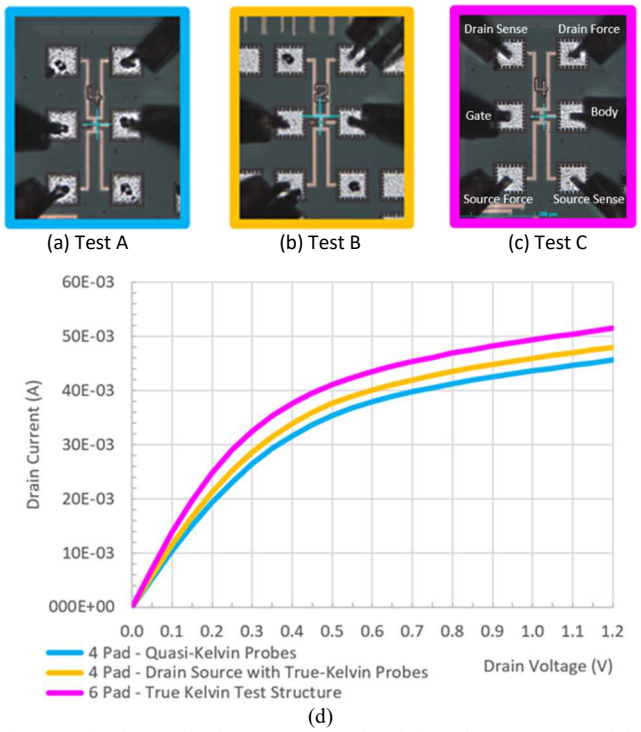


Fig. 12. Die photos showing MEMS quasi-Kelvin and MEMS true-Kelvin probes on 4-pad and 6-pad test configurations - 4 quasi-Kelvin probes Test A (a), 2 quasi-Kelvin 2 true-Kelvin probes Test B (b), 6 quasi-Kelvin probes Test C (c) and their respective Id versus Vd plots (d).

because it does not have the additional source and drain sense test pads. The device test leads should be designed with short, wide top metallization to minimize parasitic resistances. At various temperatures, characterizing additional de-embedding test structures with the source and drain test leads shorted together, allows source and drain parasitic test-lead resistor models to be developed. Having these resistor models in SPICE simulators while extracting the MOSFET model parameters will result in highly accurate MOSFET SPICE models.

B. Capacitance-Voltage (CV) Measurements

Apart from IV measurements, to qualify as a device characterization and modelling probe, the proposed true-Kelvin MEMS probe must be able to support CV and flicker noise wafer measurements. In CV tests with a LCR meter, as each true-Kelvin MEMS probe has 2 force-sense probe tips, it is possible to connect one MEMS tip or have both MEMS tips shorted to each of the "high" and "low" test terminals of the LCR meter. To validate the CV measurement performance of the true-Kelvin probes, Keysight 4284 LCR meter is used to measure the gate capacitance of the 100/0.06µm NMOSFET in a 4-pad test layout. The source and drain of the transistors are shorted together and the gate voltage is swept from -2 to 2V.

Shorting both force and sense probe tips together and connecting them to each of the instrument test terminals, Fig. 13 shows excellent correlations, with almost identical values of NMOSFET gate capacitance measured with MEMS true-Kelvin and cantilever quasi-Kelvin probes. On other hand, it is interesting to note that when only 1 tip is connected to each of the LCR meter test terminals, after an OPEN calibration, there is a systematic error with larger capacitance of about 0.01pF. This is likely due to additional capacitance introduced by the floating tip that was not corrected during the OPEN in air capacitance calibration but was measured by the LCR meter in the device measurements when this floating tip comes into electrical contact with the test pad.

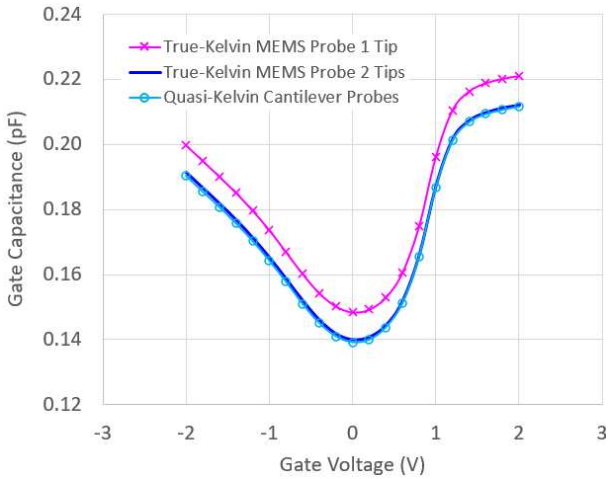


Fig. 13. Gate capacitance versus gate voltage for the 100/0.06µm NMOSFET, measured with MEMS true-Kelvin probe with 1 tip or 2 tips shorted to each of the test terminals of the LCR meter versus quasi-Kelvin cantilever probe.

C. Flicker Noise (1/f Noise) Measurements

The Primarius 9812DX(HV) flicker noise measurement system is used to qualify the true-Kelvin MEMS probes for device characterization and modelling applications. Similar to the CV test setup, there are 2 ways to connect the true-Kelvin probes. Fig. 14(a) shows the sense tip floating and only the force tip is connected to the Amplifier Filter Unit (AFU) of the flicker noise test setup using 1 single test cable each for the drain and source device terminals. Another possible setup is to short the force and sense probe tips with 2 separate test

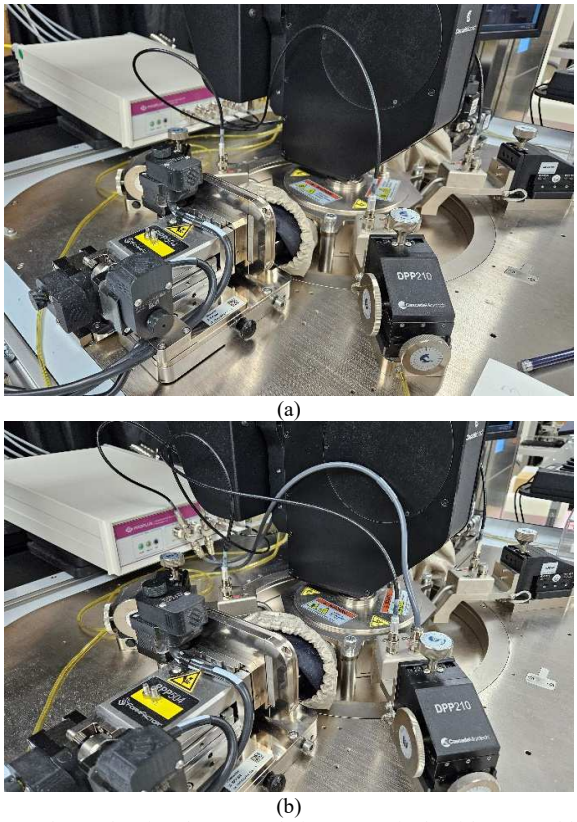


Fig. 14. Photos showing the source drain test terminals with 1 test cable (a) and 2 test cables (b) connected to the Amplifier Filter Unit (AFU) of the Primarius 9812DX(HV) flicker noise test system.

cables to the measurement channel on the AFU as shown in Fig. 14(b). This method reduces the cabling resistance but increases the parasitic capacitances of the test setup which will likely cause an early roll-off in the measured noise spectrum if a very large load resistor is selected for more accurate noise measurements. In this study, to facilitate a fair comparison to quasi-Kelvin cantilever probe, only the force tip is connected to the AFU using 1 test cable and sense tip is left floating.

Fig. 15 shows the noise spectrum, Sid versus frequency plots for a W/L=20/0.06µm NMOSFET device in a 4-pad test layout. A smaller device with lower bias currents is selected so that the device noise spectrum will be low, allowing the true-Kelvin MEMS probes to be carefully evaluated. It is observed that the noise spectrums measured by the quasi-Kelvin cantilever probes and true-Kelvin MEMS probes are well correlated. At low current levels of 6.5µA and 50µA, it is noted that the true-Kelvin MEMS probes have smaller fluctuations in the measured Sid compared to the quasi-Kelvin probes. This could be attributed to an ultra-stable MEMS probe-to-pad contact and well-matched characteristics impedance for the true-Kelvin MEMS probe. Both features are critical in allowing fluctuations of the drain current to be accurately captured by the instruments, over time, during low frequency noise characterization of these silicon-based transistors.

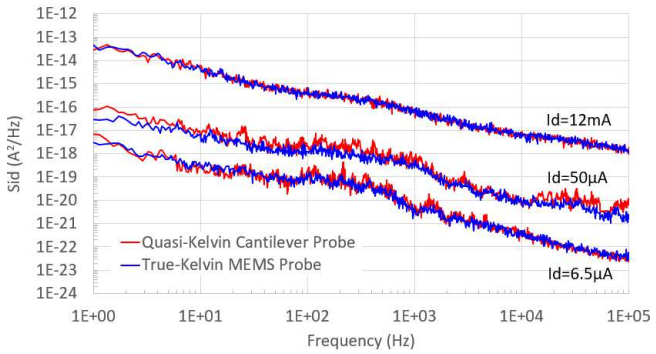


Fig. 15. Sid noise spectrum for the W/L=20/0.06µm NMOSFET biased at Vg=Vd=0.3V (Id=6.5µA), Vg=0.3V Vd=1.2V (Id=50µA) and Vg=Vd=1.2V (Id=12mA), measured with cantilever quasi-Kelvin probes and MEMS true-Kelvin probes.

V. Conclusion

A novel low-leakage DC MEMS probe with ultra small probe scrubs and true-Kelvin force-sense probe tips has been successfully demonstrated to address the test challenges of making precise and consistent device modelling wafer measurements for advanced node silicon-based transistors.

Acknowledgment

The authors would like to thank Taiwan Semiconductor Research Institute (TSRI) for rendering help in making capacitance and flicker noise measurements of the NMOSFETs with the new true-Kelvin MEMS DC probes.

References

[1] K. Kataoka, S. Kawamura, T. Itoh and T. Suga, "low contact force and compliant MEMS probe card utilizing fritting contact", MEMS 2002, pp.364-367.

[2] Bong-Hwan Kim and Jong-Bok Kim, “Design and fabrication of a high manufacturable MEMS probe card for high speed testing”, Journal of Micromechanics and Microengineering, 2008.

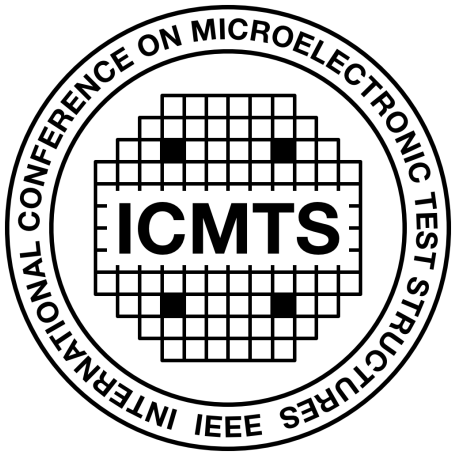
[3] T.Lai and C.Tsou, "A Claw type of MEMS probe card for the electrical testing of micro-solder ball", MEMS 2012, pp.345-348.

[4] K. Hosaka, S. Morishita, I. Mori, M. Kubota and Y. Mita, "An integrated CMOS-MEMS

[5] probe having two-tips per cantilever for individual contact sensing and Kelvin measurement with two cantilevers," 2013 IEEE International Conference on Microelectronic Test Structures (ICMTS), Osaka, Japan, 2013, pp. 3-6.

[6] C. B. Sia, "True Kelvin CMOS Test Structure to achieve accurate and repeatable DC wafer-level measurements for device modelling applications," 2017 International Conference of Microelectronic Test Structures (ICMTS), Grenoble, France, 2017, pp. 1-4.

SESSION 7
S-Parameters and De-Embedding



17th April 2024, 16:00–17:00

Session Co-Chairs: Tatsuya Ohguro *Toshiba Corp., Japan*
Carlo Cagli, *CEA-Leti, France*

Understanding the Substrate Effect on De-embedding Structures Fabricated on SOI Wafers Using Electromagnetic Simulation

Bruno Neckel Wesling[1,2], Marina Deng[2], Chhandak Mukherjee[2], Thomas Mikolajick[1,3], Jens Trommer[1], and Cristell Maneux[2]
[1]NaMLab gGmbH, Nöthnitzer Str. 64 a, 01187 Dresden, Germany
[2]IMS, 351 Cours de la Libération, Talence Cedex, 33045, France
[3]Chair for Nanoelectronics, TU Dresden, 01187 Dresden, Germany

Email: bruno.neckel-wesling@u-bordeaux.fr

***Abstract*—In this paper, we present the fabrication, characterization, and electromagnetic simulation of open pad test structures on silicon-on-insulator substrates, with an emphasis on the impact of the substrate properties on RF performance. Targeting the design of optimal RF test structures for emerging technologies, we demonstrated that a high-resistivity substrate is essential to minimize losses and parasitic capacitances in RF measurements for technologies using silicon-on-insulator wafers.**

I. Introduction

In the era of advanced neural networks permeating daily life [1], compute-in-memory is emerging as a solution to the memory access bottlenecks inherent in traditional architectures [2]. This paradigm shift towards novel architectures has prompted a reevaluation of the fundamental blocks that constitute these systems [3]. In this context, emerging devices such as the reconfigurable field effect transistors (RFETs) have shown promising potential as a future contender for the basic circuit building block development [4].

Academic research environments often serve as testing grounds for emerging device concepts, using technologies such as silicon-on-insulator (SOI) wafers to showcase innovations like RFETs and junctionless transistors [5], [6]. However, unlike industrial processes, academic level fabrication lacks the complete back end-of-line process. As a result, metal contact pads for electrical test structures are often placed directly on top of the buried oxide (BOX). For DC measurements, the BOX functions as a good insulator between adjacent devices and enables the use of the substrate as a back gate [7]. However, for analog design or even RF measurements, the substrate can lead to significant losses.

In this study, we will examine how these constraints may impact the RF performance of open test-structures utilised for de-embedding. Getting access to both extrinsic and intrinsic elements of the device, including resistances, inductances and capacitances of the small-signal equivalent circuit associated with these emerging technologies, can be quite challenging if the devices are fabricated through a non-optimised process.

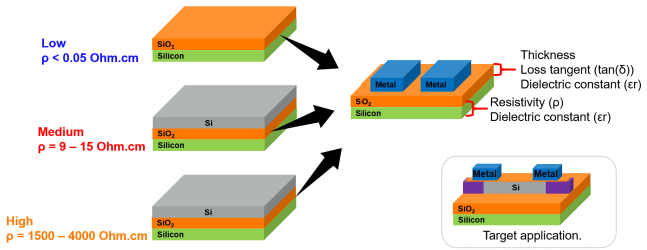


Fig. 1. Wafers having different levels of substrate resistivity used for the fabrication of the open pad structures (left) and a schematic drawing of the test structure (top right) including the necessary material properties to be defined for the EM simulations on ADS. On the bottom right a schematic of an RFET as a target application is shown [5].

To explore the frequency response of these emerging devices fabricated on SOI wafers, it is therefore necessary to first determine the structures that one requires to fabricate, in order to enable electrical contacts, such as pads and metal lines. In this context, electro-magnetic (EM) simulations can be explored as a tool to understand the effects of different wafer substrates and thus to provide predictive results on the operation of the high frequency dummies required to de-embed active devices [8], [9], [10].

II. Methodology

In this work, open pad structures were fabricated on different SOI substrate resistivity and the S-parameters were measured up to 24 GHz using |Z|−Probe® RF probes with 100 µm pitch. The measurement data was then used as calibration inputs for the EM simulation (see Figure 1). Finally, an equivalent circuit is proposed for the wafers that are most suitable and optimised for the fabrication of active devices.

A. Test Structure Fabrication

The fabrication steps for the RF pad structures are presented in Figure 2. The medium resistivity wafer corresponds to a SOI wafer with a 100 nm BOX and a substrate resistivity

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

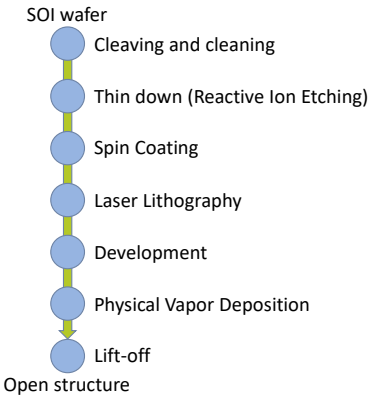


Fig. 2. Fabrication steps of an open pad structure. The thin-down step is skipped for the low resistivity wafer.

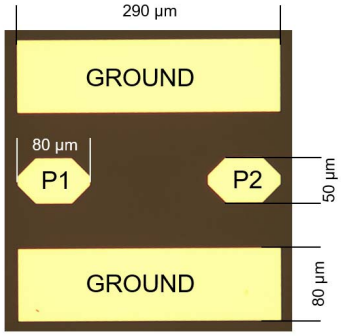


Fig. 3. Optical image of the fabricated open pad structure. The symmetry implies that $S_{11} = S_{22}$ and $S_{12} = S_{21}$.

between 9-18 Ohm·cm. The low resistivity wafer presents a 100 nm SiO_2 on top of the Si with a resistance lower than 0.05 Ohm·cm. For the high resistivity wafer, a 1000 nm BOX with a substrate resistance between 1500-4000 Ohm·cm was chosen. All the wafers used in this study are commercially available. To grow the structures directly on top of the BOX for the standard and high resistivity wafers, the silicon on top is first removed using reactive ion etching. The design of the open structure is transferred using laser direct write lithography using a positive resist.

After the exposure, the resist is developed, and a metal layer is deposited using ion beam sputtering. The metal layer consists of 90 nm of Ni and 10 nm of Pt. A lift-off process is the last step followed where the undesired metal areas are removed together with the underlying resist, leaving the designed structures on top of the SiO_2. The final structure with its associated dimensions can be seen in Figure 3.

B. Test Structure Electromagnetic Simulation

The EM simulations were performed using the Momentum simulator from PathWave Advanced Design System (ADS). The layout of the open pad was first imported into ADS using the same GDSII file that was used for the fabrication steps.

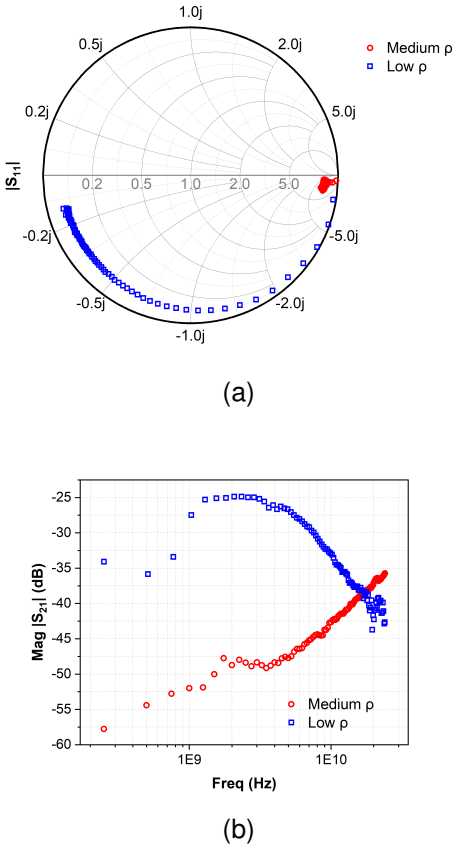


Fig. 4. Measured S-parameters for the medium (red circle symbol) and the low (blue square symbol) resistivity wafers with 100 nm SiO_2 layer. (a) Reflection coefficients S_{11}. (b) Magnitude of the transmission parameter S_{21}.

The electrical properties of the substrates were then specified based on the data-sheet of the selected wafer.

III. Results

A. Electrical Characterization versus EM Simulation of the Open Pad Test Structure

The measured reflection coefficient of the open pad on the medium-resistivity SOI wafer, shown in Figure 4a, indicates a coupling between the ports resulting in a deviation from the purely capacitive behaviour, due to the poor isolation provided by the substrate. The low resistivity wafer yields a high capacitance due to the thin BOX. Figure 4b displays a significant reduction in cross-talk from the low to the medium resistivity substrate, even though further improvement can still be achieved [11].

Figures 5 depict results obtained from the EM simulations with similar characteristics as the measured ones. Based on the EM simulation outcomes, the high resistivity wafer was deemed a reasonable solution for minimising possible cross talk at frequencies below 10 GHz and achieving small pad capacitances.

The results obtained from the high resistivity wafer are presented in Figure 6. At 1 GHz, the high resistivity wafer recorded a 12-dB reduction in cross-talk compared to the

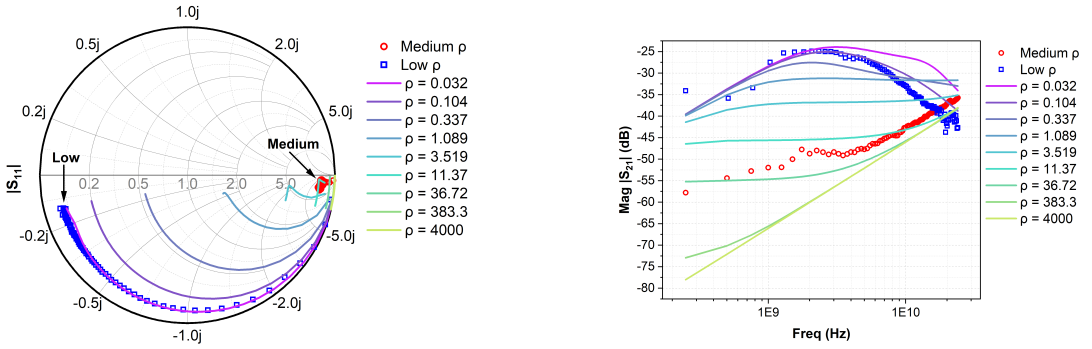


Fig. 5. S-parameters from measurements (symbols) and simulations (lines) for different resistivities ρ (Ohm·cm) with a 100 nm BOX. (a) Reflection coefficients S_{11}. (b) Magnitude of the transmission parameter S_{21}. EM simulation with a ρ of around 0.032-0.104 Ohm·cm captures the behaviour of the low resistivity wafer quite well, similar results were obtained for a ρ = 11.37-36.72 for the medium resistivity wafer.

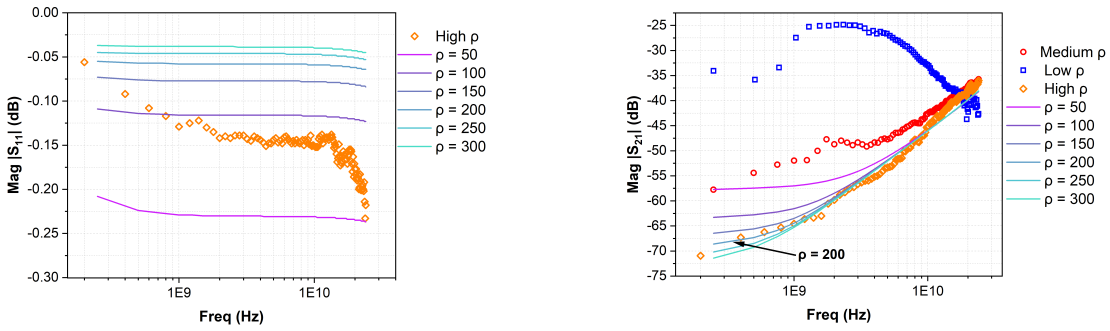


Fig. 6. S-parameters from measurements (symbols) and simulations (lines) for different resistivities ρ (Ohm·cm): (a) Magnitude of the reflection coefficients S_{11} and (b) Magnitude of the transmission coefficients S_{21}, for the high resistivity SOI wafer with a 1000 nm BOX. EM simulation results for different resistivity values obtained for a 1000 nm BOX are also compared with the results from the measured low and medium resistivity wafers with 100 nm BOX. EM simulation for the resistivity of approximately 200 Ohm·cm captures the behaviour of the high-resistivity wafer correctly.

medium-resistivity one. The measurements on the high resistivity substrate demonstrate that the expected values obtained from the EM simulation are distinctly different from that of the experimental observations, due to the presence of parasitic surface conduction at the interface of the BOX and the substrate. It has been previously demonstrated that the effective substrate resistivity at the interface is in fact lower than the one that is obtained deep within the substrate [11]. It can be observed that the measurements from the high resistivity substrate are comparable with the electromagnetic (EM) simulation results when the resistance of the substrate is approximately 200 Ohm·cm (see Figure 6b). This value is one order of magnitude lower than the intrinsic value of 1500-4000 Ohm·cm specified in the wafer data-sheet. Hence, as an additional outcome, the EM simulation can also be utilised to determine the equivalent resistivity of the substrate.

B. Equivalent Circuit of the Open Pad

An equivalent circuit and its extraction from measurements is proposed for the open structures on the medium and high resistivity wafers and is presented in Figure 7. The circuit is based on the standard π network of an open circuit consisting of three impedances Z_i, Z_t and Z_o. For the entire analyses $Z_i = Z_0$ is considered, due to the symmetry of the structure.

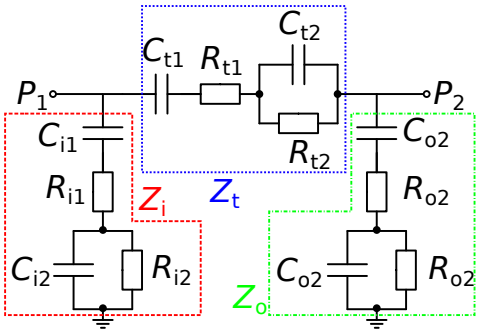


Fig. 7. Proposed equivalent circuit for the open pad structure.

Non idealities of the substrate, such as the effects of the

TABLE I
VALUES OF THE PARAMETERS UTILISED FOR THE SIMULATION OF THE EQUIVALENT CIRCUIT.

Parameters	High	Medium	Parameters	High	Medium
C_{i1}	56.5 fF	210 fF	C_{t1}	2.1 fF	7.7 fF
R_{i1}	5 Ω	5 Ω	R_{t1}	600 Ω	3k Ω
C_{i2}	6.1 fF	5.3 fF	C_{t2}	1 fF	1.8 fF
R_{i2}	4895 Ω	1 kΩ	R_{t2}	13.2 kΩ	32 kΩ

BOX, are included in the equivalent circuit, represented by a capacitor with a series resistance (C_{i1} and R_{i1}) and the substrate is represented by a resistance in parallel to a capacitor (R_{i2} and C_{i2}). The capacitance C_{i1} is first roughly estimated at low frequencies from the converted measurements using (1). From the same equation, C_{i2} is obtained, however at medium frequencies, where its value becomes frequency independent. The resistance R_{i2} is estimated from the highest point of (2) over frequency while R_{i1} is expected to have a small value that can just be determined directly at very high frequencies. C_{t1} and C_{t2} are obtained from (3) using the same methodology as C_{i1} and C_{i2} while the resistances R_{t2} and R_{t1} are determined from (4) when its value is maximum and from its value at high frequencies, respectively. With these estimated values the equivalent circuit is then simulated using ADS.

$$C_i = imag(Y_{11} + Y_{21})/\omega \tag{1}$$

$$R_i = real((Y_{11} + Y_{21})^{-1}) \tag{2}$$

$$C_t = imag(-Y_{21})/\omega \tag{3}$$

$$R_t = real((-Y_{21})^{-1}) \tag{4}$$

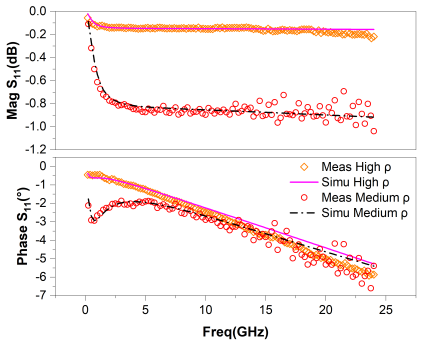
The Table I presents the extracted values of the capacitances and resistances, optimised from the equivalent circuit simulation through fitting.

The S parameters from the simulation of the final equivalent circuit and the measurements are compared in Figure 8, showcasing a good agreement between the simulation and measurement, and equation 5 was used to quantify the model accuracy. An error of 1.5 % was obtained for the medium ρ wafer while a 3.86 % error was observed for the high ρ wafer.

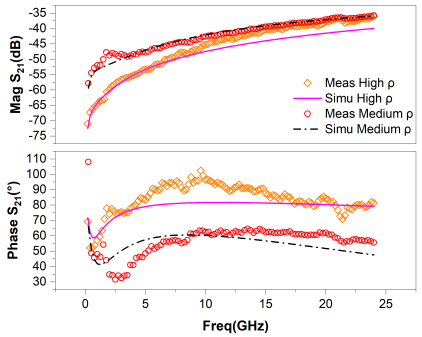
$$\varepsilon_{\text{tot}(S)} = 100 * \frac{1}{4} \sum_{ij} \left\{ \sum_{\text{freq}} \frac{|\text{meas}\,(S_{ij}) - \text{sim}\,(S_{ij})|^2}{|\text{meas}\,(S_{ij})|^2} \right\} \frac{1}{N_{\text{freq}}} \tag{5}$$

IV. CONCLUSION

In conclusion, we demonstrated the strength of EM simulations as a tool to understand the impact of the resistivity of SOI wafers followed by a methodology for the equivalent circuit extraction. Our results indicate that the layout of other dummy structures can also be optimized through predictive EM simulations before new fabrication runs are performed.



(a)



(b)

Fig. 8. S-parameters from measurements (symbols) and equivalent circuit simulations (solid and dash-dot lines) for different resistivity values ρ (Ohm·cm). (a) Magnitude and phase of the reflection coefficients S_{11}. (b) Magnitude and phase of the transmission coefficient S_{21}.

Moreover, a high resistivity wafer can offer improvements of the RF characteristics, necessary for the modelling of emerging devices fabricated on non-optimized technology platforms with constraints imposed by a single metallization layer.

ACKNOWLEDGMENT

This work was supported by the project FVLLMONTI funded by the European Union's Horizon 2020 research and innovation program under grant agreement N○101016776.

REFERENCES

[1] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, "Training language models to follow instructions with human feedback," *Advances in Neural Information Processing Systems*, vol. 35, pp. 27 730–27 744, 2022.

[2] S. Rai, M. Liu, A. Gebregiorgis, D. Bhattacharjee, K. Chakrabarty, S. Hamdioui, A. Chattopadhyay, J. Trommer, and A. Kumar, "Perspectives on emerging computation-in-memory paradigms," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2021, pp. 1925–1934.

[3] C. Maneux, C. Mukherjee, M. Deng, M. Dubourg, L. Reveil, G. Bordea, A. Lecestre, G. Larrieu, J. Trommer, E. Breyer, S. Slesazeck, T. Mikolajick, O. Baumgartner, M. Karner, D. Pirker, Z. Stanojevic, D. Atienza, A. Levisse, G. Ansaloni, A. Poittevin, A. Bosio, D. Deleruyelle, C. Marchand, and I. O'Connor, "Modelling of vertical and ferroelectric junctionless technology for efficient 3D neural network compute cube dedicated to embedded artificial intelligence," in *2021 IEEE International Electron Devices Meeting (IEDM)*. San

Francisco, CA, USA: IEEE, Dec. 2021, pp. 15.6.1–15.6.4. [Online]. Available: https://ieeexplore.ieee.org/document/9720572/

[4] T. Mikolajick, G. Galderisi, S. Rai, M. Simon, R. Böckle, M. Sistani, C. Cakirlar, N. Bhattacharjee, T. Mauersberger, A. Heinzig *et al.*, "Reconfigurable field effect transistors: A technology enablers perspective," *Solid-State Electronics*, vol. 194, p. 108381, 2022.

[5] A. Heinzig, S. Slesazeck, F. Kreupl, T. Mikolajick, and W. M. Weber, "Reconfigurable Silicon Nanowire Transistors," *Nano Letters*, vol. 12, no. 1, pp. 119–124, Jan. 2012. [Online]. Available: https://pubs.acs.org/doi/10.1021/nl203094h

[6] R. T. Doria, M. A. Pavanello, R. D. Trevisoli, M. De Souza, C.-W. Lee, I. Ferain, N. D. Akhavan, R. Yan, P. Razavi, R. Yu, A. Kranti, and J.-P. Colinge, "Junctionless Multiple-Gate Transistors for Analog Applications," *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2511–2519, Aug. 2011. [Online]. Available: http://ieeexplore.ieee.org/document/5934399/

[7] M. Simon, H. Mulaosmanovic, V. Sessi, M. Drescher, N. Bhattacharjee, S. Slesazeck, M. Wiatr, T. Mikolajick, and J. Trommer, "Three-to-one analog signal modulation with a single back-bias-controlled reconfigurable transistor," *Nature Communications*, vol. 13, no. 1, p. 7042, Nov. 2022, number: 1 Publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41467-022-34533-w

[8] B. Neckel Wesling, M. Deng, C. Mukherjee, M. De Matos, A. Kumar, G. Larrieu, J. Trommer, T. Mikolajick, and C. Maneux, "Extraction of small-signal equivalent circuit for de-embedding of 3D vertical nanowire transistor," *Solid-State Electronics*, vol. 194, p. 108359, Aug. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0038110122001319

[9] C. Yadav, M. Deng, S. Fregonese, M. Cabbia, M. De Matos, B. Plano, and T. Zimmer, "Importance and Requirement of Frequency Band Specific RF Probes EM Models in Sub-THz and THz Measurements up to 500 GHz," *IEEE Transactions on Terahertz Science and Technology*, vol. 10, no. 5, pp. 558–563, Sep. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9123692/

[10] L. Galatro and M. Spirito, "Millimeter-Wave On-Wafer TRL Calibration Employing 3-D EM Simulation-Based Characteristic Impedance Extraction," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 4, pp. 1315–1323, Apr. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7837598/

[11] D. Lederer and J.-P. Raskin, "New substrate passivation method dedicated to HR SOI wafer fabrication with increased substrate resistivity," *IEEE Electron Device Letters*, vol. 26, no. 11, pp. 805–807, Nov. 2005. [Online]. Available: http://ieeexplore.ieee.org/document/1522461/

Modified Bisection Thru-Only Deembedding Algorithm for Long Test Fixtures

Alexander Quint, Luca Valenziano, Joachim Hebeler, Thomas Zwick, Akanksha Bhutani
Institute of Radio Frequency Engineering and Electronics, Karlsruhe Institute of Technology, Germany
Engesserstr. 5, 76131 Karlsruhe, Germany
alexander.quint@kit.edu

***Abstract*—Thru-only deembedding is a simple deembedding method using a single deembedding structure. Common thru-only deembedding methods often have overshoots in the deembedded S-parameters when the deembedding structure is $\frac{\lambda}{2}$ long. This is caused in the bisection step. A modified approach is presented, which extracts the phase of the S-parameters of the deembedding structure before the bisection step, performs the bisection on the magnitude of the S-parameters and applies the phase afterwards again, therefore eliminating the overshoots. The accuracy of the proposed algorithm is verified on measured test structures.**

I. Introduction

S-parameter deembedding is a well known technique to remove the influence of test fixtures from the measured S-parameters of an arbitrary device under test (DUT). The basic principle of deembedding is shown in Fig. 1, where S_{DUT} are the true S-parameters of the DUT, $S_{\mathrm{TF,L}}$ and $S_{\mathrm{TF,R}}$ are the S-parameters of the left and right test fixture respectively and S_{tot} are the S-parameters of the embedded DUT.

The S-parameters of the fixtures are obtained by measuring deembedding structures consisting of the test fixtures. A simple example for a deembedding structure is shown in Fig. 2, with S_{Deem} being the S-parameters of the deembedding structure.

One simple variant of thru-only deembedding is the so-called bisection thru-only deembedding, where $S_{\mathrm{TF,L}}$ and $S_{\mathrm{TF,R}}$ are obtained by bisecting S_{Deem}. The bisection is usually carried out by taking the matrix square root of the $ABCD$-parameters of the deembedding structure. This deembedding method is then called matrix root based bisection thru-only deembedding. Fig. 3 shows the deembedded and measured S-parameters of a $3\,\mathrm{mm}$ GCPW $50\,\Omega$ line. The used test structures are shown in Fig. 4 and denoted in Table I. All test strutures are grounded coplanar waveguide (GCPW) lines on a $0.2\,\mathrm{mm}$ thick Rogers RO4003C substrate, manufactured on a picosecond laser.

It can be seen that using this bisection method, the dembedded S-parameters can have overshoots. These overshoots happen at frequencies where the length of the deembedding structure are odd multiples of $\frac{\lambda}{2}$ [1], [2], which will be investigated in the following.

II. Deembedding Approaches

In addition to the shortly introduced matrix root based bisection thru-only deembedding, there are many known deembedding methods. They differ in the amount and kind of deembedding structures. In some cases it is beneficial to only have one deembedding structure, e.g. when chip or printed circuit board (PCB) space is tight or expensive. Therefore, this paper only focuses on this class of deembedding methods, called thru-only deembedding.

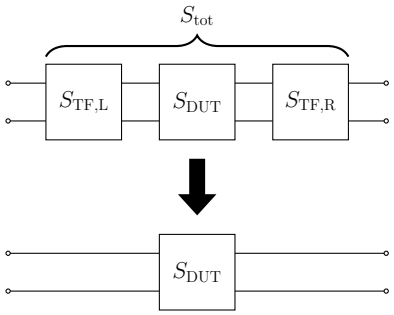


Fig. 1. Basic principle of S-parameter deembedding.

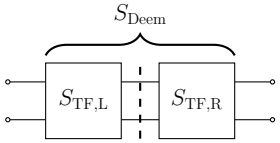


Fig. 2. Deembedding structure for thru-only deembedding.

Many methods model the test fixtures using lumped-element models [3], [4]. This assumption is however not valid, if the deembedding structure is not electrically short. For electrically long test fixtures, more general approaches are required.

A different approach to thru-only deembedding is presented in [5], in which the deembedding structure is not bisected but subtracted at the end of the embedded DUT. As no bisection is performed, there are no overshoots in the deembedded S-parameters. However, this approach requires the DUT and the transition from test fixture to DUT to be well matched. Therefore it is inaccurate if the DUT has internal reflections.

The thru-only deembedding approaches published so far have limitations, either in terms of the electrical length of the test fixture or in terms of the choice of DUT. The matrix root based bisection thru-only deembedding is the most general one. It requires the least assumptions about the test fixtures and the DUT. Therefore it is thoroughly analysed in this work

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

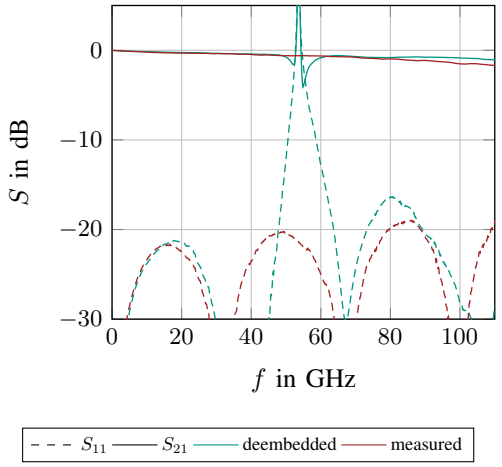


Fig. 3. Measured and deembedded S-parameters of a 3 mm GCPW 50 Ω line (test structure 3 in Fig. 4) using matrix root bisection thru-only deembedding. The deembedded S-parameter are obtained by performing bisection thru-only deembedding on a 5 mm GCPW 50 Ω line (test structure 5). The deembedding structure is a 2 mm GCPW 50 Ω line (test structure 2).

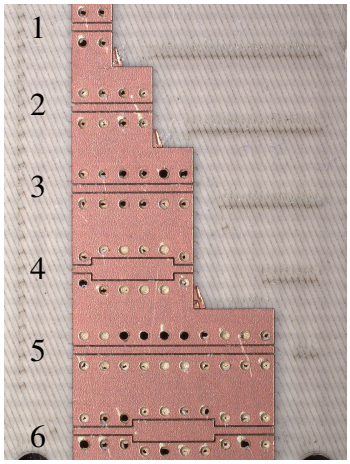


Fig. 4. Used test structures on 0.2 mm thick Rogers RO4003C.

and a modification is proposed, eliminating the overshoots at frequencies where the deembedding structure is $\frac{\lambda}{2}$ long.

III. Analysis of Matrix Root Based Bisecton Thru-Only Deembedding

The $ABCD$-parameter representation of the embedded DUT A_{tot} and the deembedding structure A_{Deem} are

$$A_{\mathrm{tot}} = A_{\mathrm{TF,L}} \cdot A_{\mathrm{DUT}} \cdot A_{\mathrm{TF,R}} \tag{1}$$

$$A_{\mathrm{Deem}} = A_{\mathrm{TF,L}} \cdot A_{\mathrm{TF,R}} \tag{2}$$

respectively, with $A_{\mathrm{TF,L}}$, $A_{\mathrm{TF,R}}$ and A_{DUT} being the $ABCD$-parameters of the left and right test fixture and the DUT respectively. Under the assumption $A_{\mathrm{TF,L}} = A_{\mathrm{TF,R}} = A_{\mathrm{TF}}$, A_{TF} can be obtained by

$$A_{\mathrm{TF}} = \sqrt{A_{\mathrm{Deem}}} \tag{3}$$

TABLE I
Test structures and their usage in the validation of the proposed deembedding method

No.	Test Structure	Usage
1	50 Ω line, $l = 1$ mm	Single test fixture
2	50 Ω line, $l = 2$ mm	Deembedding structure
3	50 Ω line, $l = 3$ mm	DUT
4	Beatty standard, $l_{\mathrm{B}} = 2$ mm, $l_{\mathrm{line}} = 1$ mm	DUT
5	50 Ω line, $l = 5$ mm	Embedded DUT
6	Beatty standard, $l_{\mathrm{B}} = 2$ mm, $l_{\mathrm{line}} = 3$ mm	Embedded DUT

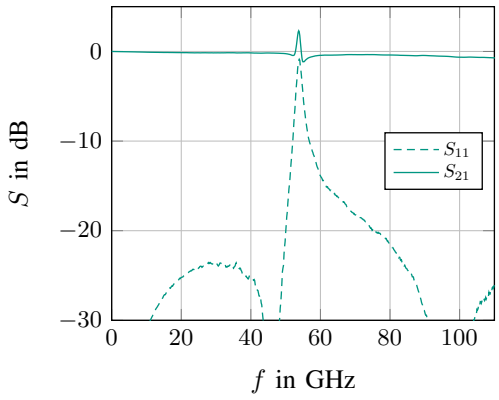


Fig. 5. S-parameters S_{TF} of a single test fixture, obtained using equation 3 from the deembedding structure, which is a 2 mm GCPW 50 Ω line (test structure 2 in Fig. 4).

Using A_{TF}, A_{DUT} and subsequently S_{DUT} can be calculated as

$$A_{\mathrm{DUT}} = A_{\mathrm{TF}}^{-1} \cdot A_{\mathrm{tot}} \cdot A_{\mathrm{TF}}^{-1} \tag{4}$$

Fig. 5 shows S_{TF} for a single test fixture obtained from the deembedding structure by taking the matrix root of A_{Deem} using Eq. 3. S_{TF} shows an overshoot at 55 GHz, where the length of the deembedding structure on the used substrate is $\frac{\lambda}{2}$. This unphysical behavior is therefore introduced in the matrix root step. Therefore, the matrix root operation for deembedding structure lengths of $\frac{\lambda}{2}$ is further analyzed.

The test fixture is in the most cases a connector and a line or a probe pad and a line. Therefore, the deembedding structure is generally a line between either two connectors or two probe pads. The S-matrix S_{Deem} for such a deembedding structure can be described as

$$S_{\mathrm{Deem}} = \begin{pmatrix} r & \sqrt{1-|r|^2} \cdot e^{-\gamma l} \\ \sqrt{1-|r|^2} \cdot e^{-\gamma l} & r \end{pmatrix} \tag{5}$$

with r being the reflection coefficient of the deembedding structure, γ being the complex propagation constant and l being the total length of the deembedding structure. The reflection coefficient r is frequency dependent, but this will not be further elaborated as the matrix root operation is only

analyzed for frequencies, where the deembedding structure is $\frac{\lambda}{2}$ long. Under the assumptions that $r \ll 1$, $l = \frac{\lambda}{2}$ and $\alpha \ll \beta$, with α and β being the real and the imaginary part of γ respectively, A_{Deem} is

$$A_{\mathrm{Deem,l}=\frac{\lambda}{2}} = \begin{pmatrix} -1 & -Z_0 r \\ -\frac{r}{Z_0} & -1 \end{pmatrix} \tag{6}$$

with Z_0 being the system impedance. Using Eq. 3, the matrix root A_{TF} is then calculated using matrix diagonalization [6], resulting in

$$A_{\mathrm{TF,l}=\frac{\lambda}{2}} = \begin{pmatrix} \frac{r}{2} & Z_0 \\ -\frac{1}{Z_0} & \frac{r}{2} \end{pmatrix} \tag{7}$$

When converting $ABCD$- to S-parameters, all parameter conversions have the same denominator, following the conversion formula

$$S_{ik} = \frac{x}{A + \frac{B}{Z_0} + CZ_0 + D} \tag{8}$$

with S_{ik} being an arbitrary S-parameter and x being the respective numerator in the conversion formula to calculate S_{ik} [7]. For all physical networks, including A_{Deem} at all frequencies, the shown denominator in Eq. 8 is ≥ 2. However, for $A_{\mathrm{TF,l}=\frac{\lambda}{2}}$, the denominator in Eq. 8 is < 2, which results in the overshoots in S_{TF} seen in Fig. 5 and therefore also in the overshoots in S_{DUT} seen in Fig. 3. Consequently, A_{TF} does not represent a physically valid network, although it is mathematically correct.

This section makes clear, that and how the overshoots in S_{DUT} are introduced in the matrix root step. Possible solutions for the overshoots need to focus on this step.

IV. Modified Bisection Thru-Only Deembedding

A. Concept

The validity of matrix root based bisection thru-only deembedding is dependent on the length of the deembedding structure and therefore on the phase of S_{Deem}. To ensure validity at all frequencies, the phase of S_{Deem} needs to be modified for the matrix root step.

Fig. 6 shows a modified bisection thru-only deembedding method based on the known matrix root based bisection thru-only deembedding, which will further be called the traditional deembedding method. As first step, the phase of all S-parameters of S_{Deem} is extracted and removed. Afterwards $S_{\mathrm{Deem,mod}}$ is converted to $A_{\mathrm{Deem,mod}}$ and $A_{\mathrm{TF,mod}}$ is calculated using Eq. 3. By removing the phases of S_{Deem} the deembedding structure has virtually zero length and no unphysical behavior is introduced in the matrix root step. Subsequently, the extracted phases of S_{Deem} are divided by 2 and applied on $S_{\mathrm{TF,mod}}$. The division by 2 is necessary because the test fixture has half the length of the deembedding structure. Finally, deembedding using Eq. 4 is performed. As all phases in S_{Deem} and therefore also in $A_{\mathrm{Deem,mod}}$ are removed during the bisection step, a small error is obviously introduced in the whole deembedding process which will be quantified and compared to the traditional deembedding method in the following.

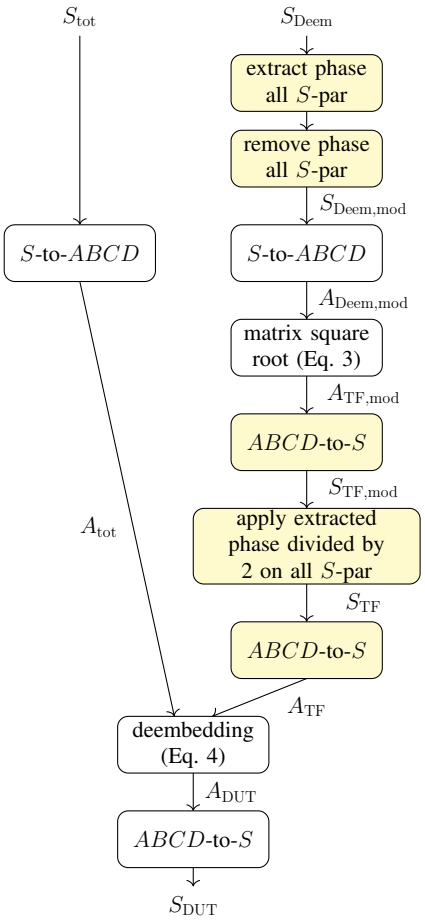


Fig. 6. Modified bisection thru-only deembedding. The white steps are also included in the known matrix root based bisection thru-only deembedding. The yellow steps are introduced in the modified bisection thru-only deembedding.

B. Validation

The presented modified bisection thru-only deembedding method is verified in simulation and measurement. For the test structures, shown in Fig. 4 and denoted in Table I, $50\,\Omega$ lines have been chosen as they are comparably simple and tolerant against manufacturing inaccuracies, which makes the comparison between deembedding and measurement easier. Additionally, Beatty standards are chosen which are widely used for PCB material characterization [8]. To validate the performance of the modified deembedding method, the deembedded S_{DUT} shall be compared to its measured counterpart. For this comparison to be as accurate as possible, the deembedded and measured DUT must be exactly the same structure. That is why all test structures don't contain explicit probe pads or connectors. The measured S-parameters are obtained using a Keysight PNA in combination with Keysights N5250 DC-110 GHz measurement system. The test structures are

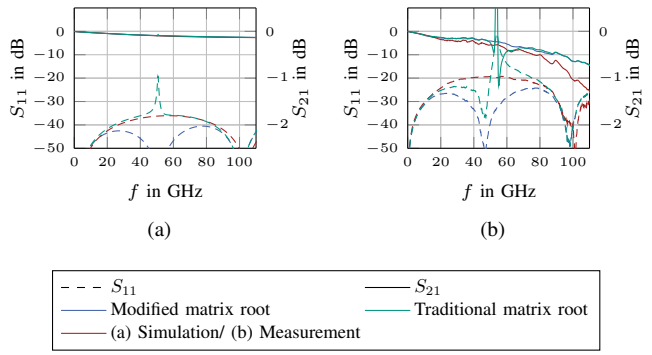


Fig. 7. S_{TF} calculated by the traditional and the modified matrix root step compared to the actual S_{TF} for test structure 2 in Fig. 4 in (a) simulation and (b) measurement.

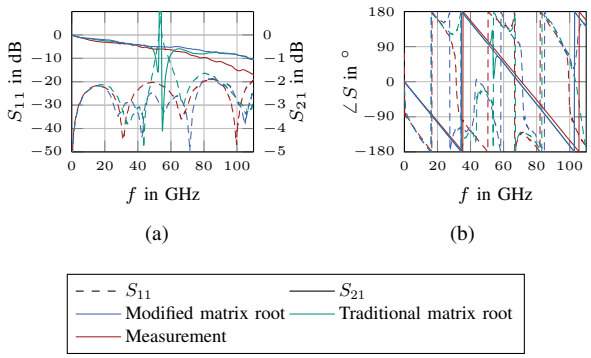


Fig. 8. S_{DUT} calculated by the traditional and the modified deembedding method compared to the measured S_{DUT}. The DUT is test structure 3 from Fig. 4, the deembedding structure is test structure 2 and and the embedded DUT is test structure 5.

contacted with probes with a 150 µm pitch. A first tier Load-Reflect-Reflect-Match (LRRM) calibration on an impedance standard substrate is performed.

1) Matrix root step: At first the result of the matrix root operation S_{TF} is evaluated on test structure 2, a 2 mm GCPW 50 Ω line, from Fig. 4, which represents the deembedding structure, to assess the validity of the modified matrix root step. The calculated S_{TF} are compared to simulation and measurement data of test structure 1, a 1 mm GCPW 50 Ω line, from Fig. 4, which represents a single test fixture. Fig. 7 shows a comparison of the calculated and the actual S_{TF}.

In simulation and measurement, there are no overshoots in S_{TF} for the modified matrix root step compared to the traditional matrix root step. At frequencies further away from where the deembedding structure is $\frac{\lambda}{2}$ long, both methods result in similar S-parameter, with the ones obtained by the traditional method being slightly closer to the actual S_{TF}. This makes sense, as obviously there is a small error introduced by omitting the phase in the matrix root step.

2) Entire deembedding method: The final deembedded S_{DUT} for the traditional and the modified deembedding method is also compared to the measured S_{DUT}. Fig. 8a shows the deembedded S_{DUT} in magnitude obtained by deembedding a 5 mm GCPW 50 Ω line (test structure 5 from Fig. 4) with a 2 mm GCPW 50 Ω line (test structure 2) as deembedding structure. Test structure 3, which is a 3 mm GCPW 50 Ω line, is the reference DUT. Fig. 8b shows the phase of the deembedded S_{DUT}. To further quantify the two matrix root methods, the errors ΔS_{ik}, ΔS_{ref}, $\Delta S_{\mathrm{trans}}$ and ΔS_{tot} on the obtained S-parameters are computed via

$$\Delta S_{ik} = \frac{1}{n_f} \sum_f |S_{ik,\mathrm{deem}} - S_{ik,\mathrm{meas}}| \tag{9a}$$

$$\Delta S_{\mathrm{ref}} = \Delta S_{11} + \Delta S_{22} \tag{9b}$$

$$\Delta S_{\mathrm{trans}} = \Delta S_{21} + \Delta S_{12} \tag{9c}$$

$$\Delta S_{\mathrm{tot}} = \Delta S_{\mathrm{ref}} + \Delta S_{\mathrm{trans}} \tag{9d}$$

TABLE II
S-PARAMETER ERRORS USING EQS. 9 FOR GCPW 50 Ω LINES.

	ΔS_{ref}	$\Delta S_{\mathrm{trans}}$	ΔS_{tot}
Traditional deembedding method	0.219	0.338	0.557
Modified deembedding method	0.094	0.282	0.376

with n_f being the number of frequency points and $S_{ik,\mathrm{deem}}$ and $S_{ik,\mathrm{meas}}$ the deembedded respectively measured S-parameters. The S-parameter error values can be seen in Table II.

In Fig. 9 the dembedded S_{DUT} are shown in magnitude and phase for a 5 mm GCPW Beatty standard with a 2 mm mismatched section (test structure 6), again deembedded with a 2 mm GCPW 50 Ω line (test structure 2). The reference DUT is a 3 mm GCPW Beatty standard with a 2 mm mismatched section (test structure 4). The S-parameter error values are shown in Table III.

From Figs. 8 and 9, it is apparent that the modified deembedding method shows no overshoots in the deembedded S_{DUT} in contrast to the traditional method. Further away from frequencies where the deembedding structure is $\frac{\lambda}{2}$, both S_{DUT} obtained by the two investigated deembedding methods show very similar values, which should be the case as the traditional method is mathematically correct at these frequencies. Furthermore, the phases of the deembedded and measured S-parameters match very well. This validates the concept of extracting the phase before the matrix root step and applying it afterwards.

Both deembedded S_{DUT} show differences to the measured S_{DUT}, which come from differences in the physicality of the deembedded and measured DUT. The measured DUT is contacted by two probes whereas for the deembedded DUT the probing is eliminated. The comparison of S_{11} is clearly affected by this. As the probes are not placed at the very edge of the DUT, there is also an electrical length difference for the measured and the deembedded DUT, which results in a mismatch in the measured and deembedded phases of S_{DUT} as can be seen in Figs. 8b and 9b. For those reasons,

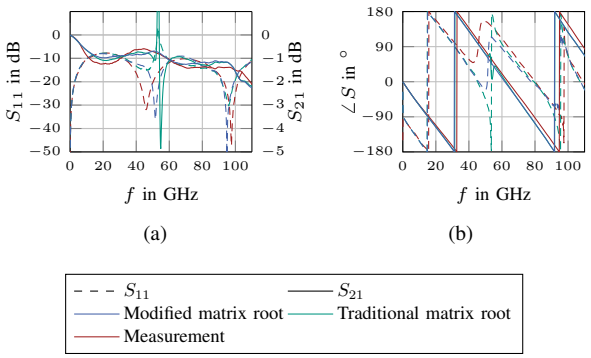


Fig. 9. S_{DUT} calculated by the traditional and the modified deembedding method compared to the measured S_{DUT}. The DUT is a 3 mm GCPW Beatty standard with a 2 mm mismatched section (test structure 4 from Fig. 4), the deembedding structure is test structure 2 and and the embedded DUT is a 5 mm GCPW Beatty standard with a 2 mm mismatched section (test structure 6.)

TABLE III
S-PARAMETER ERRORS USING EQS. 9 FOR GCPW BEATTY STRUCTURES.

	ΔS_{ref}	$\Delta S_{\mathrm{trans}}$	ΔS_{tot}
Traditional deembedding method	0.202	0.353	0.555
Modified deembedding method	0.155	0.343	0.497

the error values presented in Tables II and III can not be evaluated on their own and only serve as a comparison. For both different DUTs, the modified deembedding method has a smaller error, therefore presenting an interesting alternative on known deembedding methods.

C. Limitations

Regarding the deembedding structure, the deembedding algorithm has a limitation. Fig. 10 shows the simulated and deembedded S_{DUT} for different return losses of the deembedding structure. The DUT, deembedding structure, and the embedded DUT are all GCPW lines, and their lengths match those of test structures 3, 2, and 5 from Figure 4. To achieve the desired return loss values for all these structures, the line geometry was adjusted. The shown S-parameters are obtained by simulation.

Higher return loss values in deembedding structures lead to larger discrepancies between simulated and deembedded S-parameters. This is due to the matrix root step, where higher return loss values amplify the influence of the phase of all S-parameters on the matrix root. Accordingly, the error caused by extracting the phase is higher, which leads to the seen deviation between the simulated and deembedded S-parameters. For a tolerable error, the return loss of the deembedding structure should be preferably $-15\,\mathrm{dB}$ or better.

V. CONCLUSION

This work presents a thorough investigation on bisection thru-only deembedding for electrically long test fixtures. Known bisection thru-only deembedding methods show unphysical behavior at frequencies where the test fixtures are $\frac{\lambda}{2}$ long. As this behavior is rooted in the length and therefore the phase of the S-parameters of the deembedding structure, the traditional matrix root based bisection thru-only deembedding algorithm is modified regarding the phase of the deembedding structure, thus eliminating the mentioned unphysical behavior in S_{DUT}. Test measurements confirm the improved deembedding accuracy for the modified deembedding method compared to the traditional one. Furthermore, the modified method is analysed regarding its use-case limitations. The recommendation is to use it with deembedding structure with a low return loss.

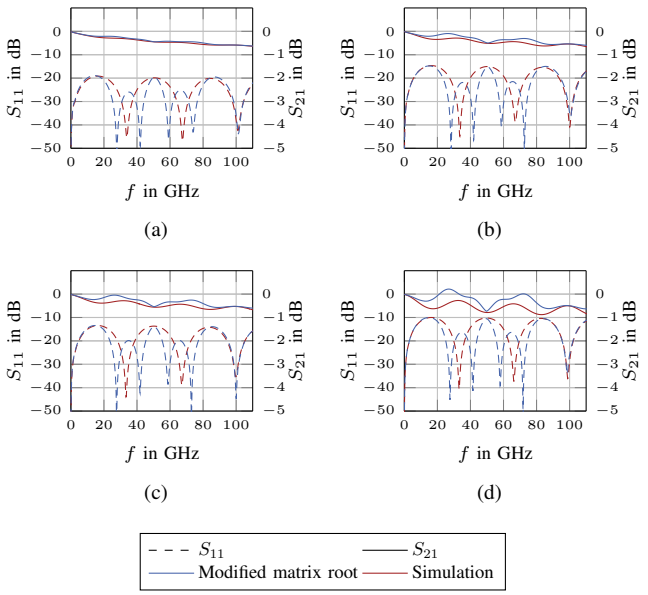


Fig. 10. S_{DUT} calculated by the the modified deembedding method for different return losses of S_{Deem}: (a) $-20\,\mathrm{dB}$, (b) $-15\,\mathrm{dB}$, (c) $-13\,\mathrm{dB}$ and (d) $-10\,\mathrm{dB}$ compared to simulated S_{DUT}. The DUT is based on test structure 3 from Fig. 4, the deembedding structure on test structure 2 and and the embedded DUT on test structure 5, all with modified geometries to reach the specified return loss values.

ACKNOWLEDGEMENT

This work is funded by the German Research Foundation (DFG) under the project "ActiOnPro" (grant number ZW 180/28-1). This work is financially supported in part by the BMBF programme Forschungslabore Mikroelektronik Deutschland (For-Lab).

REFERENCES

[1] P. Heymann, H. Prinzler, and F. Schnieder, "De-embedding of MMIC transmission-line measurements," in *1994 IEEE MTT-S International Microwave Symposium Digest*, San Diego, USA, May 1994, pp. 1045 – 1048.

[2] A. Mangan, S. Voinigescu, M.-T. Yang, and M. Tazlauanu, "De-embedding transmission line measurements for accurate modeling of IC designs," *IEEE Trans. Electron Devices*, vol. 53, no. 2, pp. 235–241, Feb. 2006.

[3] T. Sekiguchi, S. Amakawa, N. Ishihara, and K. Masu, "On the validity of bisection-based thru-only de-embedding," in *2010 International Conference on Microelectronic Test Structures (ICMTS)*, Hiroshima, Japan, Mar. 2010, pp. 66–71.

[4] T. Kolding, "On-wafer calibration techniques for giga-hertz CMOS measurements," in *1999 International Conference on Microelectronic Test Structures (ICMTS)*, Gothenburg, Sweden, Mar. 1999, pp. 105 – 110.
[5] Y. Goto, Y. Natsukari, and M. Fujishima, "New On-Chip De-Embedding for Accurate Evaluation of Symmetric Devices," *Japanese Journal of Applied Physics*, vol. 47, no. 4, pp. 2812–2816, Apr. 2008.
[6] N. J. Higham, *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9780898717778
[7] D. Pozar, *Microwave Engineering*. Wiley, 2012.
[8] R. Beatty, "2-Port lambda/4 Waveguide Standard of Voltage Standing-Wave Ratio," *Electronics Letters*, vol. 9, no. 2, pp. 24–26, Jan. 1972.

Evaluation of Lab-based Lithium Niobate Surface Acoustic Wave Test Structure Using Efficient Maskless Lithography and SMA Connection Approach for Microfluidic Applications

Mohammad Salman Parvez[1], Shadhin Hussain[1], Matthew Goeckner[1], Chadwin D. Young[1], and Jeong-Bong (JB) Lee[2]

1. The University of Texas at Dallas, Richardson, TX 75080, USA
2. Baylor University, Waco, TX
Corresponding Author
Chadwin D. Young
The University of Texas at Dallas, 800 W. Campbell Rd., RL10, Richardson, TX 75080, USA
Email: chadwin.young@utdallas.edu

***Abstract*— Surface Acoustic Wave (SAW) devices, particularly those made of Lithium Niobate ($LiNbO_3$), are extensively used in telecommunications and microfluidic applications. This work describes the fabrication of a $LiNbO_3$-based SAW test structures using maskless photolithography for rapid device dimension changes as well as introduces a cost-effective technique to solder Sub-Miniature version A (SMA) connectors to delicate substrates such as $LiNbO_3$-based SAWs, without the need to fully create printed circuit boards. The SMA connected device facilitated improved characterization results compared to simple copper tape connections. The characterization accuracy is then validated through simulation.**

Keywords*— *Surface Acoustic Wave (SAW) devices, Lithium Niobate ($LiNbO_3$), Sub-Miniature version A (SMA).

I. INTRODUCTION

Lithium niobate ($LiNbO_3$) based Surface Acoustic Wave (SAW) devices have been used in telecommunication industries [1, 2], while being explored in microfluidic devices such as sensors [3, 4] and actuators [5], for mixing [6], pumping [7], jetting, and atomization [5]; however, $LiNbO_3$ is fragile and susceptible to thermal stress. For these applications, SAW devices must function at a resonant frequency that depends on the type of piezoelectric material, the design parameters of test structure [e.g., Interdigitated Electrode (IDT) [8], electrode material, etc.]. Theory can predict this resonant frequency [9], but this resonant frequency must be based on the characterization results for these devices under test (DUTs). Characterization results are greatly influenced by the type and the quality of the DUT connection to the external generator (i.e., the network analyzer). In lab-based applications, PCB connections [10] provide very good results, but can be expensive and time consuming to make, whereas lump connections [11] are cheap, but not very accurate. Alternatively, SMA connectors can be a cost-effective way to minimize transmission loss and improve accuracy. Therefore, a novel technique using a 'filler' for soldering SMA connectors to $LiNbO_3$ substrates is explored with improved characterization results compared to a lump connection (e.g., copper tape). Moreover, we verify the accuracy of the measured resonant frequency with a simulated lithium niobate SAW device.

II. DEVICE PREPARATION

A. Fabrication by Maskless Photolithography

The fabrication process for $LiNbO_3$ wafers, (to be specific Y-cut 128º, X axis) involves several crucial steps to ensure the creation of precise and functional Interdigital Transducers (IDTs) test structures. As shown in **Fig. 1**, the wafer undergoes a meticulous cleaning process using DI water, Isopropyl Alcohol, and Acetone followed by a brief bake at 115º C for 5 minutes. To enhance bonding, a Hexamethyldisilazane (HMDS) process is incorporated prior to spin coating the wafer with LOR 5A and AZ1512 photoresist and soft bake. A pivotal maskless photolithography process is executed using the MicroLight3D Smart Print UV Maskless Lithography System for rapid adaptation of IDT structural geometries. The final steps deposit a Titanium adhesive layer (30-50 nm) and a gold layer (120-150 nm) using an E-beam evaporator. The fabrication concludes with a lift-off process. This comprehensive process ensures the successful creation of intricate structures on the $LiNbO_3$ substrate for diverse applications.

B. Sub-Miniature version A (SMA) connection

The soldering process of the SMA (Molex SMA jack edge mount 50 Ohms, EWR-11795 SMA-J/PCB) to the fragile $LiNbO_3$ substrate is a complex procedure, requiring meticulous care (**Fig. 2**). In the initial step (**Fig. 2a**), precise alignment of the signal and ground pins of the SMA connector with the corresponding pads of the fabricated Interdigital Transducers

979-8-3503-2987-2/24/$31.00 ©

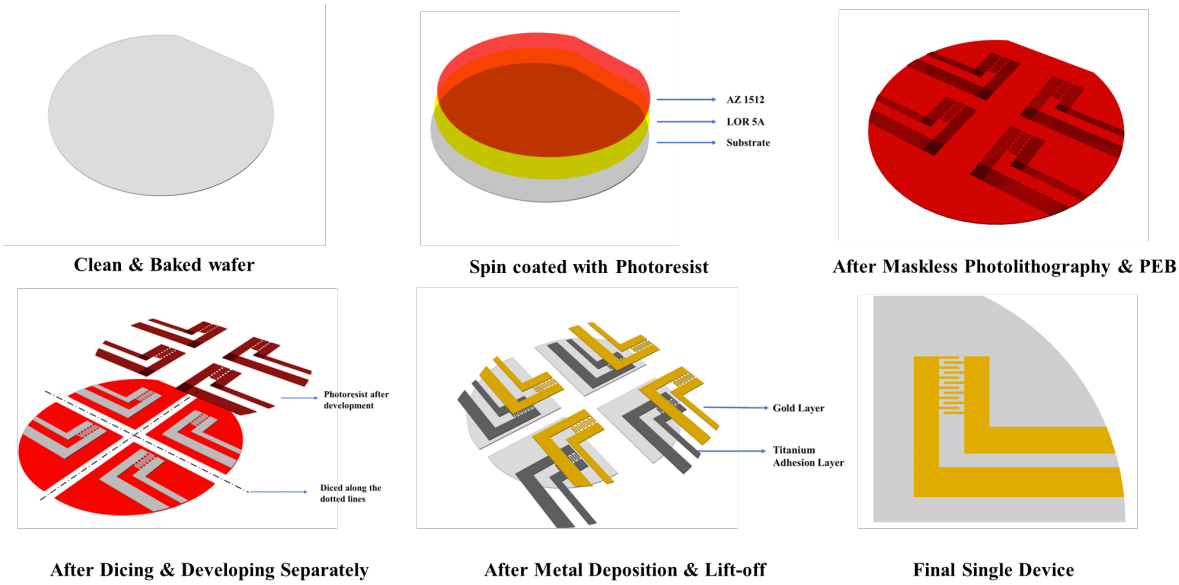


Fig 1: Fabrication Sequence of Maskless Surface Acoustic Wave Device (SAW) on a Y-cut 128º, X axis $LiNbO_3$ substrate

(IDTs) test structure is crucial. The design of the IDT pads considers the dimensions of the SMA connector pins **(Fig. 3)**, ensuring a reasonable gap to prevent shorting when soldering. To facilitate attachment, the IDT connection pads are strategically designed to touch the wafer's edge. The second step **(Fig. 2b**) involves inserting a filler between the bottom surface of the $LiNbO_3$ substrate and the SMA connector to temporarily secure it in the aligned position. The third and pivotal step **(Fig. 2c)** is the soldering of SMA connector pins to the IDT pads within a temperature range of 140 ºC – 150 ºC for the specified metal. As illustrated in **Fig. 2c (Red Circle)**, the molten solder metal has been dropped carefully on the interface of the pins and IDT pads, without the hot solder bar touching the substrate. Post-soldering, the filler is gently removed, and the SMA connector is delicately secured **(Fig. 2d)**. To enhance mechanical strength, Gorilla™ Epoxy Glue is applied **(Fig. 2e)** to connect the bottom surface of the SMA pins to the substrate. This assembly is left to cure at room temperature for 10-12 hours, thereby reducing mechanical stress from the solder connection and enhancing the device's durability.

C. Verification of the connection

After completing the soldering process, verification of the connections is necessary. First, there must be no short connection between the two pads of the IDTs resulting from the soldering and fabrication. To verify the short connections **(Fig. 4a),** the probes are connected to the two opposite IDT pads of the dummy Si sample and the resistance was checked by using a multimeter. Since there is no finite value for the resistance, the result confirms the connections are not shorted. While connecting the two opposite pads with the probes, care must be taken to not scratch off the deposited gold on the substrate. Next, to test the signal connection, one of the probes is placed at the tip of the signal connection of the SMA while another probe is connected to the IDT signal pin pad, **(Fig. 4b)**, where the resistance is a finite value which means that the connection is established. Using the same approach **(Fig. 4c)**, the ground connection has been verified by placing one of the probes on the SMA ground connector and the other one on the IDT pad with the soldered ground pin. The finite value for the resistance

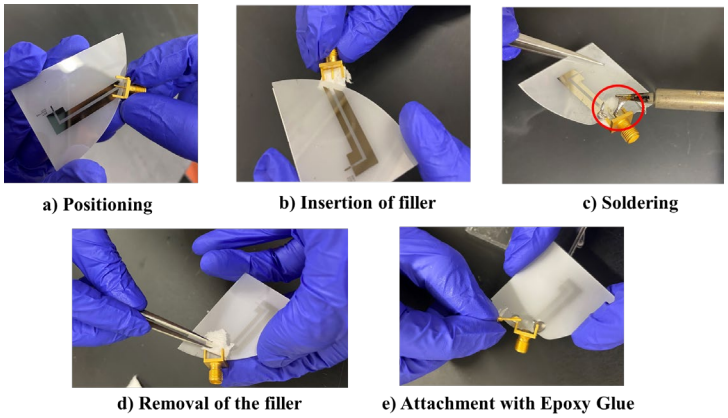


Fig. 2: Procedure for attaching SMA connector to the fragile $LiNbO_3$ substrate.

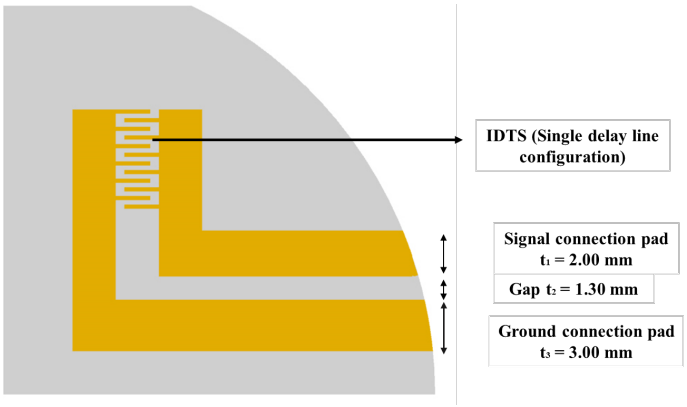
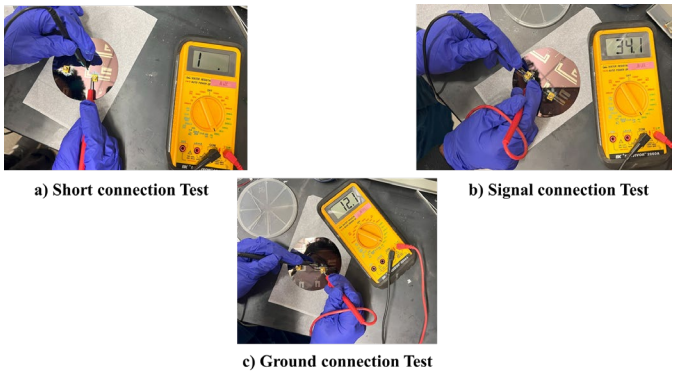


Fig. 3: Example layout of the surface acoustic wave (SAW) test structure along with associated dimensions.



a) Short connection Test

b) Signal connection Test

c) Ground connection Test

Fig. 4: Verification process of the soldered SMA connection (For picture taking purpose Si is used so that it can be used casually)

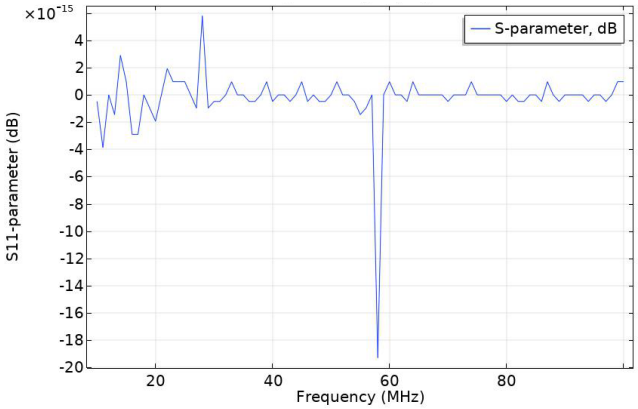


Fig. 5: S11 simulation plot for a 60µm Wavelength Device (Unit cell IDT pair)

provides confirmation of the established connection. The same process has been used to test the other samples fabricated on $LiNbO_3$ substrate.

III. Characterization Process

Next, the SAW DUTs have been tested using a Vector Network Analyzer (VNA) (Agilent technologies, E5061B, 5Hz-3GHz) for measuring the S11 parameter and the Smith chart. To maximize the data accuracy, the background impedance due to the cable and connectors were subtracted using a built-in feature of the VNA before any measurements were made. The SMA of the test structure was connected to the 50 Ω input port of the VNA by using a Bayonet Neill–Concelman or BNC connector. While connecting the test structure to the BNC port, care must be taken for a secure connection, while having no damage to the solder of the test structure. To get the output plots, a maximum number of data points (1601) has been set. The range for the testing frequency has been kept close to the approximated value of the resonant frequency of the designed device in order to ensure more accuracy in the results. Note that all the tests have been done at room temperature.

IV. Results and Discussions

Figs, 5 & 6 show the simulation results of 60 µm & 70 µm wavelength DUTs should be for the unit cell IDTs that serve as an initial reference for the resonant frequency. This resonant frequency can be approximated to be ~ 60 MHz for 60 µm wavelength DUT and above 50MHz for 70 µm wavelength DUT (dip shown in the plot). **Fig. 7** demonstrates an improperly configured test structure using copper tape connections where the S11 results have a huge shift from the approximated resonant frequency around 115 MHz. However, the SMA connected 60 µm wavelength DUT has a sharp dip of about -23 dB at the resonant frequency of 63.9 MHz, slightly more than the simulated result (**Fig. 8**). This small discrepancy from the simulated value could be due to the number of electrode pairs (39 for SMA 60 µm wavelength connected DUT), or possibly the variation in the acoustic velocity from the simulated material parameters and the real material. Apart from the simulation, the resonant frequency can be calculated using the equation [12]:

$$v = f \times \lambda \qquad (1)$$

where v is the wave velocity and f is the resonant frequency and λ is the wavelength of the IDTs.

Care was taken during the fabrication to align all the IDTs facing towards the X-propagation direction since it is a Y-cut 128º, X-propagating $LiNbO_3$ substrate to ensure the maximum power transformation [12]. However, there are other factors involved that could hinder the power transmission and cause power losses at the expected resonant frequency. Possible factors include: electrical reflection, electrical dissipation, parasitic radiation, etc. However, among all these electrical reflections has been given more importance in this finding since all these devices work in (RF) frequency range for which a matching network connection is a must [8]. The parameter which measures the amount of power coupled to the SAW DUT with respect to the power supplied by the source is called the reflection co-efficient of power (r_p)

$$r_p = \frac{|Z - Z_o|^2}{|Z + Z_o|^2} \qquad (2)$$

where Z_0 = 50 Ω and Z = Impedance of DUT [8] considering the standard characteristic impedance match of 50 Ω for Radio Frequency (RF) cables. **Fig. 7** illustrates the impedance at 63.9 MHz is 48.5 Ω to be precise and using equation (2) the reflection co-efficient of power at resonant frequency for the DUT of 60 µm is 2.3×10^{-4} (0.023%) which is almost 0. Another DUT of 70 µm wavelength observed an S11 parameter of about -30.9 dB at a resonant frequency of 54.6 MHz shown in **Fig. 10**. Looking at **Fig. 11**, the impedance at that frequency is 50.05 Ω for which the reflection co-efficient of power at the resonant frequency for the DUT of 70 µm is 2.5×10^{7} (2.5×10^{-7} %) which infers no electrical reflection. If we compare these results with Winkler et al., [8] where PCB board and gold-plated spring pins have been used for connection, the reflection co-efficient of power at the resonant frequency for a DUT of 30 µm wavelength with finger pairs of 34 & 80 are 0.008 (0.8%) & 0.181(18.1%) respectively. So, comparing these two results with the DUTs mentioned

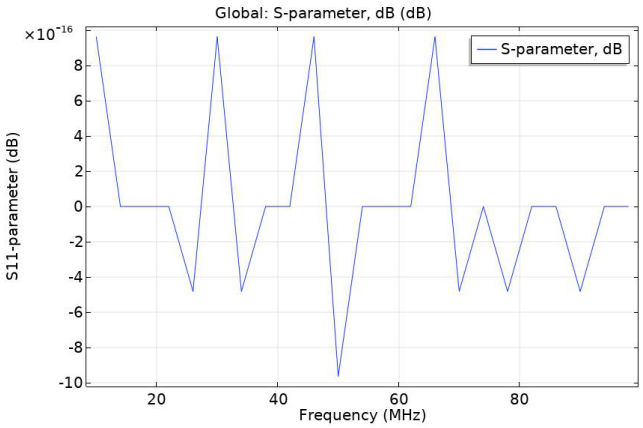


Fig. 6: S11 simulation plot for a 70µm Wavelength unit cell IDT pair)

above, direct SMA connection provides better power transmission to the IDTs.

Again, comparing the S11 curves of **Figs. 7, 8** and **10**, analysis shows that the curves of **Fig. 8 & 10** are noise free; especially at the resonant frequency. Therefore, the new SMA connection approach enables robust impedance matching, noise-free and better overall performance compared to the copper tape connection as well.

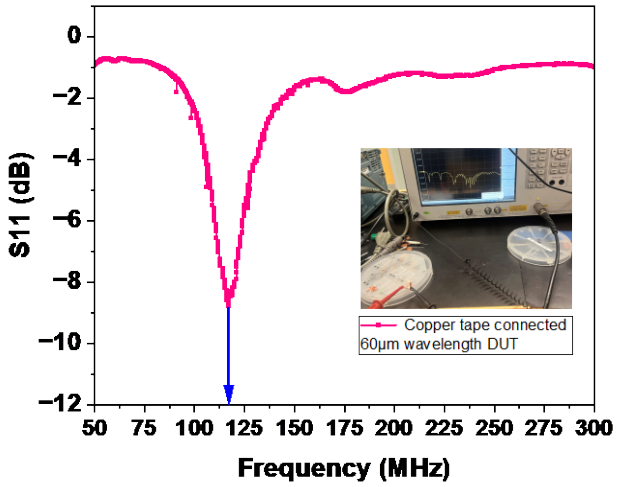


Fig: 7: S11 experimental result for a simplified copper tape connection to the 60µm Wavelength Device

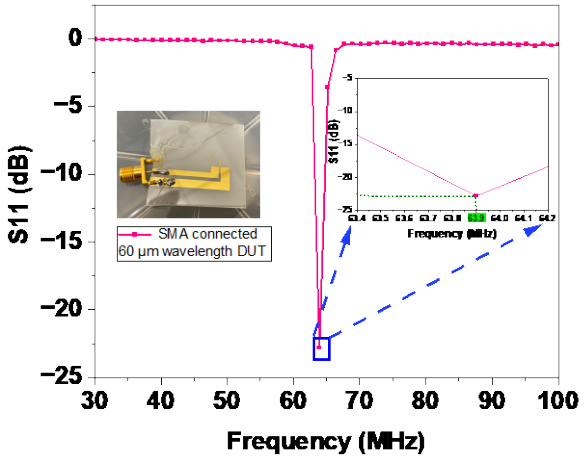


Fig. 8: S11plot for SMA connected 60µm Wavelength DUT

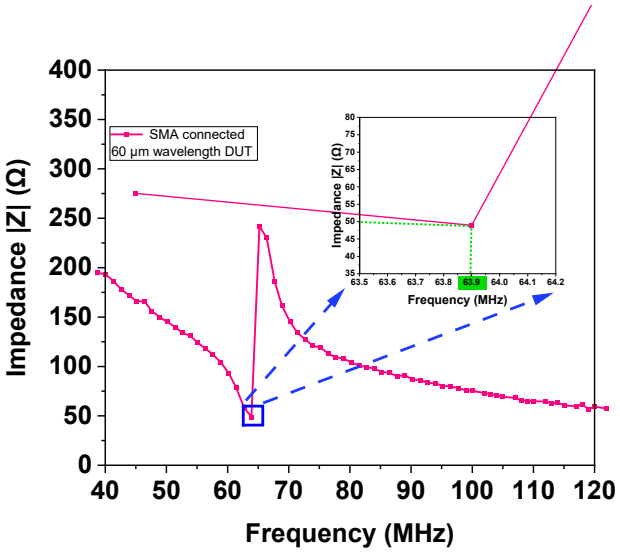


Fig. 9: Impedance plot vs. frequency for SMA connected 60µm Wavelength DUT

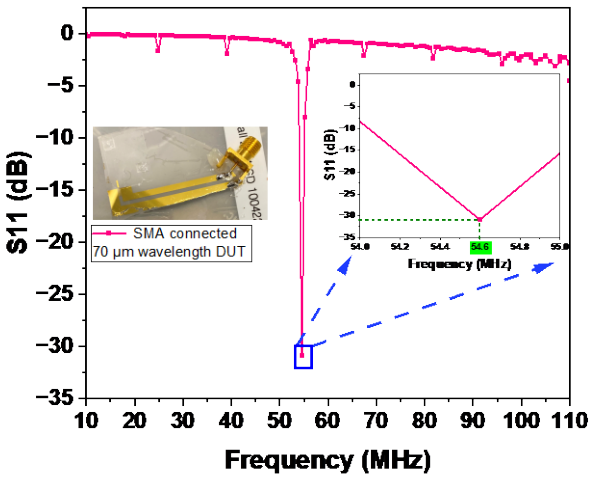


Fig. 10: S11plot for SMA connected 70µm Wavelength DUT

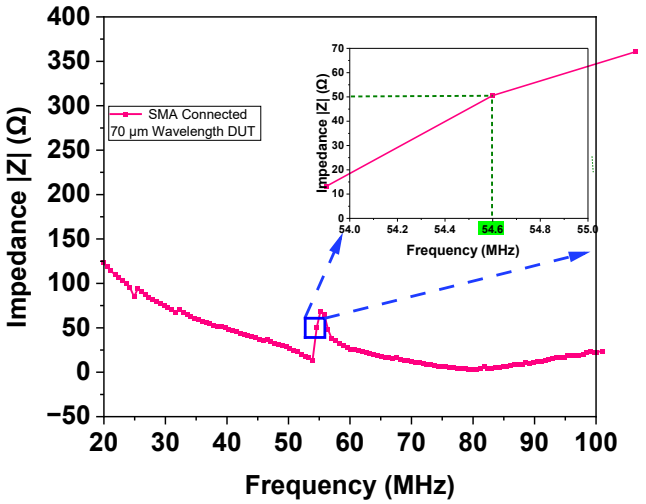


Fig. 11: Impedance plot vs. frequency for SMA connected 70µm

CONCLUSION

Surface acoustic wave test structures have been fabricated with either a copper tape connection or soldered SMA connection. The SMA connection exhibited far better results that closely matched the simulated resonant frequency coupled with better impedance matching compared to the copper tape connection. Also, these results have been compared with PCB connected test structures [12] to evaluate the impedance matching and reflection co-efficient of power, where the SMA connected test structures convincingly achieved similar results to PCB connected devices.

ACKNOWLEDGMENT

I would like to thank the members of UTD Minds lab, Dr. Young's lab and UTD cleanroom staff.

REFERENCES

[1] C. K. Campbell., Surface Acoustic Wave Devices for Mobile and Wireless Communications, San Diego, CA,: Academic, 1998.

[2] L. G. ,. S. G. ,. Yansong Yang, "Surface-Acoustic-Wave Devices Based on Lithium Niobate and Amorphous Silicon Thin Films on a Silicon Substrate," IEEE Transactions on Microwave Theory and Techniques, vol. 70, no. 11, pp. 5185 - 5194, 2022.

[3] J. Devkota, P. R. Ohodnicki, and D. W. Greve, "SAW sensors for chemical vapors and gases," Sensors (Switzerland), vol. 17, no. 4. MDPI AG, Apr. 08, 2017

[4] D. Mandal and S. Banerjee, "Surface AcousticWave (SAW) Sensors: Physics, Materials, and Applications," Sensors, vol. 22, no. 3, Feb. 2022

[5] L. Y. Yeo and J. R. Friend, "Surface acoustic wave microfluidics," Annu Rev Fluid Mech, vol. 46, pp. 379–406, Jan. 2014

[6] H. Lim, S. M. Back, H. Choi, and J. Nam, "Acoustic mixing in a dome-shaped chamber-based SAW (DC-SAW) device," Lab Chip, vol. 20, no. 1, pp. 120–125, Jan. 2020

[7] A. Zhang, Y. Zha, and J. Zhang, "A surface acoustic wave micropump to pump fluids from a droplet into a closed microchannel using evaporation and capillary effects," AIP Adv,vol.4,no.12,Dec.2014

[8] A. Winkler, R. Brünig, C. Faust, R. Weser, and H. Schmidt, "Towards efficient surface acoustic wave (SAW)-based microfluidic actuators," Sens Actuators A Phys, vol. 247, pp. 259–268, Aug. 2016

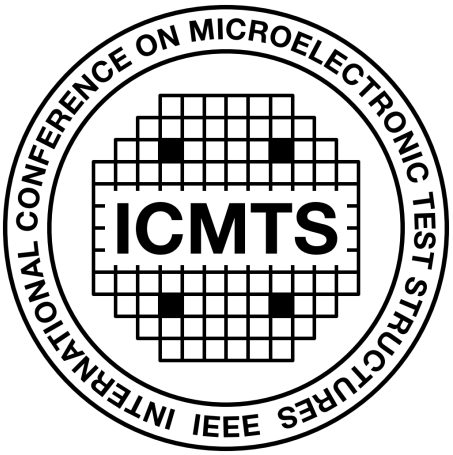
[9] R. Weser, A. Winkler, M. Weihnacht, S. Menzel, and H. Schmidt, "The complexity of surface acoustic wave fields used for microfluidic applications," Ultrasonics, vol. 106, p. 106160, Aug. 2020

[10] B. Yildirim, S. U. Senveli, R. W. R. L. Gajasinghe, and O. Tigli, "Surface Acoustic Wave Viscosity Sensor with Integrated Microfluidics on a PCB Platform," in IEEE Sensors Journal, Institute of Electrical and Electronics Engineers Inc., Mar. 2018, pp. 2305–2312

[11] R. Rimsa, "Surface acoustic wave microfluidic pumps for on-chip diagnostics," 2018.

[12] D. Mandal, and S. Banerjee, "Surface acoustic wave (SAW) sensors: Physics, materials, and applications". Sensors, 2022, 22(3), p.820.

SESSION 8
Gallium Nitride Technology



18th April 2024, 09:30–10:50

Session Co-Chairs: Johan Klootwijk, *Philips Research, Netherlands*
Kejun Xia, *TSMC, USA*

Test Structures to Investigate ESD Robustness of Integrated GaN Devices

Wei-Cheng Wang and Ming-Dou Ker*
Institute of Electronics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan
wayne8712.ee10@nycu.edu.tw, mdker@ieee.org*

***Abstract*— When more circuit functions are integrated into a single chip fabricated by GaN-on-Silicon process, the on-chip electrostatic discharge (ESD) protection design shall be provided to protect the GaN integrated circuits. In this work, ESD robustness of E-HEMT GaN devices was investigated through test structures that fabricated in a GaN-on-Silicon process. The experimental results showed that the ESD robustness is proportional to the device dimension when the GaN device was operating in the forward mode. In addition, with the gate-coupled design, the ESD level of E-HEMT GaN device can be further improved. Based on the investigation results of this work, the whole-chip ESD protection scheme can be successfully realized by E-HEMT GaN devices.**

Keywords – Electrostatic discharge (ESD), transmission line pulse (TLP), human body model (HBM), GaN, E-HEMT.

I. Introduction

With wide bandgap, high breakdown voltage, high carrier density, and high saturation velocity, Gallium Nitride (GaN) has been used in high-frequency and high-power applications. Recently, GaN gate driver and GaN switch had been fully integrated together in a monolithic single chip [1], [2]. The trend of integration with more circuit functions in a single chip will become mainstream for GaN applications. With more circuits realized with GaN devices in a single chip, they must be considered with on-chip electrostatic discharge (ESD) protection. Several ESD protection designs for stand-alone GaN high electron mobility transistor (HEMT) were reported [3], [4]. But, there was no ESD protection for fully integrated GaN circuits.

The whole-chip ESD protection scheme for GaN integrated circuits (IC) is shown in Fig. 1. The input and output pins can be protected with dual GaN diodes those connected from the pad to VCC and VSS, respectively. In addition, the power-rail ESD clamp circuit is placed between VCC and VSS power lines to provide the whole-chip protection function against ESD events happening at any pin of the GaN IC. In this work, test structures to investigate ESD robustness of GaN devices, for applying in the whole-chip ESD protection scheme, were performed with the test chip fabricated in a 0.5-μm GaN-on-Silicon technology [5].

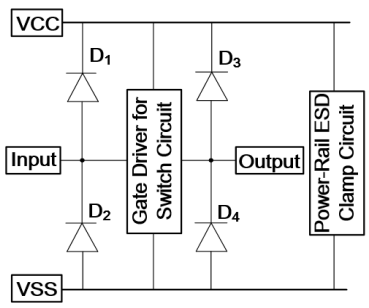


Fig. 1. Whole-chip ESD protection scheme for GaN integrated circuits.

II. Test Structures of GaN Devices

The cross-sectional view of a 12-V enhanced-mode HEMT (E-HEMT) GaN device is depicted in Fig. 2, which was realized by a GaN-on-Silicon technology [5]. The I_D-V_D curves of the fabricated 12-V E-HEMT with gate width of 5 μm and gate length of 1 μm under different gate biases are shown in Fig. 3(a). The corresponding I_G-V_G curve of this 12-V E-HEMT with V_D of 12 V is drawn in Fig. 3(b). The threshold voltage of this 12-V E-HEMT is 1.4 V.

Because there is no pure p-n diode structure provided in the GaN-on-Silicon technology, the diodes (D_1, D_2, D_3, and D_4) for using in the whole-chip ESD protection scheme can be built by the lateral field-effect rectifier (LFER) [6], which is based on the gate-source-shorted E-HEMT device. The structure of the LFER diode is illustrated in Fig. 4. One set of test structures with the E-HEMT GaN device and the LFER diode has been fabricated with different device dimensions to investigate their ESD robustness.

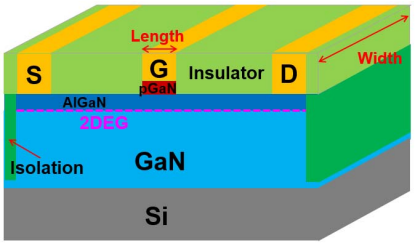


Fig. 2. The cross-sectional view of 12-V E-HEMT GaN device.

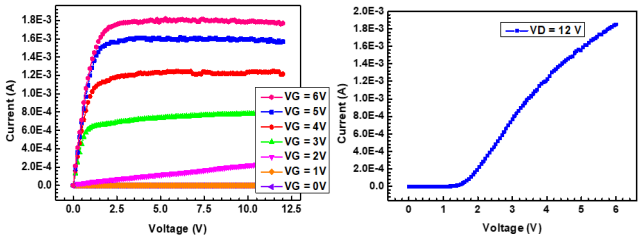


Fig. 3. The (a) I_D-V_D curves of the 12-V E-HEMT under different gate biases and (b) I_D-V_G curve of the 12-V E-HEMT with V_D of 12 V.

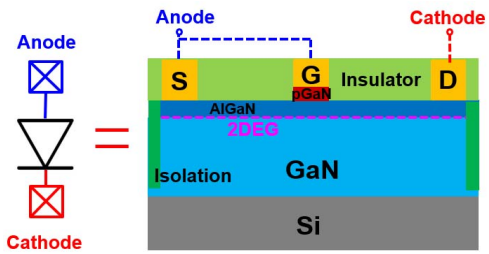


Fig. 4. The schematic and cross-sectional view of a LFER (lateral field-effect rectifier) diode built by the gate-source-shorted E-HEMT.

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

III. Experimental Results

To investigate ESD robustness of the test structures, TLP (transmission line pulse) system and HBM (human body model) ESD tester are used to examine the characteristics of the test devices. The measurement setup of the TLP and HBM ESD test on the LFER diodes under forward and reverse modes is presented in Figs. 5(a) and 5(b).

The TLP-measured *I-V* curves in both modes are depicted in Figs. 6(a) and 6(b). From the *I-V* curves, it can be observed that it is different from a classical PN diode in the forward mode. The relations between HBM ESD level and I_{t2} (secondary breakdown current) of GaN diodes with different dimensions under forward and reverse modes are shown in Figs. 7(a) and 7(b), respectively. Under the forward mode, ESD robustness (HBM ESD level and TLP-measured I_{t2}) exhibits a direct proportionality to the device dimension. However, in the reverse mode, the HBM ESD level and I_{t2} were very low, even if the device dimension (width) was increased up to 1200 μm.

Because GaN diodes only have the good characteristic of unidirectional conduction in the forward mode. The ESD protection design with only diodes (D_1, D_2, D_3, and D_4) is insufficient to establish the good ESD current path for any pin under ESD test. For example, the ESD current paths for PS mode (a positive ESD voltage appears at the I/O pin relative to the grounded VSS pin with other pins floating) and NC mode (a negative ESD voltage appears at the I/O pin relative to the grounded VCC pin with other pins floating) were not well provided by diodes. Therefore, the power-rail ESD clamp circuit was invented to provide the effective ESD current discharging paths for achieving whole-chip ESD protection deign.

To design the power-rail ESD clamp circuit with gate-coupled design, the gate-coupled effect on the E-HEMT GaN device needs to be investigated in advance. The E-HEMT GaN device with a total width of 2000 μm and a gate length of 1 μm is measured by TLP under different gate biases (V_G).

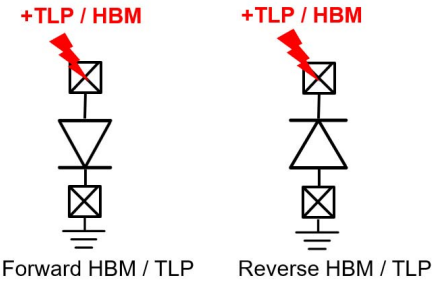


Fig. 5. The measurement setup of the TLP / HBM ESD test on the LFER diodes under (a) forward mode and (b) reverse mode.

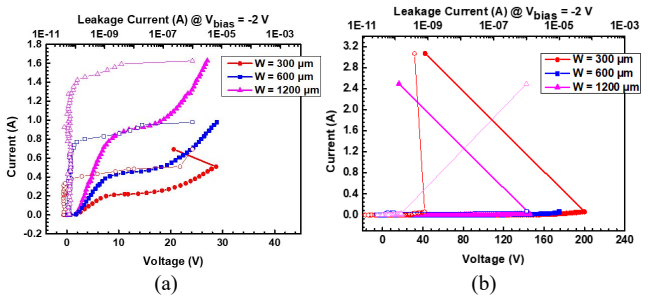


Fig. 6. TLP-measured *I-V* curves of LFER diodes with different dimensions under (a) forward mode and (b) reverse mode.

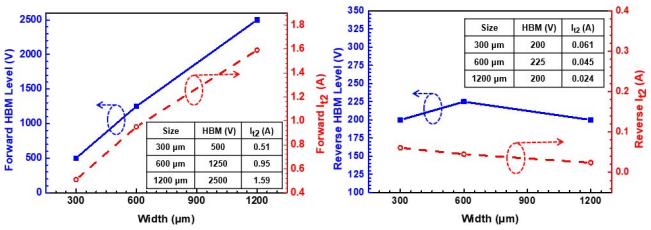


Fig. 7. The relations between HBM ESD level and TLP-measured I_{t2} of GaN diodes with different dimensions under (a) forward mode and (b) reverse mode.

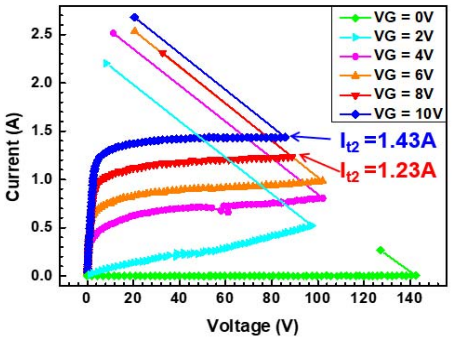


Fig. 8. The TLP-measured *I-V* curves of 12-V E-HEMT GaN device with total width of 2000 μm and gate length of 1 μm under different gate biases (V_G).

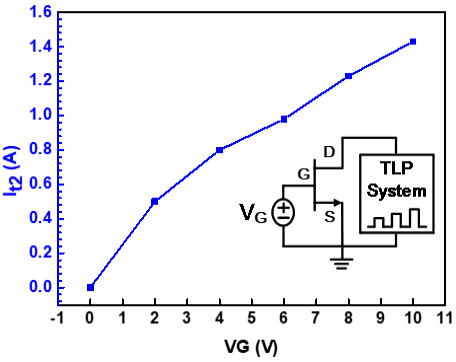


Fig. 9. The relation between V_G and TLP-measured I_{t2} of 12-V E-HEMT GaN device with total width of 2000 μm and gate length of 1 μm.

The TLP-measured *I-V* curves are shown in Fig. 8, and the relation between V_G and I_{t2} of E-HEMT GaN device is shown in Fig. 9. When the gate bias of the E-HEMT device is increasing, the TLP-measured I_{t2} clearly becomes higer. It is evident that a higher gate bias enables the E-HEMT GaN device to withstand high ESD stress. Therefore, how to increase the gate voltage in the gate-coupled circuit is an important task for ESD protection design.

Based on the aforementioned investigation results, the power-rail ESD clamp circuit realized by the gate-coupled design [7] is shown in Fig. 10. The capacitor C_1 and resistor R_1 are used to generate the gate bias to the E-HEMT GaN device during the ESD events. Increasing the capacitance of C_1 in the gate-coupled circuit will enable higher gate bias to further enhance its ESD robustness. The resistance R_1 can keep the coupled voltage to stay at the gate of E-HEMT GaN device longer in time, and therefore improve ESD level. To investigate the gate-coupled design on the power-rail ESD clamp circuit, the split parameters for test devices including the gate-grounded E-HEMT (GG-E-HEMT) and the gate-coupled E-HEMT (GC-E-HEMT) are listed in Table I. The GaN devices (Q_E) under test are drawn with a total width of 2000 μm and 4000 μm, respectively, and a gate length of 1 μm.

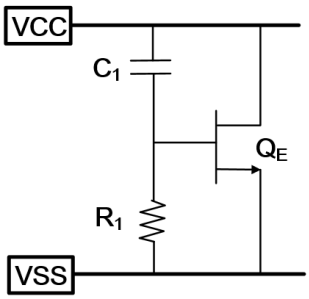


Fig. 10. The power-rail ESD clamp circuit realized by the gate-coupled design.

Table I Summary of power-rail ESD clamp circuits with gate-coupled design realized with different parameters (C_1, R_1, and Q_E *size*).

Device		C_1	R_1	Q_E size
GG-E-HEMT	A0	no	0 Ω	W = 2000 μm, L = 1 μm
GC-E-HEMT	A1	0.4 pF	1.95 MΩ	
	A2	1.6 pF		
	A3	4 pF		
	A4	0.4 pF		W = 4000 μm, L = 1 μm

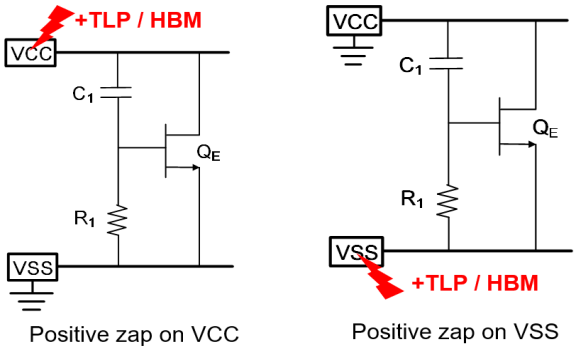


Fig. 11. The measurement setup of TLP / HBM ESD test on the power-rail ESD clamp circuits with gate-coupled design under (a) positive zap on VCC and (b) positive zap on VSS.

Fig. 11 displays the measurement setup on the power-rail ESD clamp circuits with gate-coupled design in the TLP and HBM ESD test. The TLP *I-V* curves under positive zap on VCC and VSS among the test devices are shown in Figs. 12(a) and 12(b), respectively. It can be seen that there is some saturation region along the TLP-measured *I-V* curves under postive zap on VCC. Such a phenomenon is very unique, that was not found in CMOS devices with silicon process.

The HBM ESD levels of the test circuits are measured and listed in Table II with the TLP-measured I_{t2} under two zapping modes. Since the HEMT GaN device lacks the parasitic bipolar junction transistor (BJT) to enter the snapback region (as that of silicon device) for discharging ESD current, the GG-E-HEMT with a device width of 2000 μm exhibits a poor HBM ESD level of only 275V. However, by using the gate-coupled design to generate gate bias during ESD event, the GC-E-HEMT exhibits high ESD robustness. To investigate the relation between the Q_E size and the ESD level with GC-E-HEMT, it can be inferred from the device total widths between A1 and A5. When the device width of Q_E increases from 2000 μm to 4000 μm, the corresponding HBM ESD level improves from 2600V (A1) to 4500V (A5), and the I_{t2} also enhances from 1.24A to 2.65A. As a result, the size of Q_E is almost proportional to its ESD robustness.

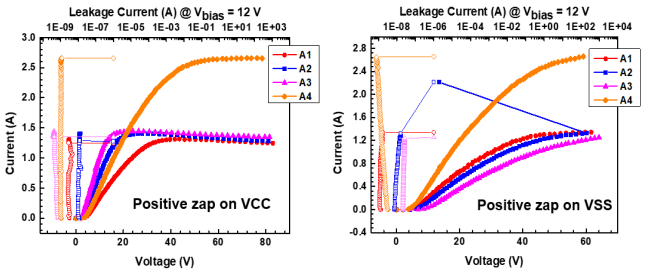


Fig. 12. TLP-measured *I-V* curves of the power-rail ESD clamp circuits with gate-coupled design under (a) positive zap on VCC and (b) positive zap on VSS.

Table II Summary of HBM ESD levels and TLP-measured I_{t2} among the power-rail ESD clamp circuits with gate-coupled design.

Device	Positive zap on VCC		Positive zap on VSS	
	HBM (V)	I_{t2} (A)	HBM (V)	I_{t2} (A)
A0	275	0.001A	3200	1.51
A1	2600	1.24	2300	1.34
A2	2700	1.27	2300	1.32
A3	2900	1.34	2200	1.25
A4	4500	2.65	4300	2.65

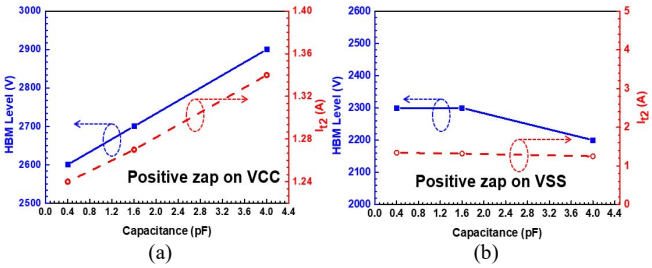


Fig. 13. The dependences of HBM ESD levels and TLP-measured I_{t2} on the gate-coupled capacitance in the power-rail ESD clamp circuits under (a) positive zap on VCC and (b) positive zap on VSS.

The impact of the capacitance on the gate-coupled circuits can be examined from the test results among the A1, A2, and A3 devices. The dependences of HBM ESD levels and TLP-measured I_{t2} on the gate-coupled capacitance among the power-rail ESD clamp circuits are clearly illustrated in Figs. 13(a) and 13(b), respectively. Larger capacitance is observed to enhance ESD robustness of the power-rail ESD clamp circuits with gate-coupled design under postive zap on VCC. However, it was almost with no influence to ESD level of the test circuits under positive zap on VSS. The reason to cause such a difference is discussed in the following section.

IV. Discussion

In Fig. 12(a), with positive zap on VCC, the current under TLP measurement was almost saturated. Such a phenomenon differs from that of the gate-coupled circuit in the silicon device, because the E-HEMT lacks the oxide layer at its gate structure. The gate leakage in the metal-semiconductor (MS) structure of GaN [8] is large. As shown in Fig. 14, the DC I_G-V_G curve of the HEMT device indicates that the gate current is almost up to 1 μA with a bias of 6V. When the bias on the gate exceeds the operating voltage (6V), the leakage current will become extremely high. Thus, the coupled voltage on the gate of Q_E in the power-rail ESD clamp circuits under positive zap on VCC was released by the gate leakage current. Comparing the I_{t2} shown in Fig. 8 and Table II, it can be estimated that the power-rail ESD clamp circuits with gate-coupled design only keep the coupled gate voltage of ~8V.

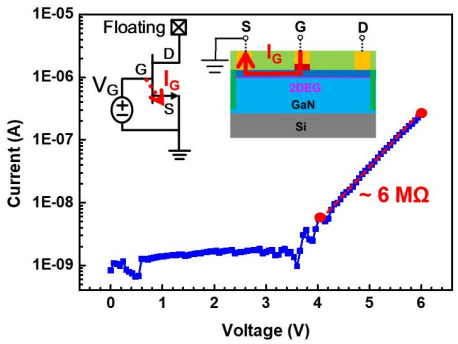


Fig. 14. The DC I_G-V_G curve of a GaN HEMT with the total gate width of 2000 μm and the gate length of 1 μm.

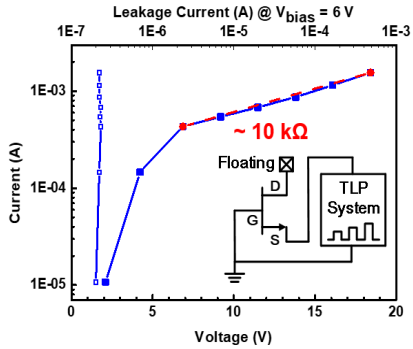


Fig. 15. TLP-measured *I-V* curve of E-HEMT (2000 μm) swept from source to gate with drain floating.

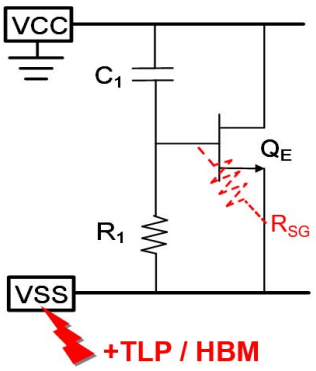


Fig. 16. The schematic of the gate-coupled circuit with an equivalent resistance (R_{SG}) between the gate and the source of Q_E under the ESD test mode with positive zap on VSS.

For the gate-coupled circuit in the CMOS silicon process, the ESD current can be discharged through the body diode of MOSFET under the ESD test mode with positive zap on VSS. In Table II, the E-HEMT lacks the parasitic BJT and body diode, but it still has somewhat good HBM ESD level and I_{t2} under positive zap on VSS. To clarify the issue, the TLP-measured *I-V* curve from the source to the gate of a stand-alone E-HEMT device is shown in Fig. 15. The measurement setup is also drawn in the inset of Fig. 15. It can be seen that the measured current from source to gate is large, and the corresponding equivalent resistance (R_{SG}) is approximately 10 kΩ, which was very small compared to R_1 (1.95 MΩ).

The schematic of gate-coupled circuit with the additional equivalent resistance (R_{SG}) under the ESD test mode of positive zap on VSS is shown in Fig. 16. The total resistance (R_{SG}‖R_1) is dominated by the value of R_{SG}. With a shorter RC time constant, the gate of Q_E in the gate-coupled circuit under the ESD test mode with positive zap on VSS will get the voltage through R_{SG} to turn on the channel of Q_E for releasing ESD current. This is the reason why the gate-coupled circuit also has good ESD robustness under the ESD test mode with positive zap on VSS, even if the Q_E device is without the body diode.

V. Conclusion

The ESD robustness of GaN diode (LFER) is almost proportional to its total gate width under forward mode, but it has very low ESD level under reverse mode. Due to this characteristic of unidirectional conduction in the GaN diodes, the power-rail ESD clamp circuit with gate-coupled design is used to establish the whole-chip ESD protection for the GaN IC. Under the ESD test mode with positive zap on VCC, ESD robustness of the power-rail ESD clamp circuit is influenced by the dimension of Q_E and the capacitance in the gate-coupled circuit. Increasing the capacitance and the size of Q_E can improve HBM ESD level and TLP-measured I_{t2} of the power-rail ESD clamp circuit. However, the gate leakage current through the MS structure in GaN HEMT reduces the impact of the capacitance in the gate-coupled circuit. Even if there is no body diode inside Q_E device, the reason to cause gate-coupled circuit with good enough ESD robustness under the ESD test mode with positive zap on VSS has been well explained. Through the test structures investigated in this work, good ESD protection on GaN IC can be well achieved.

Acknowledgment

This work was supported in part by the National Science and Technology Council (NSTC), Taiwan, under Contract of NSTC 112-2221-E-A49-147-MY3. The authors would like to thank *Taiwan Semiconductor Research Institute* (TSRI) for the supports with EDA tools and the chip fabrication through TSMC.

References

[1] Y.-Y. Kao *et al.*, "Fully integrated GaN-on-Silicon gate driver and GaN switch with temperature-compensated fast turn-on technique for achieving switching frequency of 50 MHz and slew rate of 118.3 V/Ns," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 12, pp. 3619-3627, Dec. 2021.

[2] S.-Y. Lin *et al.*, "Monolithic GaN-Based gate driver with on-chip adaptive on-time controller and negative current slope detector to prevent shoot-through," *IEEE Solid-State Circuits Letters*, vol. 6, pp. 217-220, 2023.

[3] J.-H. Lee *et al.*, "Incorporation of a simple ESD circuit in a 650V E-mode GaN HEMT for all-terminal ESD protection," in *Proc. IEEE Int. Reliability Physics Symp.*, 2022, pp. 2B.3-1-2B.3-6.

[4] C.-Y. Ke, W.-C. Wang, M.-D. Ker, C.-Y. Yang, and E.-Y. Chang, "Investigation on ESD robustness of 1200-V D-mode GaN MIS-HEMTs with HBM ESD test and TLP measurement," in *Proc. Int. VLSI Symp. on Technology, Systems and Applications*, 2023, pp. 1-2.

[5] *TSMC 0.50 μm GaN WBG E-HEMT USG AL 0P3M HKMG 650V Design Rule.* Jun. 2019.

[6] W. Chen, K.-Y. Wong, and K. J. Chen, "Single-chip boost converter using monolithically integrated AlGaN/GaN lateral field-effect rectifier and normally off HEMT," *IEEE Electron Device Lett.*, vol. 30, no. 5, pp. 430-432, May. 2009.

[7] M.-D. Ker, C.-Y. Wu, T. Cheng, and H.-H. Chang, "Capacitor-couple ESD protection circuit for deep-submicron low-voltage CMOS ASIC," *IEEE Trans. VLSI Syst.*, vol. 4, no. 3, pp. 307-321, Sept. 1996.

[8] S.-W. Tang *et al.*, "Using gate leakage conduction to understand positive gate bias induced threshold voltage shift in p-GaN gate HEMTs," *IEEE Trans. Electron Devices*, vol. 70, no. 2, pp. 449-453, Feb. 2023.

Method for Suppressing Trap-Related Memory Effects in IV Characterizations of GaN HEMTs

Johan Bremer[1], Niklas Rorsman[1] and Mattias Thorsell[1]
[1]Microwave Electronics Laboratory
Chalmers University of Technology
MC2, 412 96 Gothenburg Sweden
Email: bremerj@chalmers.se

***Abstract*—A method to suppress trapping-related effects when performing IV characterizations of field effect transistors is presented. Standard IV measurements usually utilize voltage sweeps with fixed start, stop, and step values. At high electric fields in these sweeps, the charging of electron traps with long time constants may occur. The trapped electrons cause different memory effects such as hysteresis and the kink effect. The proposed method suppresses these effects, by reordering the bias points, so to prevent charging due to high preceding electric fields. The method provides more rudimentary IV measurements, useful for e.g. technology evaluation and modeling purposes.**

I. Introduction

IV measurements are used in the evaluation of transistor technologies, to supply data for process monitoring, as well as for component modeling. To allow this, the measurements must be repeatable and independent of instrumentation and operator (i.e. laboratory independent). However, the IV characterization of gallium nitride (GaN) high electron mobility transistors (HEMTs) often suffer from electron-trap-related effects such as current slump and the kink effect [1]–[4]. The dominant cause of the activation of trap states is the high electric field at the gate contact at high drain-gate voltages. The electric field causes electrons to occupy trap states in e.g. the buffer, barrier, passivation, and their interfaces. It has been shown that the trap states may exhibit emission time constants, ranging from ns to 100s of seconds [5], [6]. The states with long time constants create long-lasting memory effects due to the virtual gate and back-gating effect [7], [8]. Consequently, the shape of the entire IV characteristic becomes dependent on measurement settings such as the maximum voltage and measurement sweep time.

This paper proposes a new method to perform the I_{ds}-V_{ds} characterization, which aims to minimize the memory effects caused by electron trapping, resulting in the most rudimentary IV characteristic. This is achieved by sorting the bias points so that the electric field in the gate-drain region increases throughout the IV sweep. The method is evaluated on two commercially available GaN HEMT technologies. The proposed method suppresses the kink effect and measures the knee region with minimal impact of trap-related memory effects. The rudimentary IV data provided by the method could be useful in the extraction of a non-linear model of the HEMT.

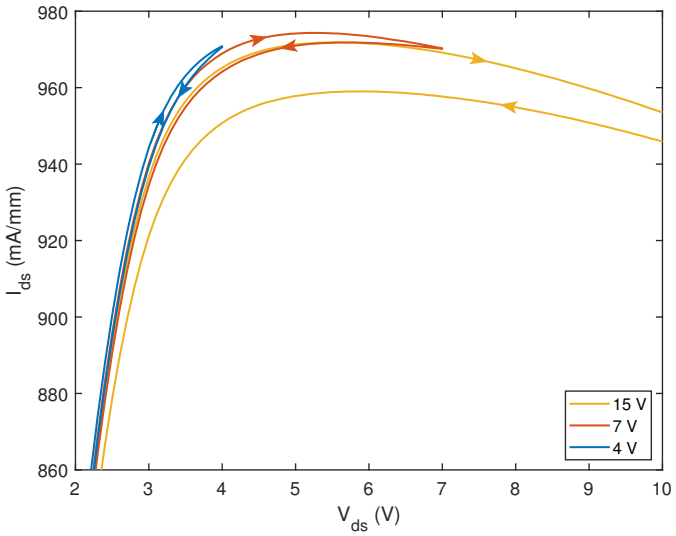


Fig. 1. Consecutive bi-directional drain-current versus drain-voltage IV sweeps, with increasing maximum drain voltage (4 to 15 V). The sweeps start at 0 V and the gate voltage is 1 V. The arrowheads indicate the direction of the sweeps. The drain voltage axis is limited to 10 V to see the hysteresis for all voltage sweeps.

II. Proposed IV Characterization Method

A. Memory Effects in IV Measurements

The memory effect in the device can be demonstrated by performing bi-directional drain-current versus drain-voltage (I_{ds}-V_{ds}) sweeps. In Fig. 1, three consecutive sweeps are performed with increasing maximum drain voltage ($V_{ds,max}$). The hysteresis (loop openings) are a distinct sign of a memory effect. Hysteresis can be observed already for $V_{ds,max}$ = 4 V, and is significant for $V_{ds,max}$ = 15 V. The forward sweep for $V_{ds,max}$ = 15 V overlaps with the reverse sweep of $V_{ds,max}$ = 7 V showing that the 15 V measurement starts with the memory state left by the previous (7 V) sweep.

With higher voltage sweeps, the kink effect is often observed in the I_{ds}-V_{ds} characteristics of HEMTs [4]. It is a hysteretic instability of the drain-source current in field-effect transistors, observed during a slow drain bias sweep. The effect can be seen between 5 and 15 V in Fig. 2, where several I_{ds}-V_{ds} characteristics of a GaN HEMT are shown. In these characteristics, the maximum drain-source voltage is varied to 5, 15, or 30 V. The kink effect becomes apparent when the maximum voltage of the measurement is increased.

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

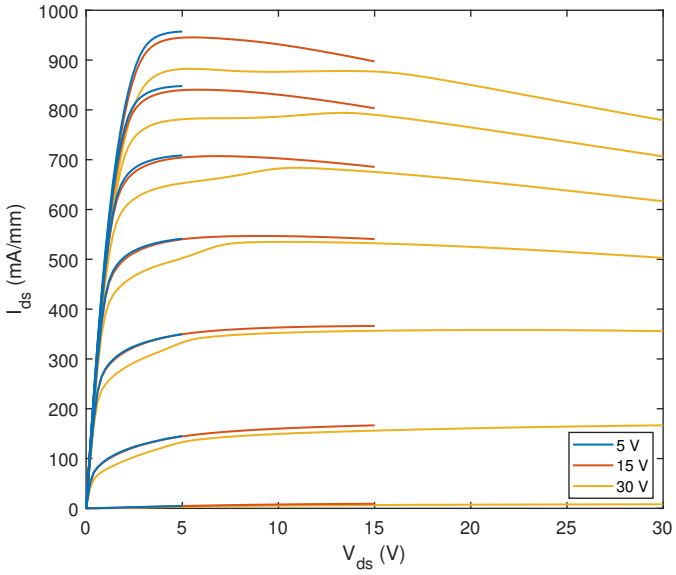


Fig. 2. I_{ds}-V_{ds} characteristics with increasing maximum drain voltage (5 to 30 V). The gate voltage is swept from -2 to 1 V with 0.5 V steps.

B. Methodology to Suppress Memory Effects

Normally, an IV characterization of field-effect transistors is defined by selecting the start-, stop-, and step-voltage of the gate and drain voltage sources. For an I_{ds}-V_{ds} characterization, V_{ds} is swept from a low-to-high value while the gate-source voltage (V_{gs}) is kept constant. The V_{ds} sweep is then repeated for different V_{gs} voltages, which are either increasing or decreasing. The characterization with increasing V_{gs} was demonstrated in Fig. 2 above, demonstrating that this standard measurement may provide a trapping-impaired I_{ds}-V_{ds} characteristic that depends on $V_{ds,max}$. We propose that such dependency can be mitigated by rearranging the gate and drain bias points so that the drain-gate electric field is strictly increasing throughout the measurement. Therefore, a method to quantify the electric field is needed. However, the field distribution is complex and requires physical modeling and simulations to quantify. By noting that the peak drain-gate electric field and extension of the field into the buffer generally increases when V_{ds}/V_{gs} increases/decreases [9], [10], we propose that the drain-gate voltage ($V_{dg} = V_{ds} - V_{gs}$) can be used to make a first-order approximation of the strength of the electric field.

The ordering of the bias points is performed by first calculating V_{dg}, and then sorting so that V_{dg} strictly increases. Fig. 3 shows the sweeps of the V_{ds} (Fig. 3a), V_{gs} (Fig. 3b), and the resulting V_{dg} (Fig. 3c) for the proposed method as well as for the standard measurements with increasing and decreasing V_{gs}. For both the standard measurements, V_{dg} repeatably changes from a low to a high value in a sawtooth shape (Fig. 3c). Therefore, most measurement points are affected by high preceding electric fields. In contrast, the proposed method results in a strictly increasing V_{dg}.

Multiple bias points may yield the same V_{dg}. In such cases, the bias points are sorted so that V_{gs} increases. In this way, the HEMT tends to change from a colder to a warmer state for each V_{dg}. Furthermore, the measurement points at e.g. $V_{gs,max}$ will be preceded by possibly several points with the same V_{dg}.

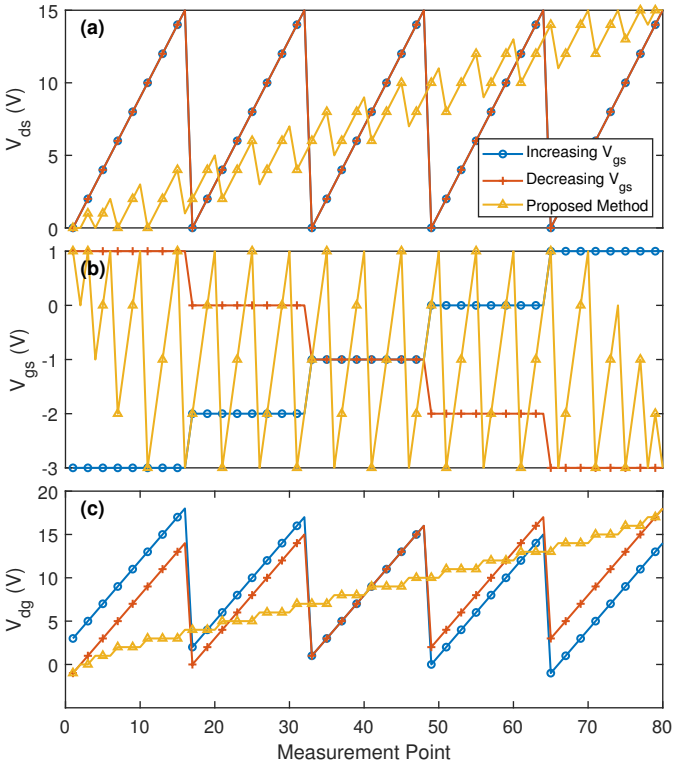


Fig. 3. Measurement lists of an I_{ds}-V_{ds} characterization (0.5 V gate voltage step, 1 V drain voltage step). (a) and (b) show the drain and gate voltages, respectively, and (c) shows the corresponding drain-gate voltage. Red and blue show bias lists for standard I_{ds}-V_{ds} characterizations, where V_{ds} is the primary increasing sweep (a), and V_{gs} is the increasing/decreasing secondary sweep (b). Yellow in (c) shows the proposed rearranged list where V_{dg} is strictly increasing throughout the measurement.

This extends the time for e.g. trap-filling for measurement points at high V_{gs}.

III. Method Evaluation

A. Devices and Measurement Setup

The characterized HEMTs are fabricated with commercially available monolithic microwave integrated circuit AlGaN/GaN on silicon carbide technologies. The first device is a 2x40 µm HEMT with 0.15 µm gate length (denoted A). The second device is a 1x40 µm HEMT with 0.25 µm gate length (denoted B).

The HEMTs were characterized using a Keysight B2912B source measurement unit at room temperature with the microscope lamp switched off. The kink effect is affected by the timing settings of the SMU [4], and hence delay times ranging from 1 to 200 ms were evaluated with an integration time fixed to 2 ms. To ensure that the characterization is performed on HEMTs with a similar initial memory state [2], a recovery time of 24 h was used between measurements.

B. I_{ds}-V_{ds} Characterization

HEMTs A and B were characterized by the proposed as well as the standard methods by using the settings in Table I, and the results are shown in Fig. 4. The increasing and decreasing V_{gs} methods generally exhibit a lower current in the knee region. The kink is seen between 5 and 15 V for HEMT A and at around 5 V for HEMT B. At $V_{gs,max}$,

TABLE I
I_{ds}-V_{ds} CHARACTERIZATION SETTINGS FOR HEMT A AND B.

$V_{gs,max}$ (V)	$V_{gs,min}$ (V)		$V_{gs,step}$ (V)	$V_{ds,max}$ (V)	$V_{ds,min}$ (V)	$V_{ds,step}$ (V)
	A	B				
1			0.5	30	0	0.2
	-2	-2.5				

TABLE II
KEY FIGURES OF MERIT OF I_{ds}-V_{ds} CHARACTERISTICS. R_{on}, V_{knee}, AND I_{knee} ARE CALCULATED AT $V_{gs} = V_{gs,max} - V_{gs,step}$ (0.5 V).

Method	Increasing V_{gs}		Decreasing V_{gs}		Proposed	
HEMT	A	B	A	B	A	B
$I_{ds,max}$ (mA/mm)	885	794	973	799	956	809
R_{on} (Ωmm)	2.48	2.81	2.52	2.79	2.33	2.68
V_{knee} (V)	2.77	2.79	2.81	2.77	2.71	2.65
I_{knee} (mA/mm)	765	664	761	666	824	706

the increasing V_{gs} method shows the largest deterioration due to the preceding sweeps starting at $V_{gs,min}$. On the contrary, the decreasing V_{gs} method shows the smallest deterioration at $V_{gs,max}$, since no preceding sweeps have taken place. The IV characteristics of the proposed method (Fig. 4a and Fig. 4b) show no obvious kink effect. Overall, the IV characteristic is improved, indicating electron trapping effects in the other approaches. However, an exception can be seen at $V_{gs,max}$ for HEMT A. This is partly due to variations between devices, but possibly also due to the measurement order of biased points with the same V_{dg} (as discussed in section II-B).

A comparison of key figures of merit of the different methods for both HEMTs is shown in Table II. The general trends are reduced on-resistance (R_{on}), higher knee current (I_{knee}), and lower knee voltage (V_{knee}) for the proposed method compared to the more standard methods. The proposed method suppresses the kink effect by minimizing the trap-related memory effects on the I_{ds}-V_{ds} characteristic . Memory effects appear to have a larger impact on the IV characteristic of HEMT A compared to HEMT B. Therefore, the improvements with the proposed method are bigger in the case of HEMT A.

C. Timing Sensitivity

The observed deterioration of the IV characteristic depends on the time constants of the traps and is therefore impacted by the time settings of the SMU. The sensitivity of the timing on the measurement was investigated by varying the delay time (t_d) between the bias output and the measurement start (t_d = 1, 100, and 200 ms). Fig. 5 shows the I_{ds}-V_{ds} characteristics of HEMT A using the proposed method (Fig. 5a) and the decreasing V_{gs} method (Fig. 5b). Evidently, the proposed method shows a smaller dependence on t_d compared to the decreasing V_{gs} method.

However, a fast sweep means a short trap filling-time at high voltages, as well as a de-trapping impacting a larger V_{ds} range at the next V_{gs} setting. Therefore, a fast sweep can result in the

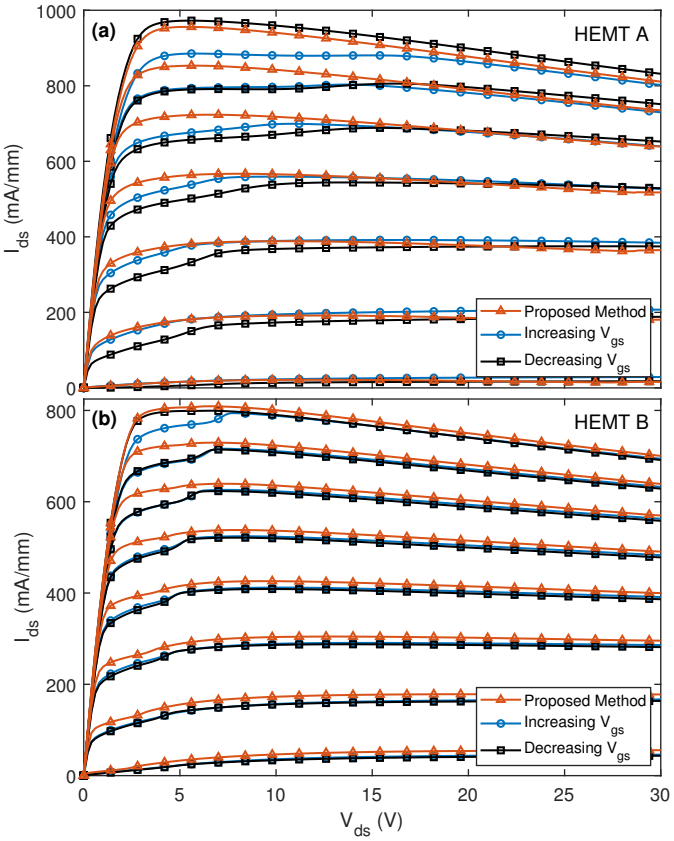


Fig. 4. HEMT I_{ds}-V_{ds} characteristics measured with the proposed method (red), increasing V_{gs} method (blue), and decreasing V_{gs} method (black). (a) shows HEMT A and (b) shows HEMT B. V_{gs} is swept from -2 to 1 V for HEMT A and -2.5 to 1 V for HEMT B. The V_{gs} step is 0.5 V and the V_{ds} step is 0.2 V.

appearance of a kink-free characteristic, as seen for t_d = 1 ms for the decreasing V_{gs} method. The current is however lower also in this case compared to the proposed method. This is indicated by the larger current difference (at the knee) between $V_{gs,max}$ and $V_{gs,max} - V_{gs,step}$ in Fig. 5b compared to Fig. 5a.

Due to the self-heating, a current reduction, which increases when t_d and/or V_{gs} increase, can be expected. This can be seen for the proposed method but not consistently for the decreasing V_{gs} method in Fig 5b. This implies that the proposed method likely provides better information on the impact and time dependence of the self-heating in the HEMT.

IV. DISCUSSION

The V_{gs} sweep of the I_{ds}-V_{ds} characterization is often swept from below to above the pinch-off voltage (V_{po}). The electric field distribution changes abruptly when V_{gs} is lower than V_{po}. In such cases, a high V_{ds} creates an off-state stress, where e.g. trap states deep in the buffer could be activated. Clearly, this is not considered by the calculation of V_{dg}. For the decreasing V_{gs} method, measurements below V_{po} do not affect the on-state characteristic. However, for the proposed- and increasing V_{gs}-methods, these off-state points could impact the entire characteristic. The dependency of the kink effect on V_{ds} and V_{gs} was studied in [2], and their results show that the kink is nearly negligible when the HEMT is stressed at an off-state. This is also in line with our tests and it indicates that $V_{gs,min}$ can be selected below V_{po} without causing any kink-related

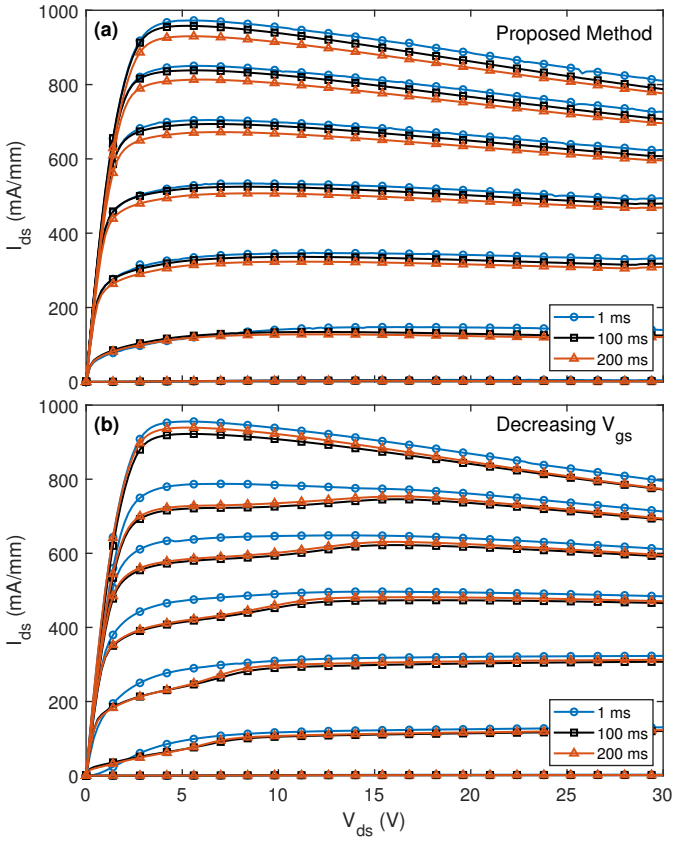


Fig. 5. I_{ds}-V_{ds} characteristics of HEMT A with different delay times. The IV measurement is performed with the proposed method in (a) and the decreasing V_{gs} method in (b).

trapping. In our case, $V_{gs,min}$ is set close to the pinch-off voltages of HEMT A (-2 V) and HEMT B (-2.8 V).

V. Conclusion

A new technique to perform I_{ds}-V_{ds} characterizations of field effect transistors is proposed. The evaluation of the method showed little to no kink in the knee region. With the proposed method, the same rudimentary IV properties of the knee region are measured regardless of how high V_{ds} is swept. For low-voltage sweeps, the impact of the kink effect is reduced, and hence the effectiveness of the method diminishes. However, the bi-directional sweeps showed hysteresis (memory effect) even with relatively low voltage sweeps. This implies that the method has a limited potential to also improve characterizations at lower voltage.

In addition, the suppression of trap-related memory effects results in an IV characteristic, which is less sensitive to the measurement time settings. Measurement results are therefore easier to recreate since the time settings can be selected more freely. It was seen that one way to reduce the kink in the knee region is to sweep fast. The effectiveness of such a strategy is however limited and unreliable since it depends on the trap characteristics of the HEMT.

GaN has a large bandgap, which allows for trap-states with long emission time constants. This makes the proposed method particularly useful for GaN-based devices. However, other technologies and devices could also benefit from the proposed method. In general, the method could be useful for any device whose IV characteristic deteriorates when subjected to high electric fields. Logic devices in CMOS technologies also suffer from kink effects, where impact ionization plays a central role [11]. In this case, the kink is exacerbated by divots in shallow isolation trenches for MOSFETs. Since impact ionization is strongly connected to the electric field, the proposed method could also be useful for these devices.

Acknowledgment

This work was performed within the Competence Center for III-Nitride Technology C3Nit-Janzén and has received financial support from the Swedish Governmental Agency for Innovation Systems (VINNOVA) via grant No. 2016-05190, Linköping University, Chalmers University of Technology, ABB, Epiluvac, Ericsson, FMV, Gotmic, Hexagem, On Semiconductor, Saab, SweGaN, United Monolithic Semiconductors (UMS), and Volvo Cars.

We would like to acknowledge WIN Semiconductors Corp. for providing access to their NP15 process and manufacturing test chips as a part of WIN's university multiproject wafer program.

References

[1] R. Ye, X. Cai, C. Du, H. Liu, Y. Zhang, X. Duan, and J. Zhu, "An Overview on Analyses and Suppression Methods of Trapping Effects in AlGaN/GaN HEMTs," *IEEE Access*, vol. 10, pp. 21 759–21 773, Dec. 2022, doi: 10.1109/ACCESS.2021.3139443.

[2] M. Wang and K. J. Chen, "Kink Effect in AlGaN/GaN HEMTs Induced by Drain and Gate Pumping," *IEEE Electron Device Letters*, vol. 32, no. 4, pp. 482–484, April 2011, doi: 10.1109/LED.2011.2105460.

[3] G. Meneghesso, F. Zanon, M. J. Uren, and E. Zanoni, "Anomalous Kink Effect in GaN High Electron Mobility Transistors," *IEEE Electron Device Letters*, vol. 30, no. 2, pp. 100–102, Feb. 2009, doi: 10.1109/LED.2008.2010067.

[4] M. Singh, M. J. Uren, T. Martin, S. Karboyan, H. Chandrasekar, and M. Kuball, ""Kink" in AlGaN/GaN-HEMTs: Floating Buffer Model," *IEEE Transactions on Electron Devices*, vol. 65, no. 9, pp. 3746–3753, Sept. 2018, doi: 10.1109/TED.2018.2860902.

[5] J. Joh and J. A. del Alamo, "A Current-Transient Methodology for Trap Analysis for GaN High Electron Mobility Transistors," *IEEE Transactions on Electron Devices*, vol. 58, no. 1, pp. 132–140, Jan. 2011, doi: 10.1109/TED.2010.2087339.

[6] D. Bisi, M. Meneghini, C. de Santi, A. Chini, M. Dammann, P. Brückner, M. Mikulla, G. Meneghesso, and E. Zanoni, "Deep-Level Characterization in GaN HEMTs-Part I: Advantages and Limitations of Drain Current Transient Measurements," *IEEE Transactions on Electron Devices*, vol. 60, no. 10, pp. 3166–3175, Oct. 2013, doi: 10.1109/TED.2013.2279021.

[7] J. Pedro, P. Tomé, T. Cunha, F. Barradas, L. Nunes, P. Cabral, and J. Gomes, "A Review of Memory Effects in AlGaN/GaN HEMT Based RF PAs," in *IEEE MTT-S International Wireless Symposium*, May 23-26, 2021, pp. 1–3, doi: 10.1109/IWS52775.2021.9499373.

[8] R. Quay, *Gallium Nitride Electronics*. Germany: Springer, Berlin, Heidelberg, 2008.

[9] J. Si, J. Wei, W. Chen, and B. Zhang, "Electric Field Distribution Around Drain-Side Gate Edge in AlGaN/GaN HEMTs: Analytical Approach," *IEEE Transactions on Electron Devices*, vol. 60, no. 10, pp. 3223–3229, Oct. 2013, doi: 10.1109/TED.2013.2272055.

[10] A. Bag, S. Das, P. Mukhopadhyay, and D. Biswas, "Observation and analysis of kink effect during drain current inception of GaN HEMT," *Superlattices and Microstructures*, vol. 120, pp. 101–107, Aug. 2018, doi: 10.1016/j.spmi.2018.04.035.

[11] H. Park, M. Bawedin, H. Choi, and S. Cristoloveanu, "Kink effect in ultrathin FDSOI MOSFETs," *Solid-State Electronics*, vol. 143, pp. 33–40, May 2018, doi: 10.1016/j.sse.2017.12.002.

A Neural Network-based Manufacturing Variability Modeling of GaN HEMTs

Fredo Chavez[1a], Danial Bavi[2a], Nicholas C. Miller[3b] and Sourabh Khandelwal[4a]
[a]School of Engineering, Macquarie University, Australia
[b]Department of Electrical and Computer Engineering, Michigan State University, USA
[1]fredo.chavez@mq.edu.au, [2]danial.bavi@hdr.mq.edu.au, [3]ncmiller@msu.edu, [4]sourabh.khandelwal@mq.edu.au

***Abstract*—A new technique to accurately model the manufacturing variability of GaN HEMT using a neural network(NN) is presented in this paper. Compact model parameters are automatically generated through Principal component analysis (PCA) parameters from variations in I-V data. Together with the bias conditions, the compact model parameters are used to train a neural network. The NN-based compact model captures the I-V behavior of 115 GaN HEMT with excellent accuracy. The trained neural network is converted to a standard Verilog-A file that can be imported to a circuit simulator. The NN-based compact model is further evaluated in terms of complexity and simulation speed. The presented technique shows great potential in developing a fast, flexible, and accurate NN-based compact model that can be applied to any device technology.**

***Index Terms*—Neural Networks (NN), GaN HEMT, compact model, Parameter Generation**

I. Introduction

The recent development of Gallium nitride-based high electron mobility transistors(GaN HEMTs) has emerged as the prevalent technology for high-frequency, high-power applications [1], [2]. As more circuits and products designed in GaN HEMT are being fabricated, modeling the variability in electrical performance due to manufacturing variations becomes a necessity. An accurate model for manufacturing variations allows sensitivity analysis of process variations and ensures that circuit designs meet design specifications. Modeling the manufacturing variations has been demonstrated on the industry-standard physics-based compact model, ASM-HEMT, by creating a statistical model [3] or by automatically extracting model parameters for every device [4].

Neural network(NN)-based compact models have gained popularity in previous years as an alternative to physics-based compact models because of their excellent fitting capabilities and fast development time [5]–[14]. Previous research has focused on creating an NN-based compact model based on its bias (gate and drain voltage) and its instance parameters, such as gate length, width, and temperature. Multiple transistors manufactured in the same process have variability in their geometrical and physical parameters that cause significant variations in the device's characteristics. The presented techniques do not allow the flexibility to model the manufacturing variations in these device's characteristics.

This paper proposes a framework for creating a flexible neural network-based compact model. The created neural network-based compact model can model the manufacturing variability of GaN HEMT devices. The rest of the paper is organized as follows: Section II shows the compact model development, which includes the generation of the dataset, parameter generation, and the training of the neural network-based compact model. Section III analyzes the trained neural network and evaluates its fitting accuracy and simulation speed. Finally, concluding remarks are presented in Section IV.

II. Model Development

A. Dataset Generation

DC IV measurements have been performed on 115 GaN HEMT devices from three wafers at room temperature. The measurements are taken on a standard ground-signal-ground(GSG) transistor test structure. These measurements cover all of the regions of device operation. This includes gate voltage V_G from -4V to 1V, with a step size of 0.1V, and drain voltage V_D from 0 to 15V, with a step size of 0.2V. Fig. 1 shows the variation in the I-V characteristics. These variations in I-V characteristics reflect the inconsistencies in the device fabrication process. The device fabrication induces variability in the geometrical parameters(gate-length, gate-width, barrier thickness, source-gate access region length, and drain-gate access region length) and physical parameters(gate work-function, barrier height, mobility, access regions source and drain resistances, and saturation velocity) [3]. The DC IV measurements are split into 80% training (92 devices), and 20% validation (23 devices). The validation dataset is not used to train the neural-network, but only used for evaluation purposes. Overall, the NN-based compact model is trained and validated on 356,592 and 89,148 data points respectively. The workflow of the development of the NN-based compact model for these devices is shown in Fig. 2.

B. Parameter Generation

NNs offer high accuracy and fast development of compact models [5]–[15]. However, previous techniques do not allow the flexibility to alter the I-V characteristics to model the manufacturing variations in devices because the NN's input is limited to its bias (V_G, V_D), geometry (L, W), and environmental conditions (temperature). To model the behavior of multiple devices, we have devised a methodology to automatically generate a unique set of new parameters P_n for every device using *principal component analysis* (PCA) from the I_D sweep data as shown in Fig. 2(a). PCA is a linear dimensionality reduction technique using Singular Value Decomposition to project data to a lower dimensional space [16], [17]. PCA has been widely used for data reduction, visualization and to speed up the training process of machine learning algorithms. In this proposed technique, the components generated through PCA

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

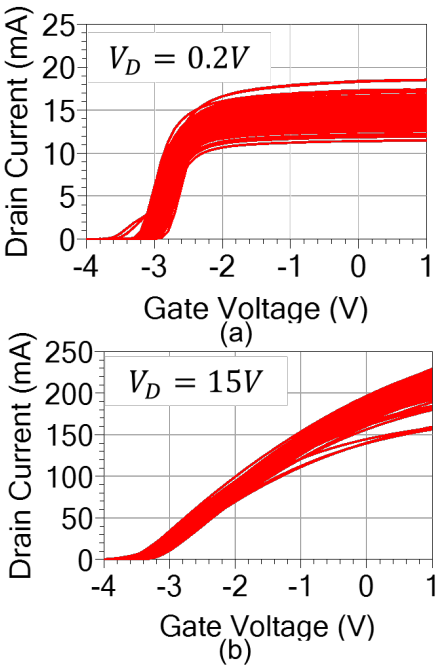


Fig. 1. The transfer characteristics of the measured 115 GaN HEMTs, which shows the process variations in (a) linear ($V_D = 0.2V$) and (b) saturation($V_D = 15V$) condition. The GaN HEMT has four fingers and single finger width of 50 μm i.e (4x50 μm).

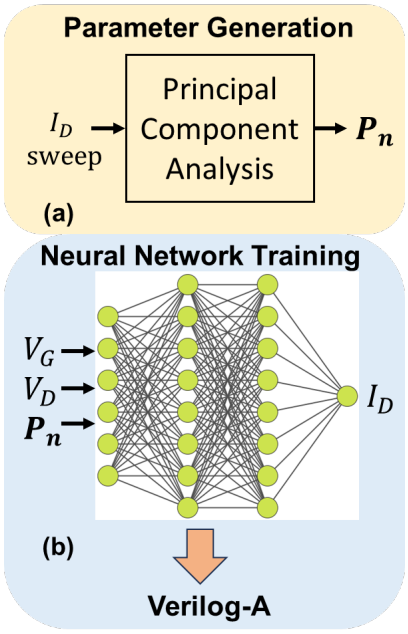


Fig. 2. The workflow in creating the NN-based compact model. First, new parameters P_n are automatically generated in (a), then the neural network is trained using the bias conditions and the P_n in (b). Finally, the trained NN is converted into a standard Verilog-A file that can be imported to a circuit simulator.

will be the new compact model parameters P_n, where n is the number of components.

To create the PCA-derived parameters P_n, the training I_D measurements are grouped into an array of 92x3876, which corresponds to the number of devices (92), and V_G and V_D bias combinations (51x76=3876). The I_D measurements are scaled using min-max normalization and then used to fit the PCA model. PCA is implemented in scikit-learn [18], a Python-based open-source machine learning library. By using this technique, we have observed that PCA requires $n = 3$ components to explain 98% of the variance in I_D, and $n = 4$ components to explain 99% of the variance in I_D as shown in Fig. 3. We have chosen to proceed with $n = 4$ components to enable better differentiation for all devices. This corresponds to parameters P_n, where $P_n = (P_0, P_1, P_3, P_4)$. The fitted PCA model is used to extract the parameters P_n for every device in the training and validation devices.

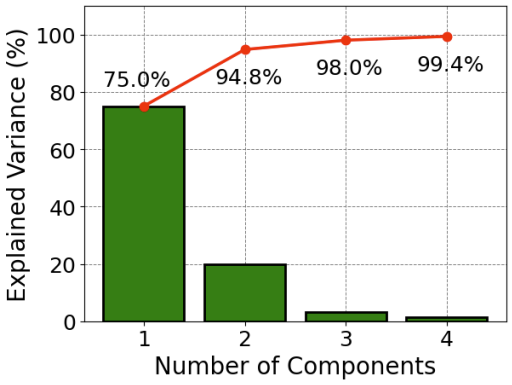


Fig. 3. A Pareto chart of the percentage of explained variance against the number of components created in PCA.

C. Neural Network-based Compact Model

The NN is depicted in Fig. 2 (b). The neural network is trained to extract the relationship between I_D with the bias conditions V_G and V_D and the PCA-derived parameters P_n, expressed as,

$$I_D = F_{NN}(V_G, V_D, P_n) \tag{1}$$

where F_{NN} is the neural network formulation for the I_D.

The NN is created and trained in an open-source machine learning library, Keras and TensorFlow [19]. The NN is trained for 92 devices and validated with 23 devices. Since each measured device has 3876 bias combinations, the training data has over 356k training and 89k validation samples. The number of hidden layers (2,3,4), neurons (16,32,64) and activation function(Tanh, ReLU) of the NN are varied and studied upon deployment to a circuit simulator. The output layer uses the linear activation function. The NN is created and trained in Tensorflow [19]. During training, the NN minimizes the *mean-squared error* (MSE) between the measured and the predicted I_D. The learning rate is held fixed at $1e^{-3}$ and the NN is trained until 30 epochs.

After the NN has been trained, the weights and biases are obtained to generate a standard Verilog-A file that can be imported to a circuit simulator.

III. Results and Discussion

Fig. 4 shows a comparison of Tanh and ReLU activation functions when used to train the NN. Tanh has been used in NN-based compact models [5], [8], [11], [13] because it can provide smooth I_D function, but ReLU has increased in popularity in different applications because it offers faster convergence and solves the vanishing gradient problem [20], [21]. The minimum recorded MSE for the validation dataset in 30 epochs for different hidden layers, neurons, and activation functions is shown, as well as the time to simulate 377 thousand points. The difference in accuracy is small, but we can observe slightly better results in using ReLU when the structure becomes more complex. Similarly, the Verilog-A file generated using ReLU offers faster simulation time because it can solve convergence issues faster, unlike Tanh which has the possibility of having convergence issues at deeper and larger structures. The convergence issues can be avoided by training the NN with a finer step size.

Number of hidden layers	Neurons per hidden layer	Validation MSE (mA)		Simulation time (s)	
		Tanh	ReLU	Tanh	ReLU
2	16	0.551	1.611	24.52	23.79
	32	0.493	0.481	26.69	24.59
	64	0.508	0.467	33.89	28.86
3	16	0.438	0.844	25.81	24.67
	32	0.553	0.552	43.55	27.55
	64	0.452	0.397	212.19	40.48
4	16	0.502	0.740	26.07	25.44
	32	0.479	0.431	43.62	29.48
	64	0.538	0.424	213.08	45.26

Fig. 4. The minimum recorded MSE for the validation dataset in 30 epochs for different hidden layers, neurons, and activation functions. Every trained NN structure is written into a Verilog-A file, imported in Cadence AWR, and recorded the time to simulate 377 thousand data points.

The lowest ReLU in the validation dataset is found by using 3 hidden layers, 64 neurons per hidden/layer, and using the ReLU activation function. Fig. 5 shows the excellent I-V fitting of the NN against the measured data. With the help of the P_n, all devices are modeled with high accuracy. Fig. 6 shows the summary of the error across V_G and V_D in the entire validation dataset. At maximum V_G and V_D, the observed maximum error is 3mA, and the maximum percentage error is 3.4%. The accuracy of the NN-based compact model can still be improved by increasing the training dataset and applying regularization techniques. Although the focus of this paper is to demonstrate the capability of a flexible NN-based compact model to capture variations in IV data, the limits of its achievable accuracy require further investigation.

IV. Conclusion

This paper has presented a new technique to model the manufacturing variability of GaN HEMT using a neural network. We have presented a technique to automatically generate parameters from the variations in I-V data from devices. The results show excellent accuracy in modeling all devices. The trained neural network can be converted to a verilog-a file that can be used by circuit simulators that can be used for IC design. The presented technique can be applied to any device technology and shows the great potential of creating fast and

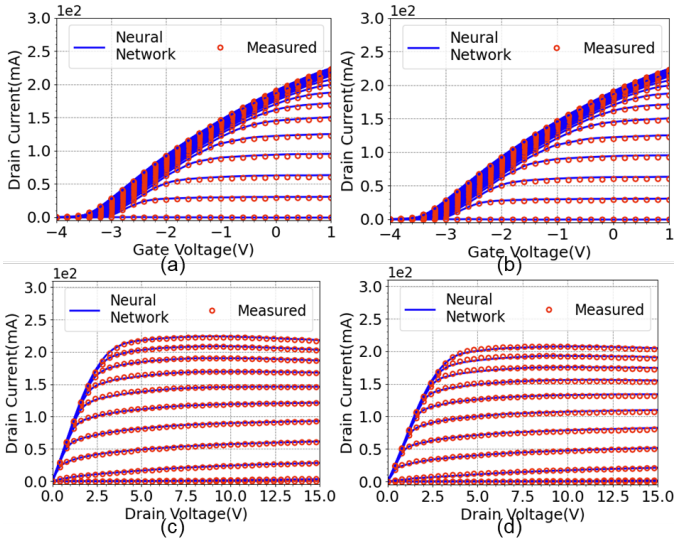


Fig. 5. The transfer characteristics (a)(b) and the output characteristics (c)(d) of the GaN HEMT device and the NN. A random GaN-HEMT device is selected in the validation dataset for (a) and (c), while the worst observed error between the measured and predicted GaN-HEMT is shown in (b) and (d). The NN has 3 hidden layers, 64 neurons per layer and uses the ReLU activation function.

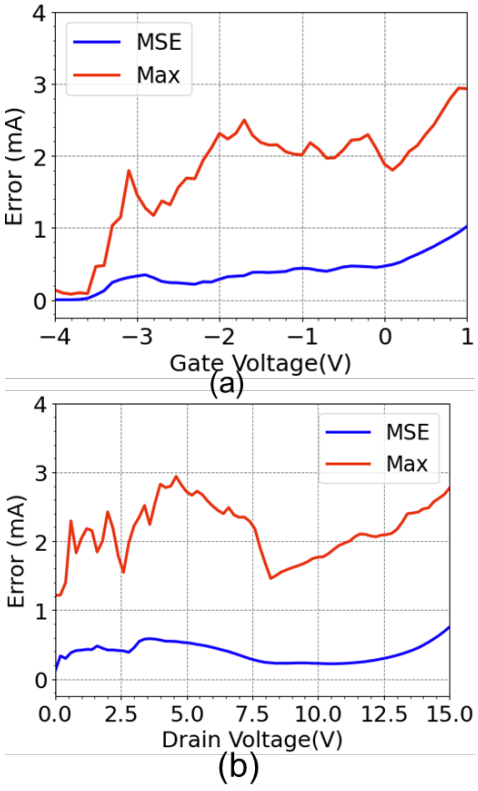


Fig. 6. The observed error in the 23 devices in the validation dataset across (a)V_G and (b) V_D.

flexible neural network-based compact models. The technique allows the neural network-based compact model to be used for yield analysis and for gaining insight into unexpected performance issues due to process variations.

REFERENCES

[1] J. Kim, Y. Kim, S. Oh, J. Choi, D.-H. Lee, K. Cho, S. Lee, and C.-H. Ahn, "A 20-w wide bandwidth gan hemt power amplifier for vhf/uhf applications," *IEEE transactions on industrial electronics*, vol. 67, no. 12, pp. 10 905–10 910, 2019.

[2] K. H. Hamza and D. Nirmal, "A review of gan hemt broadband power amplifiers," *AEU-International Journal of Electronics and Communications*, vol. 116, p. 153040, 2020.

[3] F. Chavez, N. C. Miller, D. T. Davis, and S. Khandelwal, "Statistical modeling of manufacturing variability in gan hemt iv characteristics with asm-hemt," in *2022 IEEE/MTT-S International Microwave Symposium-IMS 2022*. IEEE, 2022, pp. 375–377.

[4] F. Chavez, D. T. Davis, N. C. Miller, and S. Khandelwal, "Deep learning-based asm-hemt iv parameter extraction," *IEEE Electron Device Letters*, vol. 43, no. 10, pp. 1633–1636, 2022.

[5] M. Li, O. İrsoy, C. Cardie, and H. G. Xing, "Physics-inspired neural networks for efficient device compact modeling," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 2, pp. 44–49, 2016.

[6] L. Zhang and M. Chan, "Artificial neural network design for compact modeling of generic transistors," *Journal of Computational Electronics*, vol. 16, pp. 825–832, 2017.

[7] H. Habal, D. Tsonev, and M. Schweikardt, "Compact models for initial mosfet sizing based on higher-order artificial neural networks," in *Proceedings of the 2020 ACM/IEEE Workshop on Machine Learning for CAD*, 2020, pp. 111–116.

[8] Q. Yang, G. Qi, W. Gan, Z. Wu, H. Yin, T. Chen, G. Hu, J. Wan, S. Yu, and Y. Lu, "Transistor compact model based on multigradient neural network and its application in spice circuit simulations for gate-all-around si cold source fets," *IEEE Transactions on Electron Devices*, vol. 68, no. 9, pp. 4181–4188, 2021.

[9] J. Wang, Y.-H. Kim, J. Ryu, C. Jeong, W. Choi, and D. Kim, "Artificial neural network-based compact modeling methodology for advanced transistors," *IEEE Transactions on Electron Devices*, vol. 68, no. 3, pp. 1318–1325, 2021.

[10] K. Nishiguchi, T. Kawasaki, and M. Tanomura, "Neural network based gan hemt modelling for millimeter wave power amplifiers," in *2022 IEEE/MTT-S International Microwave Symposium-IMS 2022*. IEEE, 2022, pp. 382–385.

[11] J. Wei, H. Wang, T. Zhao, Y.-L. Jiang, and J. Wan, "A new compact mosfet model based on artificial neural network with unique data preprocessing and sampling techniques," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 4, pp. 1250–1254, 2022.

[12] G. Qi, Q. Yang, X. Chen, Z. Zhang, P. Zhou, W. Bao, and Y. Lu, "The device compact model based on multi-gradient neural network and its application on mos 2 field effect transistors," in *2022 6th IEEE Electron Devices Technology & Manufacturing Conference (EDTM)*. IEEE, 2022, pp. 88–90.

[13] C.-T. Tung and C. Hu, "Neural network-based bsim transistor model framework: Currents, charges, variability, and circuit simulation," *IEEE Transactions on Electron Devices*, vol. 70, no. 4, pp. 2157–2160, 2023.

[14] W. Dai, Y. Li, B. Peng, L. Zhang, R. Wang, and R. Huang, "Benchmarking artificial neural network models for design technology co-optimization," in *2023 International Symposium of Electronics Design Automation (ISEDA)*. IEEE, 2023, pp. 423–427.

[15] Q. Chen and G. Chen, "Artificial neural network compact model for tfts," in *2016 7th International Conference on Computer Aided Design for Thin-Film Transistor Technologies (CAD-TFT)*. IEEE, 2016, pp. 1–1.

[16] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analyzers," *Neural computation*, vol. 11, no. 2, pp. 443–482, 1999.

[17] T. Minka, "Automatic choice of dimensionality for pca," *Advances in neural information processing systems*, vol. 13, 2000.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[19] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[20] M. A. Mercioni and S. Holban, "The most used activation functions: Classic versus current," in *2020 International Conference on Development and Application Systems (DAS)*. IEEE, 2020, pp. 141–145.

[21] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, 2022.

Use of DC Probes for Multi-MHz Measurements of Crosstalk and Substrate Coupling in Gallium Nitride Power Integrated Circuits

Miao Cui and Sang Lam*
Department of Electrical and Electronic Engineering
Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu Province, China 215123
*Email: s.lam.cn@ieee.org

***Abstract*—With simple compact pads, DC probes were used for measurements of crosstalk and substrate coupling in gallium-nitride (GaN) power integrated circuits (ICs). By proper electrical measurement calibration of the DC probing system, a crosstalk voltage down to 4.4 mV and substrate coupling up to -38 dB were measured up to 25 MHz for a GaN-on-Si power IC, using an oscilloscope and a spectrum analyzer respectively. Despite some accuracy limitations, it is a low-cost viable method for determining especially substrate coupling in power ICs.**

I. Introduction

Signal integrity is important in integrated circuits (IC), as crosstalk and noise coupling through the semiconductor substrate can jeopardize the performance of especially high-frequency mixed-signal circuits integrated on the same chip. There were plentiful research works on crosstalk and substrate coupling in ICs on the semiconducting silicon (Si) substrate [1]-[4]. S-parameter measurements using a vector network analyzer (VNA) are commonly used for investigating crosstalk and substrate coupling [3]-[4] or microwave signal isolation [5]-[6]. This way however requires IC layout of ground-signal-ground (GSG) testing structures [5] surrounding devices and circuits for on-wafer measurements. Much chip area would be used up for such GSG testing structures which must be included in the IC layout design prior to the fabrication. Over the past decade, there has been investigation of crosstalk and substrate coupling in the emerging gallium nitride (GaN) IC technology [7]-[10]. So far, little research has been reported on using on-wafer measurements to determine crosstalk and substrate coupling in GaN-based power ICs, especially in the GaN-on-Si technology. In this work, we try using a DC probing system to investigate the crosstalk and substrate coupling in GaN ICs but without using a VNA and GSG layout structures. This makes the measurements feasible for obtaining useful results even though advanced microwave measurement facilities are unavailable.

II. Device Structures & Measurement Set-up

A. GaN-on-Si Device Structures

Fig. 1 shows the device structures fabricated in GaN IC technology. The GaN IC consists of AlGaN/GaN metal-insulator-semiconductor high-electron mobility transistors (MIS-HEMTs) fabricated on a GaN-on-Si substrate. The AlGaN/GaN epitaxy consists of a 25-nm AlGaN barrier and a 5-µm GaN buffer layer on a Si wafer. The device structure and fabrication details are more or less the same as reported in previous works [11]-[12]. In our experimental investigation, the MIS-HEMTs are electrically isolated and separated by a distance of ≈700 µm. An interference source of switching transients is assumed to be at the source/drain of one MIS-HEMT; and then the crosstalk voltage and power is measured at the source/drain of another MIS-HEMT. The detected signal with the same frequency as that of the interference source can then be used to estimate the substrate coupling in the GaN IC.

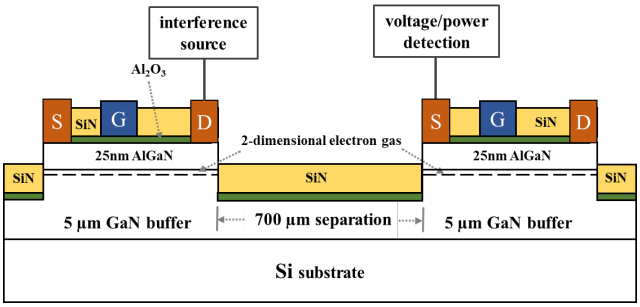


Fig. 1. Diagram showing the structures of devices on GaN-on-Si substrate for measurements of crosstalk and substrate coupling, with the interference source applied at a location separated by 600-700 µm from the point of detection by an oscilloscope or spectrum analyzer.

B. On-Wafer DC Probing Set-up

Fig. 2 shows a diagram illustrating the experimental set-up of a DC probing system for the on-wafer measurements of crosstalk and substrate coupling. It can be seen that only two inexpensive DC probes are needed and each of them requires only a small probe pad (100 µm × 100 µm in our work) for the on-wafer electrical contact. This saves the complex GSG layout requirement. In our GaN power ICs, the circuits had been designed to operate below 1 MHz. For this reason, it was unnecessary for adopting the GSG layout for any of our measurements planned in the design stage. As a result, we had no choice but to use DC probes in on-wafer measurements reported in this work. As for the electrical measurement instruments, a signal generator and a digital oscilloscope as common electronics laboratory equipment are sufficient. To measure the signal power at different frequencies, a spectrum analyzer is needed. A fairly inexpensive spectrum analyzer model was used in our measurements. Table I lists the

instruments and some key information related to the measurements. Both the spectrum analyzer and the signal generator have a 50-Ω impedance at the input and ouput connections respectively. This has an important implication for estimating the power level from the voltage amplitude of the electrical signal and vice versa.

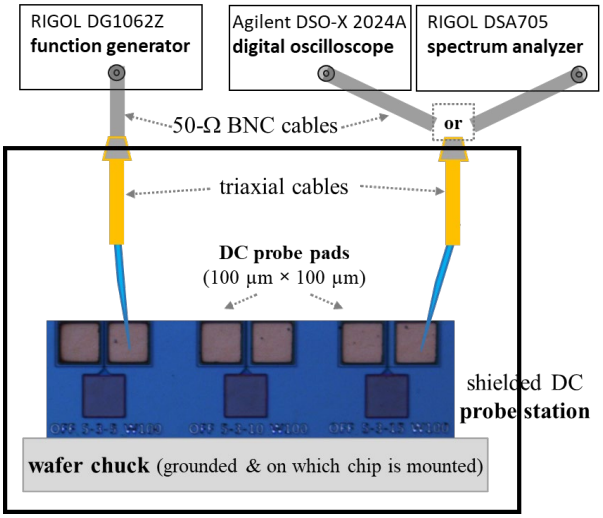


Fig. 2. Diagram illustrating the experimental set-up for on-wafer measurements of crosstalk and substrate coupling, using a function/signal generator as the interference source and a digital oscilloscope or a spectrum analyzer for detection.

TABLE I. ELECTRICAL INSTRUMENTS USED IN THE CROSSTALK MEASUREMENTS AND THEIR KEY SPECIFICATIONS

Instrument Used	Model Number	Input/Output Impedance	Frequencies of Interest
signal generator	RIGOL DG1062Z	50 Ω	1 µHz -60 MHz
digital oscilloscope	Agilent DSO-X 2024A	1 MΩ; 11 pF (in parallel)	0 - 200 MHz
spectrum analyzer	RIGOL DSA705	50 Ω	100 kHz - 500 MHz

III. SIGNAL LOSS CALIBRATION OF DC PROBING SYSTEM

A. *Calibration to Account for Signal Loss*

Before the actual measurements of crosstalk and substrate coupling, calibration of the experimental set-up was performed first. This is highly necessary as DC probing systems are typically for DC or low frequency measurements (below 1 MHz). The frequency dependence of the signal propagation in DC probing is almost sure, even though the wavelength of a multi-megahertz electrical signal (e.g. 3 MHz or below) would be much larger (10 times) than the size of the whole DC probing system (of about 2 m in each dimension). In electrical measurements using a VNA, it is routine to do calibration of the measurement set-up to automatically eliminate the signal loss and phase change associated with the set-up itself (including the cables and connections). In our work, the calibration of the DC probing system covers the signal loss but not the phase change, using a clearly set waveform output from a signal generator.

B. *Voltage Waveform & Spectral Contents*

A square waveform signal of a 3.0-V peak-to-peak voltage (V_{pp}) was used to serve as the interference source of the switching transients, which are common in GaN power ICs [12]. The voltage signal is fed to one DC probe and the other DC probe was placed on the same probe pad for picking up the signal and then detected by either the digital oscilloscope or the spectrum analyzer. Fig. 3 shows the waveform of a 400-kHz signal after propagating through the signal path (including the BNC and triaxial cables in addition to the DC probes). The peak-to-peak voltage can be measured from the waveform.

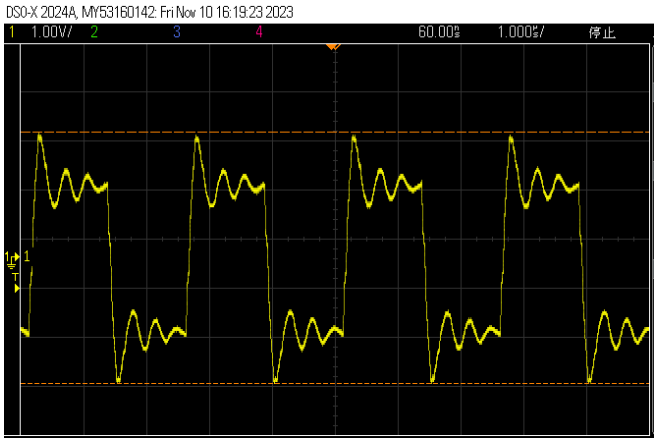


Fig. 3. Trace of a voltage signal measured by the digital oscilloscope when a 400-kHz square waveform of 3 V_{pp} is fed to one DC probe with another DC probe on the same pad picking up the signal.

The voltage signal in the frequency domain can be observed by using a spectrum analyzer. The spectral contents of the detected signal can be seen more clearly (Fig. 4), revealing precisely the power level of each spectral peak. The power level of signal at the fundamental frequency (f_0) is mainly used in our work to determine the crosstalk and substrate coupling.

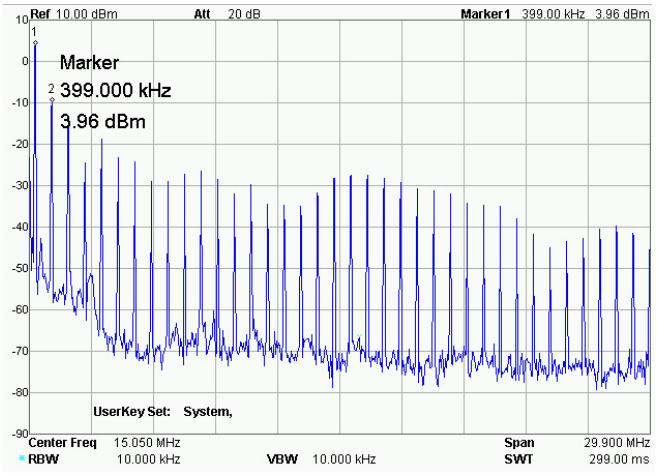


Fig. 4. Spectrum of a voltage signal measured by the spectrum analyzer when a 400-kHz square waveform of 3 V_{pp} is fed to one DC probe with another DC probe on the same pad picking up the signal (the same probing conditions as that for Fig. 3).

Fig. 5 and Fig. 6 show the frequency response (in voltage and power levels respectively) of our on-wafer measurement set-

up when both DC probes are in electrical contact by being placed on the same probe pad. The results were obtained by feeding one DC probe with square waveform signals of increasing frequencies up to 25 MHz. Attenuation is clearly seen especially at multi-MHz frequencies and beyond. In Fig. 5, the detected signal at 2 MHz and below has an even larger peak-to-peak voltage than that specified in the signal generator. This is partly due to the overshoot in the square waveform. The trend is overall the same when a sinusoidal signal was used (as shown by the dark green curve in Fig. 5). The "voltage gain" mainly at 2 MHz is likely caused by the resonance of some parasitic inductance of the DC probes and the probe pad capacitance.

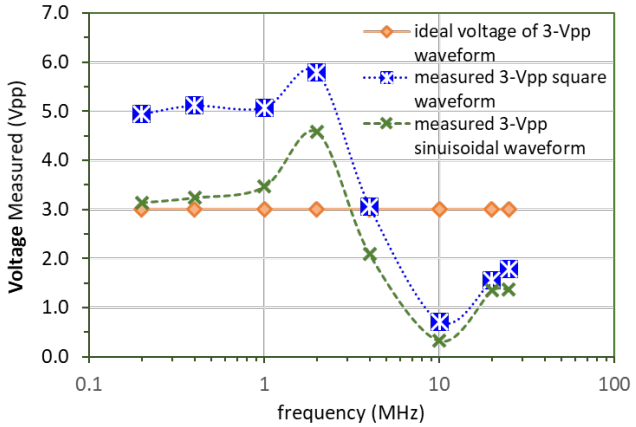


Fig. 5. Measured peak-to-peak voltage when 3-V_{pp} voltage signals of increasing frequencies (from 200 kHz to 25 MHz) were fed to one DC probe with another DC probe on the same pad picking up the signals – tested with the square waveform input as well as a sinuisoidal input.

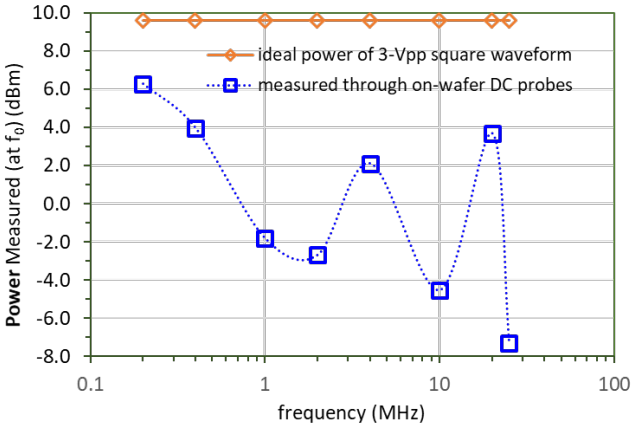


Fig. 6. Measured power levels when 3-V_{pp} square waveform signals of varying frequencies (from 200 kHz to 25 MHz) were fed to one DC probe with another DC probe on the same pad picking up the signals.

C. Fluctuating Signal Path Loss from 200 kHz to 25 MHz

Fig. 7 shows the voltage and power ratios of the signals propagating through the DC probing signal path. These data are essential for system calibration so that the later measured voltage and power levels at the other GaN transistor could be used to deduce the crosstalk and substrate coupling. The fluctuations of the signal propagation loss (Fig. 7) over even a small frequency span (from 1 MHz to 10 MHz) indicate one shortcoming of the DC probing: the measured crosstalk voltage at certain frequencies might not be accurate unless the peak-to-peak voltage at finely-resolved frequencies is consistently measured in calibration. In other words, the magnitude of the crosstalk voltage measured at one frequency must be compared with the voltage at the same frequency measured in the calibration – the voltage ratio matters but not the voltage itself. Otherwise, the measured crosstalk voltage would not be very useful in telling about the substrate coupling in the power IC chip.

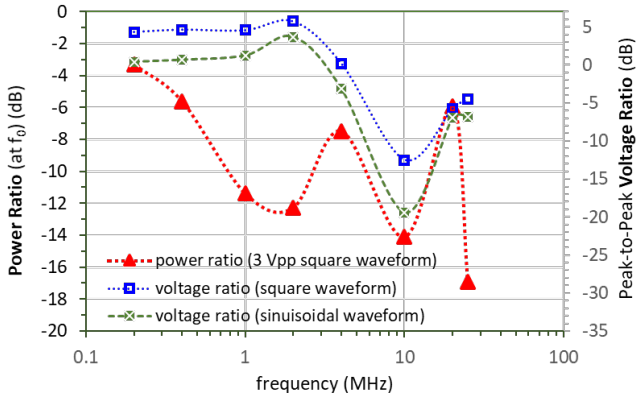


Fig. 7. Power ratio and voltage ratio of the measured signals to that of the input signals of frequencies from 200 kHz to 25 MHz, showing the loss of signal path in the DC probing system (including cable and connection loss).

There is another shortcoming in the crosstalk voltage measurement for determining the substrate coupling when the voltage signal is weak. In using the digital oscilloscope to measure the crosstalk voltage, signals of less than 1 mV_{pp} seemed to be corrupted by noise and their waveform cannot be displayed properly. Fig. 8 shows an example of a noisy waveform of about 800 μV_{pp}. Such a noisy waveform of 800 μV_{pp} is equivalent to a power level of about -64 dBm in a 50-Ω load (as in a typical spectrum analyzer). Using the inexpensive spectrum analyzer model (Table I), the minimum detectable power level can be down to -75 dBm (Fig. 4). This can be even lower by using a smaller resolution bandwidth (RBW) but at the expense of displaying the trace slowly.

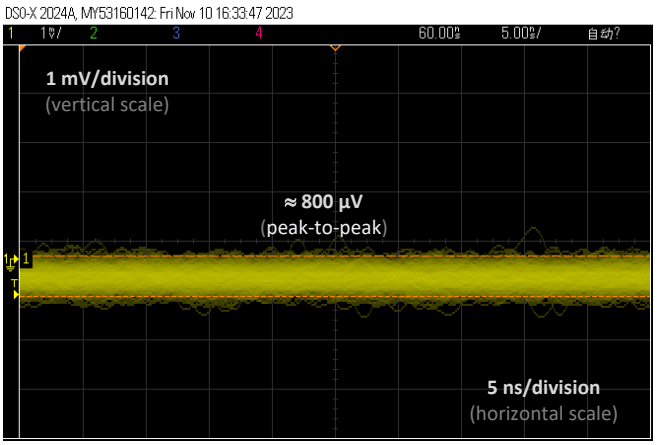


Fig. 8. Screenshot of the digital oscilloscope showing a noise pattern of ≈ 800 μV_{pp} when the voltage signal is too weak for detection in the time domain for seeing the waveform.

D. Signal Path Loss from Measurement of Power Levels

To complement the signal path loss calibration measurements reported above, the power levels of the harmonics

of a square waveform as the interference source (Fig. 1) can be measured conveniently using a spectrum analyzer. Fig. 9 shows a 400-kHz square waveform of 3 V_{pp} voltage signal, fed through a 0.3-m coaxial cable from the signal generator to a digital oscilloscope. Comparing Fig. 9 with Fig. 3, it can be seen clearly about the distortion of the square waveform when the voltage signal propagates through the DC probing system. It implies the voltage signal loss of the DC probing system at arbitrary frequencies. Instead of increasing the frequency of the square waveform signal, the power levels of the harmonics (more than 35) can be measured for the 400-kHz signal which has a similar spectrum to that in Fig. 4 but simply with different power levels for the harmonics.

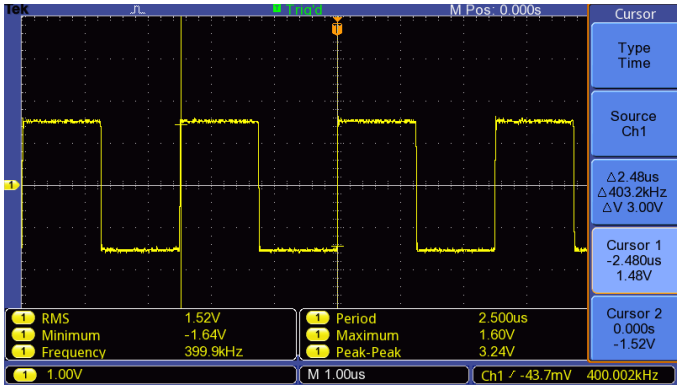


Fig. 9. Trace of a 400-kHz square waveform of 3 V_{pp} signal fed directly from the signal generator through a 0.3-m coaxial cable to the digital oscilloscope.

Fig. 10 shows the power levels of the harmonics of the 3-V_{pp} 400-kHz square waveform signal, comparing the ideal case and the actual case through a 0.3-m coaxial cable with that of the DC probing system. The detected power levels of the harmonics through the DC probing system are those in the spectrum shown in Fig. 4. As expected, the detected power levels (red) of the harmonics fall short of the ideal values and actual power levels.

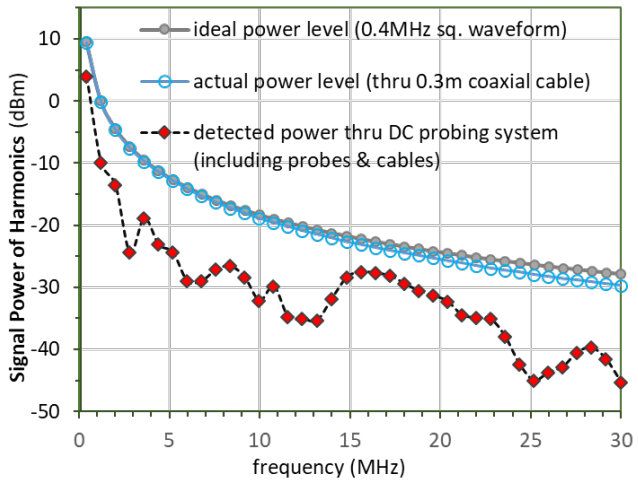


Fig. 10. Comparison of the signal power levels of harmonics of the 3-V_{pp} 400-kHz square waveform signal in three cases: the ideal case without any signal loss, the case of connection through only a 0.3-m coaxial cable, and the case of the DC probing system.

A square waveform $V(t)$ has infinitely many harmonics:

$$V(t) = V_{DC} + \frac{4}{\pi}\sum_{n=1}^{\infty}\left(\frac{V_0}{n}\right)\sin[n(2\pi f_0 t)],\ n = 1, 3, 5 \ldots \quad (1)$$

where $V_{DC} = 0$ V, $V_0 = 1.5$ V being the voltage amplitude (with 3.0 V_{pp}) and $f_0 = 400$ kHz being the fundamental frequency for the square waveform signal in Fig. 9. Accordingly, the value of the voltage amplitude is $6/\pi$ at the fundamental frequency (f_0), $2/\pi$ for the third harmonic ($3f_0$), and $1.2/\pi$ for the fift harmonic ($5f_0$), and so on and so forth. With the 50-Ω impedance of the signal generator and spectrum analyzer, the voltage amplitude detected would be half of the values: $3/\pi$ at f_0, $1/\pi$ at $3f_0$, $0.6/\pi$ at $5f_0$ etc. The detected power by the spectrum analyzer for the harmonics should be ideally as below:

$$P = \left(\frac{V_{hn}}{\sqrt{2}}\right)^2/50 = \left[\frac{1}{\sqrt{2}}\left(\frac{2V_0}{n\pi}\right)\right]^2/50\ ,\ n = 1, 3, 5 \ldots \quad (2)$$

where V_{hn} is the voltage amplitude (across the 50-Ω impedance) of the corresdponding n^{th} harmonic of the square waveform signal. Equation (2) gives $P_0 = 9.12$ mW (equivalent to 9.6 dBm) at f_0, $P_3 = 1.01$ mW (equivalent to 0.1 dBm) at $3f_0$, $P_5 = 0.37$ mW (equivalent to -4.4 dBm) at $5f_0$, and so on. These power levels of the harmonics in the ideal case (the grey dot curve in Fig. 10) were almost the same as those actual measured power levels (the pale blue circle curve in Fig. 10) when the square waveform signal was fed through a 0.3-m coaxial cable. With these data, the frequency-dependent signal path loss of the DC probe system can be determined, as shown in Fig. 11. The fluctuation of the DC probing system loss is smilar to that in Fig. 7 but the frequency points are more finely resolved in Fig. 11. The fluctuation can be more than 10 dB and therefore the system loss calibration is critical for using the DC probing system for any sensible measurements of the substrate coupling.

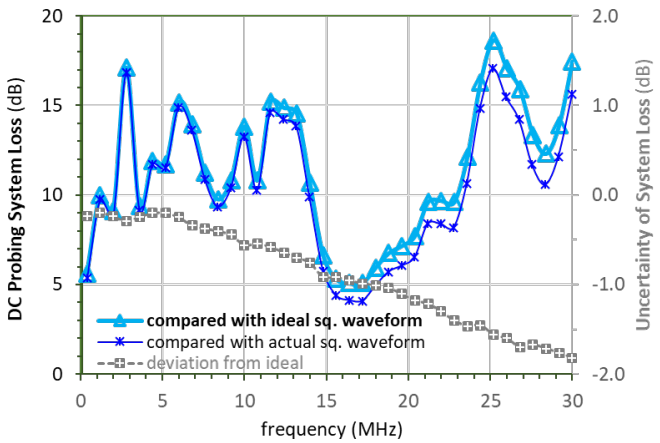


Fig. 11. The fluctuating signal path loss of the DC probing system up to 30 MHz, as determined by measuring the power levels of the harmonics of the 400-kHz square waveform signal.

IV. Crosstalk & Substrate Coupling Results

After proper calibration of the DC probing signal path of the experimental set-up, the signal detected at the other GaN MIS-HEMT separated by 600-700 µm from the first transistor can reveal how strong the crosstalk voltage and the substrate coupling are. Fig. 12 and Fig. 13 show the crosstalk

measurement results, in voltage and power levels respectively. The ratio of the "gross" crosstalk voltage to voltage through DC probes is the substrate coupling (in voltage), shown in dB on the secondary axis in Fig. 12. It is similar when looking at the power levels of the crosstalk signal: the substrate coupling is the difference in the power levels (expressed in dBm). Based on the measurement results, it seems more reliable for the substrate coupling measurements using a spectrum analyzer to detect the power levels of the crosstalk signal. In both voltage and power level measurements, the substrate coupling results of -70 dB to about -35 dB are comparable to those of silicon ICs [3],[6]. This indirectly supports the validity of our measurements using a DC probing system up to 25 MHz, despite the unavailable GSG test structures on our GaN power IC chips. Verification by computational electromagnetics has been in progress to determine the substrate coupling in the Ga-on-Si power ICs. Comparison of the EM simulation results with the measurement results obtained using the DC probing system is to be presented.

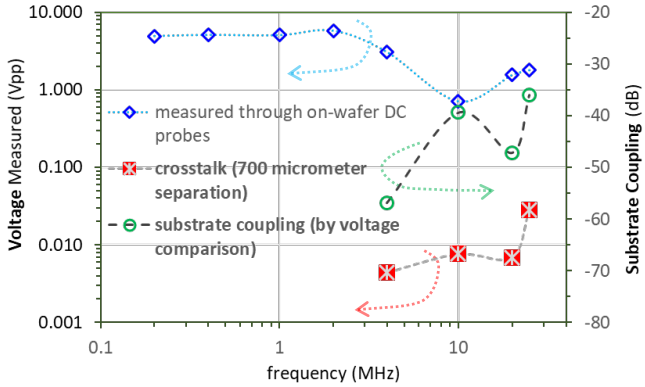


Fig. 12. On-wafer measurement results of crosstalk and substrate coupling by measuring the voltage waveform (in V_{pp}) with the digital oscilloscope, showing the substrate coupling between -57 dB at 4 MHz to -35 dB at 25 MHz. (Note the three decimal places of the numbers on the *y*-axis).

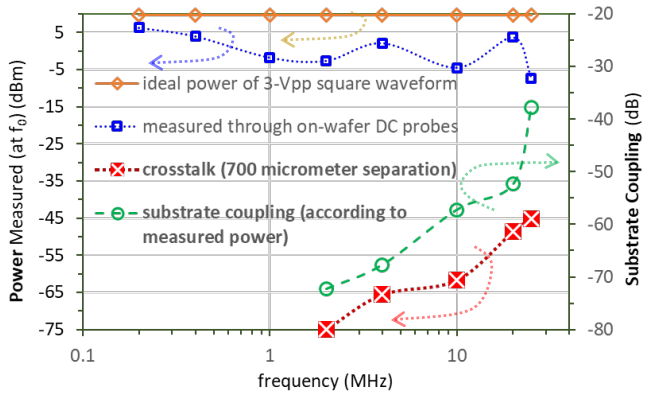


Fig. 13. On-wafer measurement results of crosstalk and substrate coupling by signal detection using the spectrum analyzer, showing an increasing substrate coupling from -70 dB at 2 MHz to -38 dB at 25 MHz.

V. Conclusion

We have demonstrated the use of DC probes with system calibration for measurements of crosstalk and substrate coupling up to 25 MHz in GaN-on-Si power ICs. Only simple and compact probe pads were good enough for the measurements, requiring no IC layout of GSG testing structures. Being spare from using VNA or other expensive microwave equipment, the method is helpful for estimating substrate coupling in power ICs with monolithic integration of both sensitive circuits and high-power/voltage ones, at frequencies extendable to ≈100 MHz.

Acknowledgment

The authors acknowledge the funding support from EEE Department, School of Advanced Technology (SAT), Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu Province, China. S. Lam appreciated the project work done by his formerly co-supervised students: Mr. Yi Shen, Mr. Ang Li, and Ms. Ziqian Li with whom the research discussions and advising helped stimulate various ideas for research on GaN devices and circuits that paved the work reported here in this paper.

References

[1] David K. Su, M.J. Loinaz, S. Masui, and Bruce A. Wooley, "Experimental results and modeling techniques for substrate noise in mixed-signal integrated circuits," *IEEE Journal of Solid State Circuits*, vol. 28, no. 4, pp. 420-430, April 1993.

[2] M. Pfost and H.-M. Rein, "Modeling and measurement of substrate coupling in Si-bipolar IC's up to 40 GHz," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 4, pp. 582-591, March 1998.

[3] Alan L. L. Pun, Tony Yeung, Jack Lau, François J. R. Clément, and David K. Su, "Substrate noise coupling through planar spiral inductor," *IEEE Journal of Solid State Circuits*, vol. 33, no. 6, pp. 877-884, June 1998.

[4] Min Xu, David K. Su, Derek K. Shaeffer, Thomas H. Lee, and Bruce A. Wooley, "Measuring and modeling the effects of substrate noise on the LNA for a CMOS GPS receiver," *IEEE Journal of Solid State Circuits*, vol. 36, no.3, pp. 473-485, March 2001.

[5] Sang Lam, Philip K. T. Mok, Ping K. Ko, and Mansun Chan, "High-isolation bonding pad design for silicon RFIC up to 20 GHz," *IEEE Electron Device Letters*, vol. 24, No. 9, pp. 601-603, September 2003.

[6] Wen-Kuan Yeh, Shuo-Mao Chen, and Yean-Kuen Fang, "Substrate noise-coupling characterization and efficient suppression in CMOS technology," *IEEE Transactions on Electron Devices*, vol. 51, no. 5, pp. 817-819, May 2004.

[7] V. Unni, H. Kawai, and E.M.S. Narayanan, "Crosstalk in monolithic GaN-on-Silicon power electronic devices," *Microelectronics Reliability*, vo. 54, pp. 2242-2247, August 2014.

[8] Qimeng Jiang, Zhikai Tang, Chunhua Zhou, Shu Yang, and Kevin J. Chen "Substrate-coupled cross-talk effects on an AlGaN/GaN-on-Si smart power IC platform," *IEEE Transactions on Electron Devices*, vol. 61, no. 11, pp. 3808-3813, November 2014.

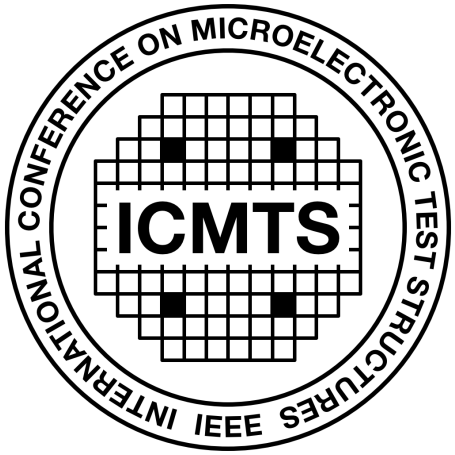
[9] S. Aamir Ahsan, S. Ghosh, S. Khandelwal, and Y. S. Chauhan, "Analysis and modeling of cross-coupling and substrate capacitances in GaN HEMTs for power-electronic applications," *IEEE Transactions on Electron Devices*, vol. 64, no. 3, pp. 816-823, March 2017.

[10] Ruize Sun, Jingxue Lai, Wanjun Chen, Chao Liu, Fangzhou Wang, Jianjun Zhou, Zhaoji Li, and Bo Zhang, "Crosstalk suppression in monolithic GaN devices based on inverted E-field decoupling," *IEEE Transactions on Electron Devices*, vol. 68, no. 4, pp. 1542 - 1549, April 2021.

[11] Miao Cui, Yutao Cai, Sang Lam, Wen Liu, Chun Zhao, Ivona Z. Mitrovic, Stephen Taylor, Paul R. Chalker, and Cezhou Zhao, "Characterization of Transient Threshold Voltage Shifts in Enhancement- and Depletion-mode AlGaN/GaN Metal-Insulator-Semiconductor (MIS)-HEMTs", *2018 IEEE International Conference on Electron Devices and Solid-State Circuits* (*EDSSC 2018*), Shenzhen, China, June 2018.

[12] Miao Cui, Qinglei Bu, Yutao Cai, Ruize Sun, Wen Liu, Huiqing Wen, Sang Lam, Yung C. Liang, Ivona Z. Mitrovic, Stephen Taylor, Paul R. Chalker, and Cezhou Zhao, "Monolithic Integration Design of GaN-based Power Chip Including Gate Driver for High-Temperature DC-DC Converters," *Japanese Journal of Applied Physics*, vol. 58, no. 5, 056505, 25 April 2019.

SESSION 9
Process/Material Monitoring



18th April 2024, 11:20–12:20

Session Co-Chairs: Jonathan Terry, *The University of Edinburgh, UK*
Chadwin Young, *University of Texas at Dallas, USA*

An Add-in Test Structure Chip to Unitedly Assess PVD Material Properties in University Open Nanotechnology Platform

Shun Yasunaga[1], Kei Misumi[1], Ayako Mizushima[1], Atsushi Toyokura[1], Etsuko Ota[1], Yurie Inoue[1], Makoto Fujiwara[1], Noriko Kawai[1], Mitsuhiro Yoda[1], Shinji Tsuboi[1], Tomoki Sawamura[1], Akio Higo[1], Ryosho Nakane[1], Yukinori Ochiai[1], and Yoshio Mita[1]

[1]Department of Electrical Engineering, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

Email: teammems@if.t.u-tokyo.ac.jp

***Abstract*—Open nanotechnology platforms at universities and research institutes are becoming essential for agile and rapid microelectronic device research and development. To further extend its capability with more users, process data acquisition and digital transformation is of high priority. In order to uniformly acquire process data in an open platform, where users' process charts are all different, we propose a "chamber sharing system", in which a small test chip designed for easily measuring necessary parameters is loaded in the process chamber together with user's sample, which will be submitted to the platform. As the first realization of such a test chip, we designed and fabricated an 8×8 mm² chip for physical vapor deposition, enabling the film's thickness, stress, and conductivity measurement. With easy preparation requiring no cleanroom environment, results compatible with previously reported values were obtained.**

I. Introduction

Conducting semiconductor microdevice research is expensive as it requires a series of fabrication apparatuses. To efficiently continue research and development in this domain, a multi-use open platform has appeared since the beginning of the 21st century, such as NNIN (USA), RENATECH (France), EuroNanoLab (EU), and ARIM (Japan). This approach of enabling academic researchers to afford the costly research environment has seen success. For example, Takeda Super Cleanroom [1] in The University of Tokyo, Japan, hosts 1092 individual users from 200 independent research groups as of December 2023 as a hub facility of the ARIM project with 80 semiconductor fabrication and test instruments equipped in a 600 m^2 cleanroom. This scale of research capacity is what one cannot expect without such an open platform, accelerating not only individual user's research activity but also collaboration between groups.

Getting a successful process result in the first trial is not just a desire but a necessity when machines are almost entirely occupied with further increasing demand awaiting as a result of activated use. A comprehensive recipe set that can cope with all materials and parameters is essential. However, preparing such recipes in open platforms is not easy. In most use cases, the process result will be lost from the cleanroom because the real silicon devices are for the users, being subject to be brought back to their laboratory for further experiments. In addition, time slots of the instrument for process development by the staff will be and should be limited to maximize the time for users. Our proposal here is to introduce a "chamber sharing system" (Fig. 1). A small test chip (ideally less than 1 cm^2), prepared by the platform, will be introduced in the same process chamber while the user who consents to do so conducts their process. The user's wafer and the test chip thereby undergo the same process. The platform keeps "the real silicon" containing the process result and performs whatever necessary analyses while the user continues with their chip.

Such an effort to evaluate the process by introducing another test chip into the process chamber has been considered previously for specific processes such as supercritical fluid deposition [2]. However, this paper proposes that the platform acquires numerous results of highly diverse processes in a uniform manner. In this scheme, every process produces one chip to be analyzed by the platform staff. The measurement must be as easy as possible to cope with the number of samples without further process or fine manipulation. The association of the test result with the process conditions logged in the equipment is also important because the staff will measure the chip among other chips well after the process finishes, rather than the users measure theirs for themselves right away. Such a test chip should be designed for each process group, such as chemical vapor deposition (CVD), physical vapor deposition (PVD), and reactive ion etching (RIE). In this paper, we have designed and evaluated a test chip for PVD as the first realization. The proposed chip enables us to measure the thickness, stress, and conductivity of the deposited thin film without further processing, only using instruments that do not necessitate a cleanroom environment.

This work was jointly supported by ARIM project JPMXP1223UT1212, "UTokyo Agile-X" X-Nics project JPJ011438, JST SICORP JPMJSC22H1, and JST CREST JPMJCR20T2, all by the Ministry of Education, Culture, Sports, Science and Technology (MEXT).

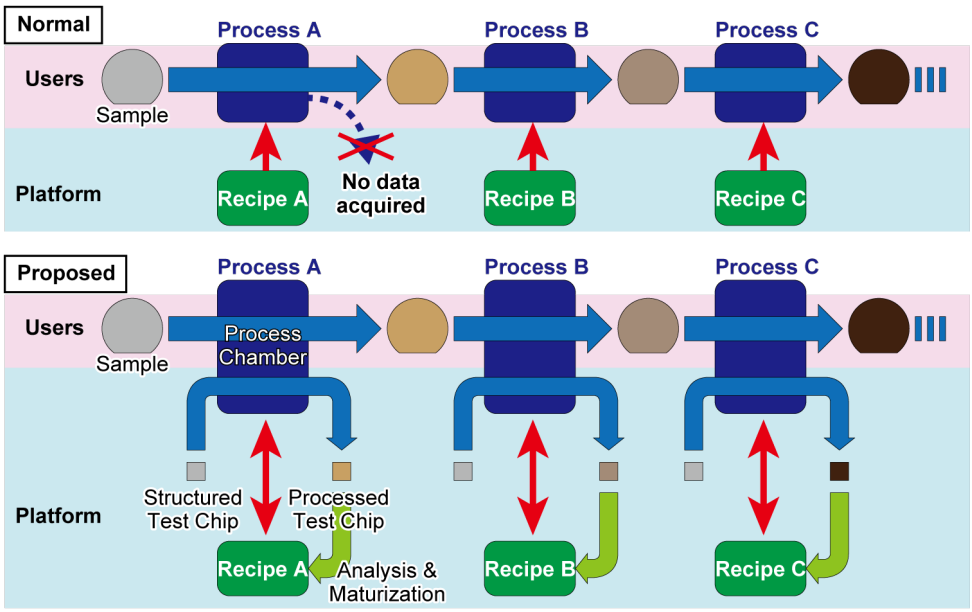


Fig. 1. Concept of "Add-in" process proof extraction scheme. User goes with a set of element processes (from A to C in this figure), while the platform receive and analyze each process with test structure.

II. Design and Fabrication

A. Design of the Test Structures

Three parameters characterizing thin films are to be collected: Thickness, residual stress, and electrical conductivity. The structure for each measurement is placed within an 8 × 8 mm^2 chip area, so it does not affect the area for the user's sample to be processed but is large enough for convenient handling. For the traceability of the chip, each test chip should be uniquely identifiable. To fabricate a prototype of such a test chip, we used direct-writing lithography and silicon etching to pattern silicon-on-insulator (SOI). Then, the structure is released (the sacrificial layer is etched away) if necessary.

For thickness measurement, the stylus profiler or the laser profiler are the most probable tools for versatile processes. Interferometers are another primary tool, but their application is limited to transparent materials. Considering the variety of processes in the open platform cleanroom, the logging chip should be designed to measure using the stylus or laser profiler with as easy preparation as possible. We designed two types of suspended beam structures, as shown in Fig. 2(a). They must be removed before their measurement to expose the undeposited substrate. Design A is ladder-like arrayed suspended beams anchored on both sides, manually broken one by one after deposition using a manipulator. Design B is also ladder-like beams embedded in a 4 × 4 mm^2 sheet, almost entirely released except for the supporting anchors at corners. The sheet and the beams are altogether removed by tape, breaking the anchors. Design A requires less area, while Design B offers faster and easier measurement preparation.

The measurement of residual stress of thin film uses various methods, including X-ray diffraction, Raman spectroscopy, ultrasonic, indentation, and measurements using micro electromechanical systems (MEMS) structures [3]. Among them, the curvature measurement of a cantilever is suitable for amorphous PVD films. The thin film deposited on a microcantilever bends it concave if the film has tensile stress and convex if compressive. It has been known that from the curvature of the cantilever and the thickness of the thin film, the residual stress of the thin film can be obtained using the Stoney formula [4]

$$\sigma = \frac{E_s h_s^2}{(1-\nu_s)6h_f R}, \qquad (1)$$

where E is the Young's modulus, h is the thickness, and ν is the Poisson's ratio. Subscript "s" denotes the substrate, and "f" is the thin film. Although some modification is required in the case of a silicon substrate, which is elastically anisotropic or when the deposition is relatively thick [4,5], the residual stress can be obtained by only measuring the height profile of the cantilever after the thickness is obtained by the structure mentioned above. For the convex deformation by compressive stress, the tip of the released cantilever touches the handle layer of the SOI wafer, hindering further bending if the deformation is larger than the thickness of the oxide layer. We removed the handle layer beneath the cantilevers to let them bend beyond the thickness of the box layer.

For the test structure for conductivity, we placed the conventional Greek-cross structure on the chip [6], which was not measured in this work, simply ensuring the area for this test structure. Together with a QR Code® and an ITF (Interleaved 2 of 5) barcode for a six-digit identification number ("123456" as a dummy number), all the structures were fitted in the 8 × 8 mm^2 chip area. To avoid isolated patterns in the QR Code® getting removed away after vapor HF release, we patterned the code over grids.

B. Device Fabrication

Fig. 3(a) shows the schematic process flow of the test chip. The device layer of SOI wafer (device, box, handle layers are 2, 1, 550 µm thick, respectively), which was diced into 20 mm squares, was directly patterned by rapid electron beam lithography system F5112+VD01 (ADVANTEST, Japan) or the laser photolithography system DWL 66+ (Heidelberg Instruments, Germany) and etched by Bosch process using deep RIE machine (MUC-21 ASE Pegasus, SPP Technologies, Japan). After removing the resist by O_2 plasma etching, the device layer was protected by spincoated photoresist, and its backside (handle layer) was patterned by photolithography. The oxide layer on the backside was first etched by inductively coupled plasma RIE using CHF_3 gas (CE 300I, ULVAC, Japan), followed by deep RIE of the handle layer. After removing the resists, the box layer was isotropically etched by vapor phase fluoric acid (VPE-100, idonus, Switzerland) for 4.5 hours to

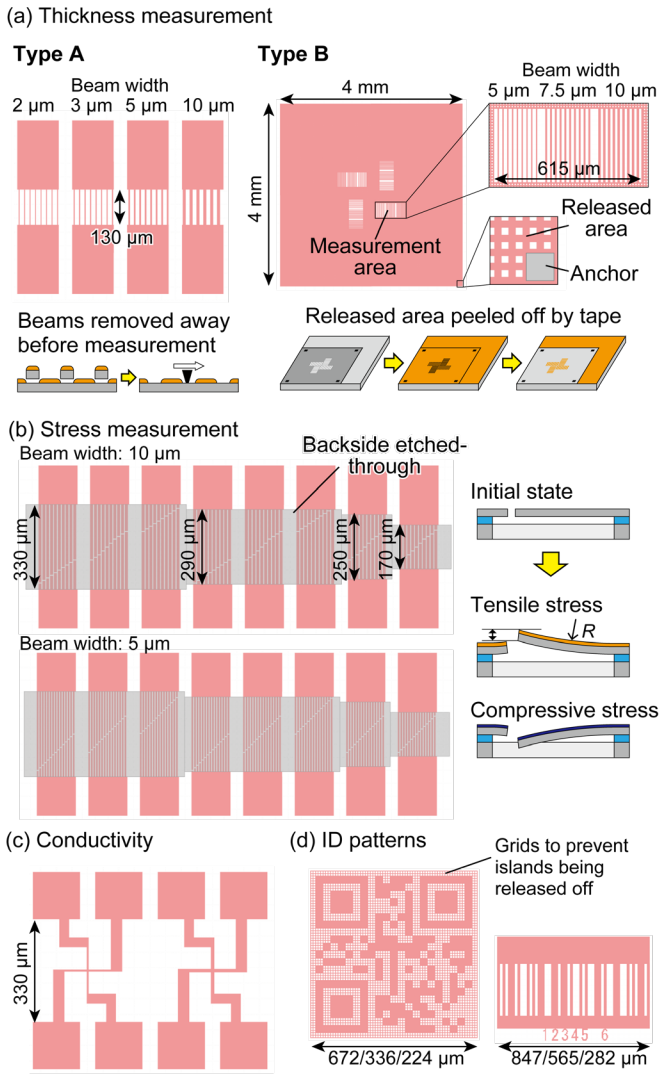


Fig. 2. Design of the test structures for (a) thickness measurement, (b) stress measurement, and (c) conductivity measurement, and (d) QR Code ® and barcode for identification.

release 10 μm-wide structures. Dicing into 8 × 8 mm^2 chips was done simultaneously during the frontside etching, backside etching, and vapor HF etching [7]. In Fig. 3(b), we show the completed chip. The cantilevers for the stress measurement are placed in two areas, oriented orthogonal to each other. Onto the completed test chip, chromium (Cr) and 99:1 aluminum-silicon (AlSi) were sputtered for evaluation.

III. Results and Discussion

The thickness measurement was done for Design A and B after sputtering 220 nm of chromium. Suspended beams in Design A were broken and removed using aluminum wire of a wire bonder. As shown in Fig. 4(a), all the fabricated beams with a width of up to 10 μm were successfully removed to reveal the undeposited surface. For Design B (Fig. 4(b)), although some anchors for fixing the released sheet on the chip were broken off by the film's stress after deposition, the sheet functioned as a stencil as designed and was removed only using polyimide tape. Across the shadowed traces, a stylus profiler, DektakXT (Bruker, USA), with a 12.5 μm-radius tip, measured the profile. The

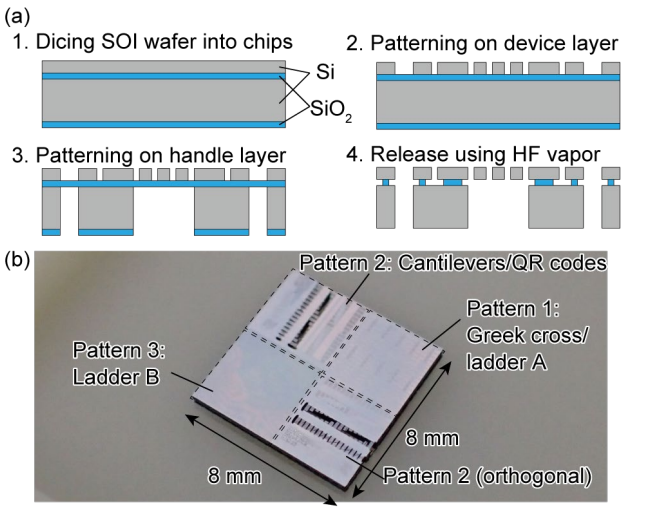


Fig. 3. Fabrication of the test chip. (a) Process flow and (b) completed test chip.

traces were observed as valleys for each beam width; however, the width influenced the measurement. A width of 10 μm gave the valley depth consistent with the actual thickness observed by scanning electron microscope (SEM), but the narrower beams saw shallower valleys accordingly. This result could be attributed to the width of the stylus tip, preventing the stylus from reaching the bottom of the valley. Also, sputtering deposition does not have directionality, meaning that deposition will be made under the suspended beam to some extent. Supposing the deposition is directional, and thus the shadow is a rectangular trench, the 10 μm-wide beams will allow for up to 1.0 μm of deposition. For a thick film deposition, the beams must be widened.

For the stress measurement, bent cantilevers were observed after the 220 nm Cr and 200 nm AlSi deposition, respectively. The SEM images in Fig. 5(a) show the result after Cr deposition by 220 nm. The profiles of 10 μm-wide cantilevers were measured by a laser microscope LEXT OLS5000 (Olympus, Japan). The

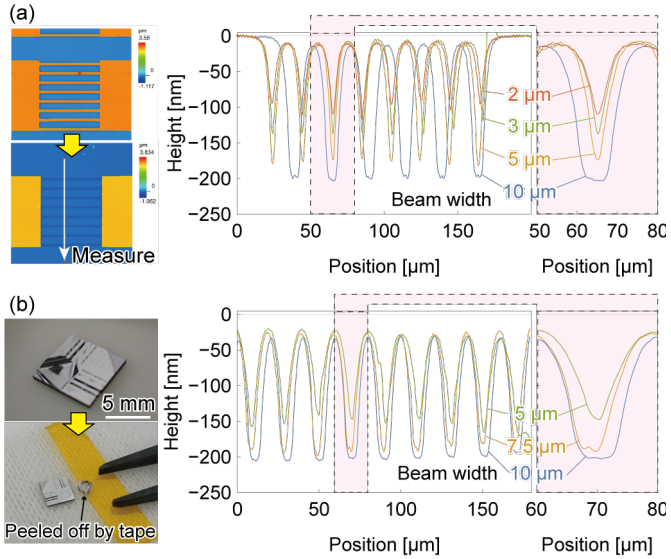


Fig. 4. Results of the thickness measurement using (a) type A and (b) type B, plotted by the width of beams. In the plot, red-shadowed sections are shown magnified in the right.

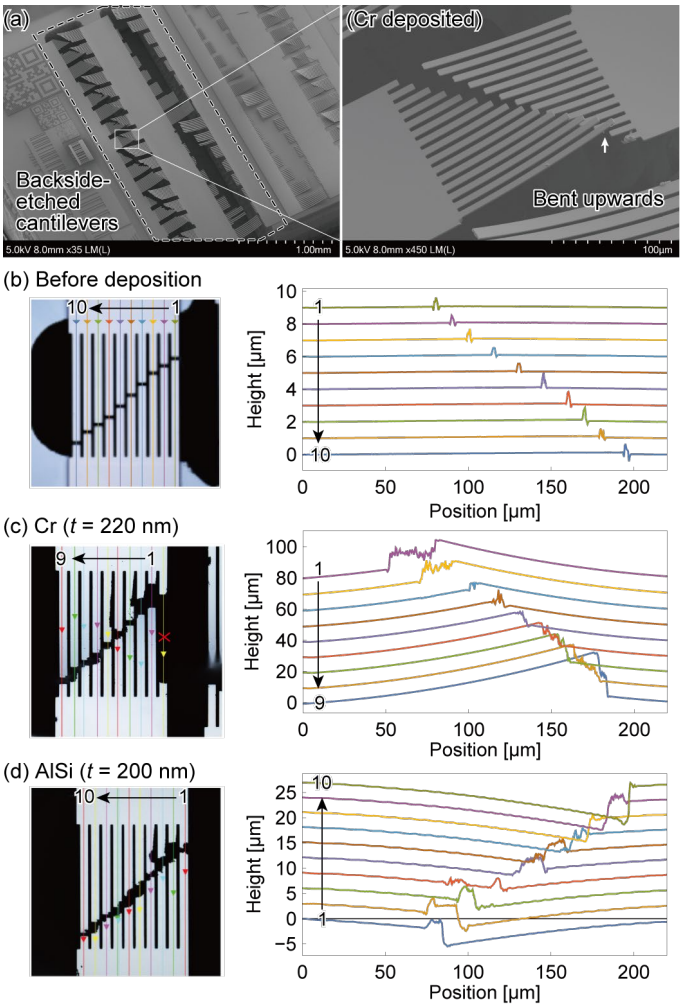


Fig. 5. Stress measurement. (a) SEM image after Cr deposition by 220 nm, (b) height profile before deposition, (c) after Cr deposition, and (d) after AlSi deposition. (b)-(d) are measured using 10μm-wide cantilevers with 170 μm anchor-to-anchor separation.

initial state before deposition had negligible deformation (Fig. 5(b)), proving that these deformations are the result of sputtering deposition. As in Fig. 5(c) and (d), the deformation was concave for Cr and convex for AlSi. Fitting a circle to 100 μm or longer profiles yielded a curvature of $1/R = 1.32 \pm 0.01$ mm^{-1} and -0.25 ± 0.02 mm^{-1} for Cr and AlSi, respectively. Using Atkinson's modified form of Stoney's formula [5]

$$\sigma = \frac{E_s h_s^3}{6h_f (h_s + h_f)(1-\nu_s)R}, \qquad (2)$$

we obtain σ_{Cr} = 652 MPa and σ_{AlSi} = −140 MPa, which are reasonable values compared with the ref [8]. In this measurement, the shortest cantilevers (170 μm anchor-to-anchor distance) were used, and longer cantilevers can detect smaller deformation covering the range of lower stress. With typical PVD film stress in the range of 1 GPa or less, the set of cantilevers prepared in this design should be sufficient for versatile films deposited in the university cleanroom.

The readability of the QR Code ® was also tested. The patterned codes displayed on a monitor were readable by

Fig. 6. Patterns coding the dummy ID number "123456".

smartphone applications, down to the smallest 224 μm pattern, whose grid width in design was 670 nm (Fig. 6). Although the pattern was too small to be completely resolved by photolithography, excessive deep RIE (condition for 5 μm etching applied to 2 μm device layer) etched the surface to become less reflective. For the barcodes, the result did not have a large contrast between the pattern and its outside, making it difficult for a decoder to read it directly. With its robustness against noise and readable size, QR Code ® should be a preferable tool to encode IDs on the chip.

IV. Conclusion

To facilitate a set of comprehensive process recipes in university cleanrooms, we proposed "chamber sharing system", where small test chips are loaded in the process chamber together with the user's samples and collected and analyzed by the platform staff. To realize such a test chip, we designed and demonstrated a numbered 8 × 8 mm2 test chip for PVD, capable of thickness, stress, and conductivity measurements. By preparing such a test chip for all process types, open platforms like university cleanrooms can develop the processes without taking up time for users.

Acknowledgment

The electron beam lithography was performed using EB writer F5112+VD01 donated by Advantest Corporation. The authors would like to thank the support team of Takeda Super Cleanroom, The University of Tokyo.

References

[1] Y. Mita, et al., "Opportunities of "CMOS-MEMS integration through LSI foundry and open facility," Jpn. J. Appl. Phys., vol. 56, article no. 06GA03, May 2017.

[2] N. Usami, E. Ota, A. Higo, T. Momose, and Y. Mita, "Drop-in test structure to evaluate residual stress in conformally grown films," IEEE Trans. Semicond. Manuf. Vol. 34, pp. 270-277, August 2021.

[3] S. Dutta, and A. Pandey, "Overview of residual stress in MEMS structures: Its origin, measurment, and control," J. Mater. Sci: Mater Electron, vol. 32, pp. 6705-6741, 2021.

[4] G.C.A.M. Janssen, M.M. Abdalla, F. van Keulen, B.R. Pujada, and B. van Venrooy, "Celebrating the 100th anniversary of the Stoney equation for film stress: Developments from polycrystalline steel to single crystal silicon wafers," Thin Solid Films, vol. 517, pp. 1858-1867, 2009.

[5] C.A. Klein, "How accurate are Stoney's equation and recent modifications," J. Appl. Phys., vol. 88, pp. 5487-5489, 2000.

[6] S. Enderling, et al., "Sheet resistance measurement of non-standard cleanroom materials using suspended greek cross test structures," IEEE Trans. Semicond. Manuf. Vol. 19, pp. 2-9, February 2006.

[7] I. Sari, I. Zeimpekis, and M. Kraft, "A dicing free SOI process for MEMS devices," Microelectron. Eng., vol. 95, pp. 121-129, 2012.

[8] R.M.R. Pinto, V. Chu, and J.P. Conde, "Amorphous silicon self-rolling micro electromechanical systems: From residual stress control to complex 3D structures," Adv. Eng. Mater., vol. 21, article no. 1900663, 2019.

Electrical behavior of ALD-molybdenum films in the thin-film limit

Kees van der Zouw, Simone D. Dulfer, Antonius A.I. Aarnink, and Alexey Y. Kovalgin
MESA+ Institute for Nanotechnology
University of Twente, 7500 AE Enschede, The Netherlands
Email: k.vanderzouw@utwente.nl

***Abstract*—Test structures were designed and fabricated to investigate the electrical properties of ultra-thin molybdenum films obtained by atomic layer deposition. The films were incorporated in conventional Van der Pauw and circular transmission line method test structures to explore the effect of film thickness on the resistivity, temperature coefficient of resistance, contact resistivity, and external electric field applied. The resistivity was shown to depend strongly on film thickness, while the temperature coefficient of resistance changed from positive to negative, indicating a change in the dominant conduction mechanism. A modest field effect was observed for the films in their thickness limit.**

I. Introduction

Atomic layer deposition (ALD) is a thin film deposition technique that has become indispensable for semiconductor fabrication schemes [1-3]. Its attributes of sequential self-limiting surface reactions allow for thin film formation with excellent thickness control, uniformity, and conformality [4-6]. Due to continuous miniaturization, interconnect RC delay and power consumption have increased and quickly become significant determinants of chip performance [7-9]. Solutions to this so-called interconnect bottleneck are sought in novel metal ALD processes [10-14].

Molybdenum (Mo), a refractory metal with a low coefficient of thermal expansion, high electrical conductivity and melting point, and relatively short electron mean free path, could be considered an excellent alternative to replace copper as the interconnect in IC logic and tungsten in 3D NAND memory devices [15-21]. However, a reliable ALD process for enabling metallic Mo films within the CMOS back-end-of-line (BEOL) temperature limits was still missing [21-24]. Recently, Mo was successfully deposited from the molybdenum oxychloride ($MoCl_2O_2$) precursor and atomic hydrogen (at-H) co-reactant in a radical-enhanced ALD process between 150 and 450 °C, well within the BEOL limits [25].

Several properties of the Mo ALD films have already been investigated, like their chemical composition, large-scale thickness uniformity, and conformality in high-aspect-ratio (HAR) structures [25]. However, a detailed analysis of the electrical characteristics of these thin films as a function of film thickness has not yet been conducted but is warranted as metallic films are generally used for their conductive properties, and confinement effects may play an essential role at the length scales for which ALD is a suitable deposition technique.

In this work, electrical test structures were designed and fabricated to characterize various electrical properties of the Mo films while approaching the thin-film thickness limit (i.e., <10 nm). Among others, four-point collinear probe, Van der Pauw, and (circular) transmission line method structures were employed, allowing for the determination of sheet and contact resistance, transfer length, temperature coefficient of resistance, and a possible field effect. In addition to evaluating the suitability of ALD Mo for application in metal interconnects, the measured electrical properties may contribute to understanding the early film-growth stages of the ALD process.

II. Experimental

Electrical test structures for Mo ALD film characterization were designed, fabricated, and investigated. A process flow was developed with two lithography steps to keep the fabrication procedure concise and prevent excessive processing on a vulnerable ALD layer. The first lithography step was used to define electrodes, while the second was used to pattern the Mo ALD film. The result is an inverted design, where the functional layer buries the electrodes. This does not pose a problem for probing due to the extreme thinness of the ALD film. A schematic representation of the fabrication process is provided in Fig. 1. Highly doped p-type 100 mm (100) Si wafers (0.01-0.025 Ω/cm), with 100 nm thermally grown SiO_2, were used as substrate. The photoresist was applied and patterned according to the desired electrode shapes. The 10 nm thick titanium adhesion and 60 nm thick platinum (Pt) electrode layers were sputter-deposited before a lift-off step defined the Pt electrodes. The Mo ALD film was grown in a home-built high vacuum ALD system with a low-volume reaction chamber. A base pressure of ~10^{-8} mbar was obtained using a turbopump, whereas an operating pressure of ~10^{-3} mbar was obtained with an 8 SCCM N_2 carrier and purge gas flow. To ensure sufficient vapor pressure, the vessel of the solid $MoCl_2O_2$ precursor was heated to 75 °C. Reactant condensation (in the gas lines, etc.) was prevented by heating the reactor's corresponding gas-inlet system and outer chamber to 85 and 100 °C, respectively. The inner reaction chamber was kept at 350 °C during the experiments. An ALD process cycle was developed, consisting of 1 s $MoCl_2O_2$ and 29 s at-H dosing times, separated by 15 s N_2 purges. This ALD cycle supplied a saturation dose of the reactants while preventing a parasitic CVD component due to insufficient purging times [25]. The ALD cycle was repeated until the desired Mo film thickness was obtained. To prevent native oxidation of Mo in air, an approx. 10 nm thick amorphous silicon (aSi) capping layer was grown by the pyrolysis of

979-8-3503-2987-2/24/$31.00 © 2024 IEEE

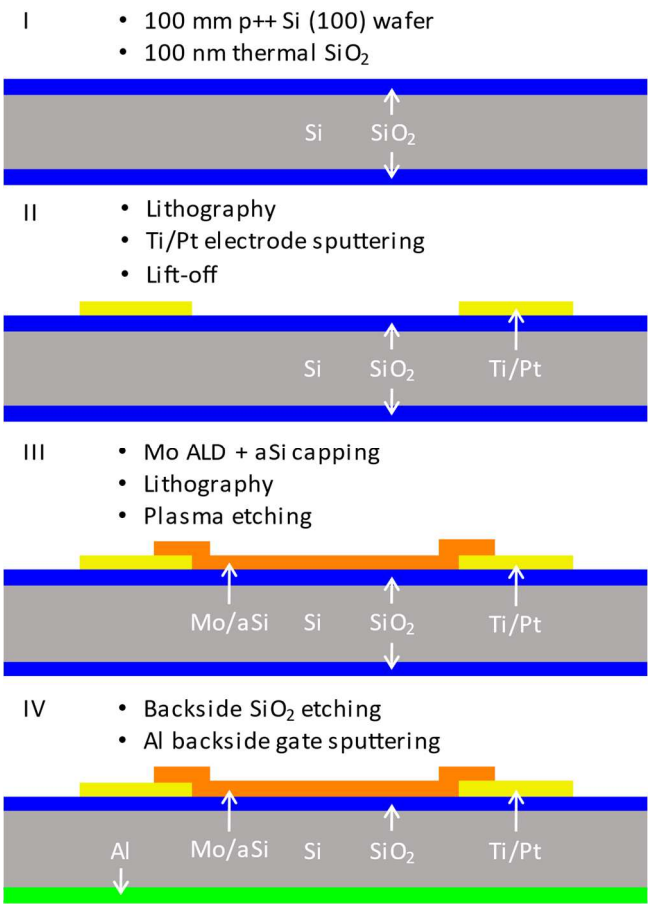


Figure 1 – Schematic of the process flow to fabricate the test structures used in this work, with: I) Isolating layer (i.e., gate dielectric) formation, II) Pt electrode formation, III) Mo ALD film formation, and IV) Backside gate electrode formation.

trisilane (Si_3H_8) at 375 °C without a vacuum break. At his temperature, reactive silicide formation should be limited [26]. Patterning of the Mo ALD film was achieved by a second photolithography step, after which the Mo and aSi layers were locally etched away by physical dry-etching to obtain the required pattern. A backside Al contact was manufactured to use the highly doped Si substrate as a gate electrode. For this, a new photoresist layer was applied to the front surface. The thermal SiO_2 on the backside of the wafer was subsequently removed in buffered hydrogen fluoride (HF), and a 400 nm thick aluminum gate electrode was sputter-deposited to finalize the test structures.

Each 100 mm wafer is facilitated with a central square (1×1 cm^2) for in situ film thickness monitoring by spectroscopic ellipsometry (SE). Located on the left and right sides of the wafers are the alignment markers and several structures (e.g., Vernier) to monitor processing accuracy (see Fig. 2a). A 1x1 cm^2 process evaluation module (PEM) consisting of various test structures was designed, and repeatedly fabricated across the wafer (see Fig. 2b). At the center of the PEM, a 4 mm wide strip of the Mo ALD film was realized for additional characterization purposes (e.g., SE and X-ray diffraction). The PEM further contained four-point colinear probe, Van der Pauw, (circular)

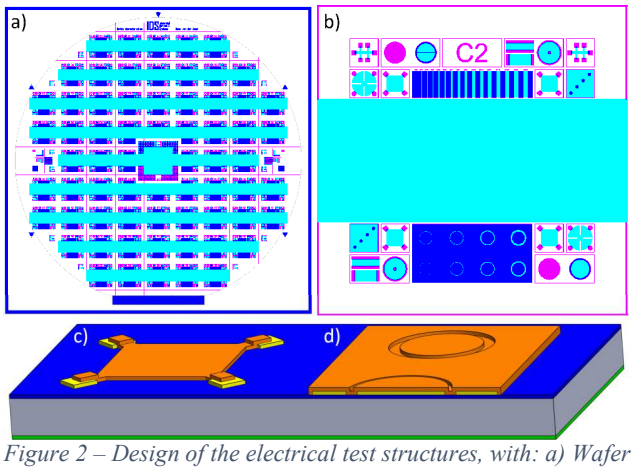


Figure 2 – Design of the electrical test structures, with: a) Wafer scale mask layout, b) Process evaluation module design, c) A three-dimensional representation of the Van der Pauw test structure, and d) A three-dimensional representation of the CTLM test structure.

transmission line method [(C)TLM], Greek-cross, Hall, and CV profiling structures. For each type of test structure, several practical design considerations were made. For instance, the Van der Pauw equation for the sheet resistance assumes a structure with infinitesimally small contacts at the circumference of the sample [27,28]. This condition can hardly be met in practice. This is why the design of the Van der Pauw structures used in this work (see Fig. 2c) is based on analytical work by Versnel, as the Van der Pauw equation should still hold with high accuracy for an octagonal design if the ratio between the contact length and the total circumference of the structure is kept below 0.1 [29]. A minor adaptation has been implemented by adding short arms that connect the electrodes with the active layer and should prevent alignment issues during fabrication from affecting the measurement results. Analysis of the CTLM structures (see Fig. 2d) can be simplified significantly by choosing a radius for the inner electrode that is much larger than the gap width to the outer electrode and four times the transfer length [30,31]. While the gap width is chosen during the design process of the test structures, the transfer length is unknown beforehand, as the distance over which most of the current transfers between the contact and the active layer depends on the specific contact resistivity and sheet resistance. This work assumed a suitable inner electrode radius of 200 µm for the CTLM structures.

Film growth was monitored in situ with a J. A. Woollam Co. M-2000 spectroscopic ellipsometer operating in the 245-1690 nm spectral range. The CompleteEASE v6.46 software package was used to operate the SE and model the layer stack characteristics to extract the Mo thin film properties. The optical properties of thin Mo films were modeled with a Drude and two Lorentz oscillators. The Drude term describes the intraband electron transitions, i.e., the existence of conduction electrons, while the Lorentz oscillators account for the interband transitions [32]. The accuracy of SE thickness determination was verified by alternative characterization techniques (e.g., SEM and TEM) for films exceeding a thickness of 5 nm [25]. Below 5 nm film thickness, the otherwise passive oscillators of the SE model started to shift in energy, indicating a change in the film's

optical properties as a function of thickness. This meant that an accurate determination of the thickness for films thinner than 5 nm would require verification by other characterization techniques, each with its own limitations. For all subsequent calculations, the SE model was still assumed to be valid for the films below 5 nm in thickness. Six wafers were fabricated with different thicknesses of the Mo ALD film: 1.5, 1.8, 2.5, 3.8, 4.8, and 8.9 nm. Additionally, a reference wafer without a Mo film was made.

Electrical measurements were conducted using an MPI TS300-SE Probe System with a high-performance local environmental chamber for ultra-low noise measurements. The system was equipped with four guarded tungsten Kelvin triax probes with external preamplifiers and a Kelvin triax gold-plated nickel thermal chuck, operating in a temperature range between 20 and 200 °C. A Keithley 4200A-SCS-PKB parameter analyzer with Clarius GUI-based software was used to operate the probe station and analyze the measurement results.

III. Results and Discussion

A. Resistivity

The electrical resistivity of the Mo ALD films was determined through Van der Pauw sheet resistance and SE thickness measurements. Under the assumption that the resistivity is independent of depth in the film, it can be determined by multiplying the sheet resistance by the thickness. Figure 3 shows the electrical resistivity as a function of film thickness. There is a clear trend of increasing resistivity with decreasing film thickness. This can be explained by various effects dominating the nanometer-thickness range. Specifically, the electron mean free path (MFP) in pristine Mo is 11.2 nm [33], which means that electron scattering at surfaces and interfaces becomes more pronounced as soon as the thickness of the layer becomes comparable to the MFP [34-37]. The type of film growth mode can also determine the resistivity during the early stages of ALD. A film formed by the Volmer-Weber growth mode (i.e., island formation) may exhibit a higher resistivity when compared to a film grown in the Frank van der Merwe growth mode (i.e., layer-by-layer growth) for the same thickness. This is due to an increase in the conduction path if film islands have yet to reach coalescence across the entire substrate area to form a continuous film.

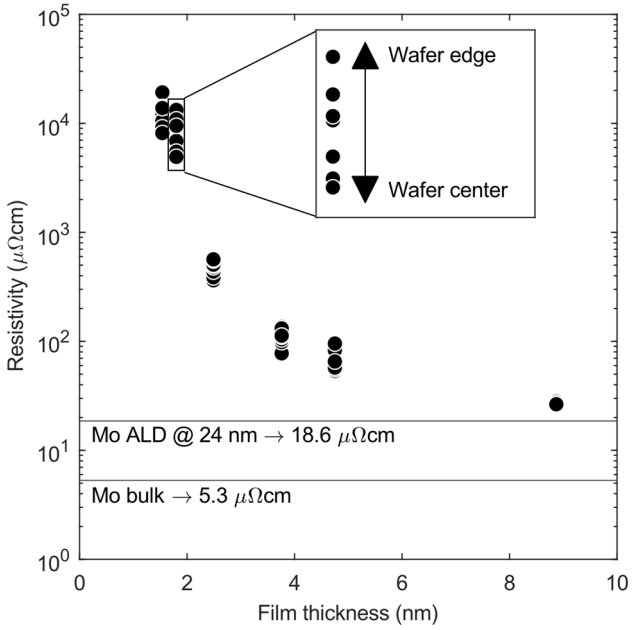


Figure 3 – Electrical resistivity of the Mo ALD films as a function of the film thickness. The resistivity of bulk Mo and a 24 nm thick Mo ALD film are shown as a reference. The inset indicates the resistivity distribution depending on the measurement location on the wafer.

The lowest resistivity of this work was measured at 27 μΩcm for the 8.9 nm Mo ALD film, which is approx. five times higher than the theoretical limit of 5.3 μΩcm for pristine single crystalline Mo [38]. As the resistivity is still thickness-dependent below 10 nm, 27 μΩcm does not represent the lowest resistivity value expected from this specific ALD process. Our earlier work obtained a resistivity value of 18.6 μΩcm from a 24 nm thick Mo ALD film [25]. Mo films formed by other ALD processes show similar resistivities; however, those films were obtained at higher processing temperatures: 13 μΩcm (at 600 – 650 °C) [24], 15 μΩcm (at 400 – 500 °C) [22], 16 μΩcm (at 450 – 590 °C) [21], and 124 μΩcm (at 120 °C) [23]. This work demonstrates the lowest resistivity obtained in a Mo ALD process at a temperature as low as 350 °C.

The measured resistivity shows a distribution across the wafer, an example of which is given in the inset of Fig. 3. The resistivity generally increases towards the edge of the wafer, being the lowest in the center. This can be explained by the surface temperature non-uniformity, i.e., the temperature decreasing towards the edge. Alternatively, a higher resistance can occur due to the increasing recombination of at-H towards the wafer edge (because of collisions), limiting its supply and, thus, the reaction kinetics. The lifetime of at-H in the reaction chamber is limited. It has been shown previously that our Mo ALD process did not occur at 350 °C if at-H was replaced by H_2. Indeed, film compositional analysis performed by X-ray photoelectron spectroscopy (XPS) revealed an increasing oxygen share in the films towards their edge. Further, X-ray reflectivity (XRR) measurements confirmed a lower film density at the edge, consistent with higher oxygen share and resistivity [25].

B. Temperature coefficient of resistance

The temperature coefficient of resistance (TCR) was determined by Van der Pauw sheet resistance measurements at temperatures ranging from 25 to 200 °C. The TCR is defined as the relative change of the resistance corresponding to a given change in temperature. The resistance at 25 °C was taken as the reference value. Figure 4a depicts the sheet resistance as a function of temperature for two samples with differently thick Mo films: 2.5 and 8.9 nm. It can be seen that the resistance increases with increasing temperature for the 8.9 nm thick Mo ALD film, suggesting a positive TCR. This is typical for metals, as electron-phonon scattering is the primary mechanism responsible for the electrical resistivity in bulk metal, and it increases at elevated temperatures [39]. In contrast, the TCR of

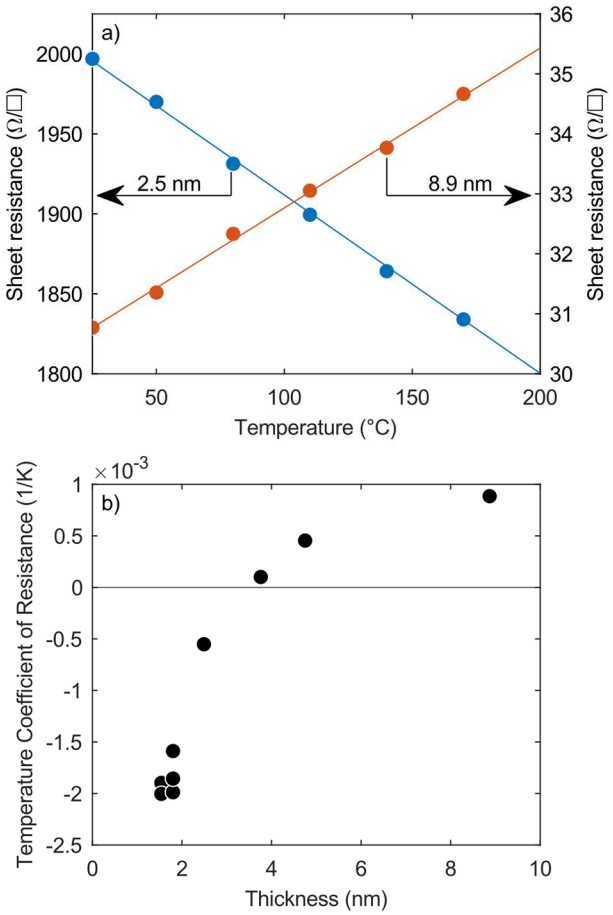


Figure 4 – a) The influence of temperature on the sheet resistance of Mo ALD films of 2.5 and 8.9 nm thick, and b) The TCR as a function of film thickness.

the 2.5 nm film is negative, which is typical for semiconductors as their conduction relies on charge carrier generation, which is enhanced at higher temperatures [40]. The transition from negative to positive TCR in Mo ALD films, depending on their thickness, is further demonstrated in Fig. 4b. It is explained by the non-trivial dependence of film thickness and temperature on the mobility and concentration of the charge carriers. Even for the 8.9 nm film with positive TCR, its value is significantly smaller than the bulk value expected for Mo ($5.2 \cdot 10^{-3}$ K^{-1}). When the film thickness reduces such that the film stays only partially percolated, the conduction mechanism can change entirely, e.g., to a hopping conduction between discrete metallic islands. This conduction type can be described by introducing additional energy barriers for the current to flow, with activation energy in the corresponding Arrhenius expression. A higher T can then promote charge transport between islands that are in contact, resulting in a negative TCR.

C. *Contact resistivity and transfer length*

CTLM structures were utilized to investigate the contact resistance and transfer length between the Pt electrodes and the Mo ALD thin film. The total resistance between the inner and outer electrodes was determined for different gap widths. After applying a correction factor to account for the circular geometry of the device, the contact resistance and transfer length were extracted from linear extrapolation of the resistance as a function of the gap width. The specific contact resistance was then determined by multiplying the contact resistance with the contact surface area, specifically the area over which the current is transferred between the electrode and the film, the size of which depends on the transfer length. All resistance measurements resulted in linear current-voltage curves, indicating Ohmic contacts. Fig. 5a shows that the contact resistance increases with decreasing film thickness, which can be explained by increased confinement effects and current crowding. The type of the growth mode (Volmer-Weber or Frank van der Merwe) can also play a role, as an island-type growth could prevent proper current transfer between the

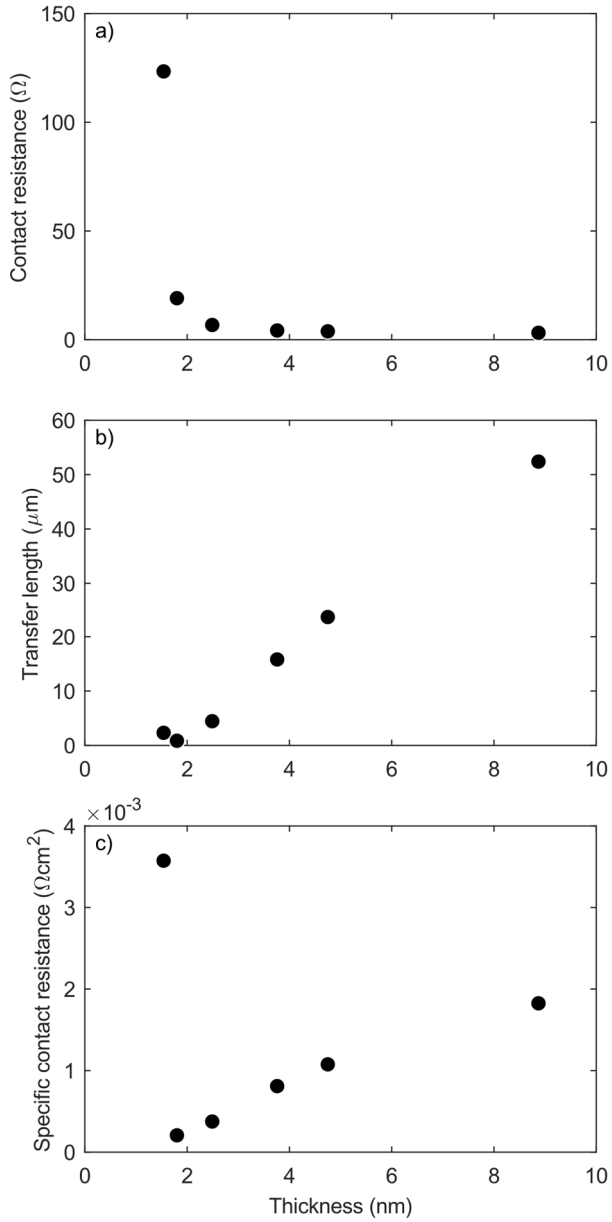


Figure 5 – The a) contact resistance, b) transfer length, and c) specific contact resistance as a function of Mo ALD film thickness.

electrodes and the film. The decrease in transfer length with decreasing film thickness (see Fig. 5b) is explained by the corresponding increase in resistivity of the Mo film. As the relative difference in resistivity between the Pt electrodes and the film becomes more significant, the length scale over which most of the current is transferred becomes smaller. The current will choose the path of the least resistance, only entering the Mo ALD film at the outer edge of the electrode. Interestingly, the specific contact resistance shows a minimum as a function of film thickness, as depicted in Fig. 5c. Its increase in the thin-film limit can be explained by the sharp increase in contact resistance. The gradual increase in the specific contact resistance for thicker films is attributed to the rapid increase of the transfer length, which is expected to be caused by the limited difference in sheet resistance between the Pt electrodes and the Mo thin film. The sheet resistance of the Pt contacts cannot be neglected anymore, resulting in an attenuated transfer length and an inability to accurately determine the specific contact resistance [41-43].

D. Electric field effect

The electric field effect in the Mo ALD thin films was investigated using Van der Pauw structures. The field effect is defined as the relative change of the sheet resistance caused by a given change of the gate voltage (applied to the substrate) while keeping the source-drain bias constant. Generally, metals do not exhibit an electric field effect due to the small penetration depth of an external electric field being essentially confined to a few first atomic layers. However, gate bias variation can still modulate sheet resistance of the films in their thickness limit, i.e., when the electron concentration is lower. Figure 6a shows the sheet resistance as a function of the gate-source bias for a Van der Pauw structure having a 1.5 nm Mo ALD layer. The sheet resistance is slightly lowered if a negative gate voltage is applied, whereas it is somewhat increased for a positive gate voltage. Figure 6b shows the magnitude of the field effect as a function of film thickness. Mo ALD films of ≥4.8 nm do not exhibit an observable field effect. The field effect becomes more pronounced with decreasing film thickness, still being relatively small. The maximum field effect observed is approx. 0.01% per volt of gate bias for the thinnest 1.5 nm Mo film.

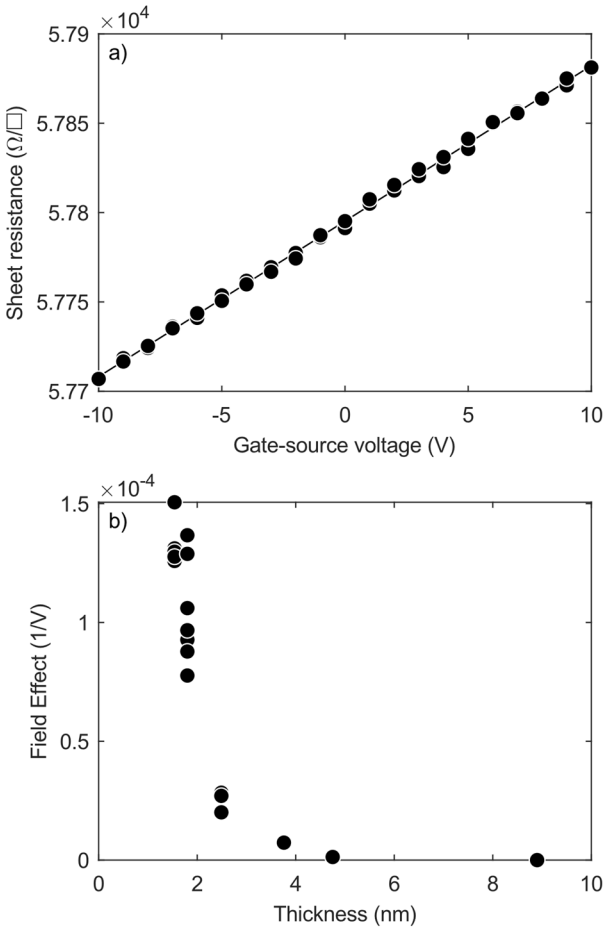


Figure 6 – The electric field effect in Mo ALD films, with a) change of the sheet resistance as a function of applied gate voltage, and b) magnitude of the field effect as a function of film thickness.

IV. Conclusion

The electrical performance of ultra-thin molybdenum films (1.5 – 9 nm thick) obtained by ALD at process conditions suitable for BEOL CMOS processing has been evaluated. The resistivity, TCR, contact resistivity, transfer length, and field effect have been investigated. The resistivity of 18.6 μΩcm was comparable to that of other Mo ALD processes but was obtained at a lower temperature. The resistivity increased rapidly with decreasing film thickness due to various effects dominating the nanometer-thickness range. The magnitude and sign of the TCR exhibited a clear dependence on the film thickness. Thick (i.e., ≥ 3.8 nm) films exhibited a positive TCR, which is expected for metals. Thin (i.e., ≤ 2.5 nm) films showed a negative TCR, indicating a non-trivial interplay between carrier concentration and mobility with changing film thickness. The contact resistance between the Pt electrodes and Mo films decayed rapidly with increasing film thickness. At the same time, an electric field effect was not observed for Mo films thicker than 3.8 nm while increasing progressively for films of 2.5 nm and less.

Acknowledgment

The authors thank H. Veltkamp for his help with the fabrication of the electrical test structures. The authors thank R.A.M. Wolters, D.J. Gravesteijn, and J.H. Klootwijk for fruitful discussions.

References

[1] M. Schumacher, P.K. Baumann, and T. Seidel, "AVD and ALD as two complementary technology solutions for next generation dielectric and conductive thin-film processing," Chemical Vapor Deposition, vol. 12, pp. 99-108, March 2006.

[2] R.D. Clark, "Emerging applications for high k materials in VLSI technology," Materials, vol. 7, pp. 2913-2944, April 2014.

[3] C. Zhao and J. Xiang, "Atomic layer deposition (ALD) of metal gates for CMOS," Applied Sciences, vol. 9, pp. 2388, June 2019.

[4] R.L. Puurunen, "Growth per cycle in atomic layer deposition: a theoretical model," Chemical Vapor Deposition, vol. 9, pp. 249-257, October 2003.

[5] S.M. George, "Atomic layer deposition: an overview," Chemical Reviews, vol. 110, pp. 111-131, January 2010.

[6] V. Cremers, R.L. Puurunen, and J. Dendooven, "Conformality in atomic layer deposition: current status overview of analysis and modelling," Applied Physics Reviews, vol. 6, pp. 021302, June 2019.

[7] P. Kapur, J.P. McVittie, and K.C. Saraswat, "Technology and reliability constrained future copper interconnects – part I: resistance modeling," IEEE Transactions on Electron Devices, vol. 49, pp. 590-597, April 2002.

[8] M.R. Baklanov, C. Adelmann, L. Zhao, and S. De Gendt, "Advanced interconnects: materials, processing, and reliability," ECS Journal of Solid State Science and Technology, vol. 4, pp. Y1-Y4, December 2014.

[9] Zs. Tőkei et al., "Inflection points in interconnect research and trends for 2nm and beyond in order to solve the RC bottleneck," International Electron Devices Meeting San Francisco, CA, pp. 5.3.1-5.3.4, December 2020.

[10] C. Adelmann et al., "Alternative metals for advanced interconnects," IEEE International Interconnect Technology Conference San Jose, CA, pp. 173-176, May 2014.

[11] K. Sankaran, S. Clima, M. Mees, and G. Pourtois, "Exploring alternative metals to Cu and W for interconnects applications using automated first-principles simulations," ECS Journal of Solid State Science and Technology, vol. 4, pp. N3127-N3133, 2015.

[12] C. Adelmann et al., "Alternative metals: from ab initio screening to calibrated narrow line models," IEEE International Interconnect Technology Conference Santa Clara, CA, pp. 154-156, June 2018

[13] D. Gall, "The resistivity bottleneck: the search for new interconnect metals," International Symposium on VLSI Technology Systems Hsinchu, Taiwan, pp. 112-113, August 2020.

[14] D. Gall, "The search for the most conductive metal for narrow interconnect lines," Journal of Applied Physics, vol. 127, pp. 050901, February 2020.

[15] A. Gupta et al., "Buried power rail metal exploration towards the 1 nm node," IEEE International Electron Devices Meeting San Francisco, CA, pp. 22.5.1-22.5.4, December 2021.

[16] V. Founta et al., "Properties of ultrathin molybdenum films for interconnect applications," Materialia, vol. 24, pp. 101511, July 2022.

[17] A. Ajaykumar et al., "First demonstration of ruthenium and molybdenum word lines integrated into 40nm pitch 3D-NAND memory devices," Symposium on VLSI Technology Kyoto, Japan, pp. 1-2, June 2021.

[18] H. Köstenbauer et al., "Resistance reduction of Mo metallization by W seed layer," SID Symposium Digest of Technical Papers, vol. 51, pp. 200-203, August 2020.

[19] P. Kreiml et al., "Electro-mechanical behavior of Al/Mo bilayers studied with *in situ* straining methods," Thin Solid Films, vol. 665, pp. 131-136, November 2018.

[20] D. Tierno et al., "Reliability of barrierless PVD Mo", IEEE International Interconnect Technology Confrerence Kyoto, Japan, p. 1, July 2021.

[21] E. Zoubenko et al., "Impact of chemical bonding difference of ALD Mo on SiO_2 and Al_2O_3 on the effective work function of the two gate stacks," Journal of Vacuum Science and Technology A, vol. 39, pp. 043201, May 2021.

[22] M. Juppo, M. Vehkamäki, M. Ritala, and M. Leskelä, "Deposition of molybdenum thin films by an alternate supply of $MoCl_5$ and Zn," Journal of Vacuum Science and Technology A, vol. 16, pp. 2845-2850, September 1998.

[23] D. Seghete, G.B. Rayner, Jr., A.S. Cavanagh, V.R. Anderson, and S.M. George, "Molybdenum atomic layer deposition using MoF6 and Si2H6 as the reactants," Chemistry of Materials, vol. 23, pp.1668-1678, April 2011.

[24] B.-J. Lee, K.-B. Lee, M.-H. Cheon, D.-W. Seo, and J.-W. Choi, "Study on the deposition characteristics of molybdenum thin films deposited by the thermal atomic layer deposition method using MoO_2Cl_2 as a precursor," Coatings, vol. 13, pp. 1070, June 2023.

[25] Van der Zouw et al., "Low-resistivity molybdenum obtained by atomic layer deposition", Journal of Vacuum Science and Technology A, vol. 41, pp. 052402, September 2023.

[26] C.M. Doland and R.J. Nemanich, "Phase formation during reactive molybdenum-silicide formation," Journal of Materials Research, vol. 5, pp. 2854-2864, December 1990.

[27] L.J. van der Pauw, "A method of measuring specific resistivity and hall effect of discs of arbitrary shape," Philips Research Reports, vol. 13, pp. 1-9, February 1958.

[28] L.J. van der Pauw, "A method of measuring the resistivity and Hall coefficient on lamellae of arbitrary shape," Philips Technical Review, vol. 20, pp. 220-224, 1958.

[29] W. Versnel, "Analysis of symmetrical Van der Pauw structures with finite contacts," Solid-State Electronics, vol. 21, pp. 1261-1268, 1978.

[30] A.J. Willis and A.P. Botha, "Investigation of ring structures for metal-semiconductor contact resistance determination," Thin Solid Films, vol. 146, pp. 15-20, 1987.

[31] J.H. Klootwijk and C.E. Timmering, "Merits and limitations of circular TLM structures for contact resistance determination for novel III-V HBTs," IEEE International Conference on Microelectronic Test Structures, vol. 17, pp. 247-252, 2004.

[32] H. Fujiwara, "Spectroscopic Ellipsometry: Principles and Applications," (Wiley, New York, 2007).

[33] D. Gall, "Electron mean free path in elemental metals," Journal of Applied Physics, vol. 119, pp. 085101, 2016.

[34] K. Fuchs, "The conductivity of thin metallic films according to the electron theory of metals," Mathematical Proceedings of the Cambridge Philosophical Society, vol. 34, pp. 100-108, 1938.

[35] E.H. Sondheimer, "The mean free path of electrons in metals," Advances in Physics, vol. 1, pp. 1-42, 1952.

[36] A.F. Mayadas, M. Shatzkes, and J.F. Janak, "Electrical resistivity model for polycrystalline films: the case of specular reflection at external surfaces," Applied Physics Letters, vol. 14, pp. 345-347, 1969.

[37] A.F. Mayadas and M. Shatzkes, "Electrical-resistivity model for polycrystalline films: the case of arbitrary reflection at external surfaces," Physical Review B, vol. 1, pp. 1382-1389, 1970.

[38] P.D. Desai, T.K. Chu, H.M. James, and C.Y. Ho, "Electrical resistivity of selected elements," Journal of Physical and Chemical Reference Data, vol. 13, pp. 1069-1096, October 1984.

[39] F. Bassani, G.L Liedl, and P. Wyder, "Encyclopedia of condensed matter physics," (Academic Press, Cambridge, Massachusetts 2005).

[40] S.M. Sze and K.K. Ng, "Physics of Semiconductor Devices," (John Wiley & Sons, Inc., Hoboken, New Jersey 2007)

[41] D.B. Scott, W.R. Hunter, and H. Shichijo, "A transmission line model for silicided diffusions: impact on the performance of VLSI circuits," IEEE Transactions on Electron Devices, vol. 29, pp. 651-661, April 1982.

[42] G.K. Reeves and H.B. Harrison, "Contact resistance of polysilicon-silicon interconnections," Electronics Letters, vol. 18, pp. 1083-1085, December 1982.

[43] G.K. Reeves and H.B. Harrison, "Determination of contact parameters of interconnecting layers in VLSI circuits," IEEE Transactions on Electron Devices, vol. 33, pp. 328-334, March 1986.

Test Structure to Assess Bump Shape Influence on Hybrid Bonding

Ayako Mizushima[1], Kei Misumi[1], Shun Yasunaga[1], Akio Higo[1], Ryosho Nakane[1], Kazumichi Tsumura[2], Kazuyuki Higashi[2], Yukinori Ochiai[1], and Yoshio Mita[1]

[1]*Graduate School of Engineering, The University of Tokyo,* Tokyo, Japan

[2]*TOSHIBA CORPORATION,* Kawasaki, Japan

***Abstract*— Bump bonding is widely used in CMOS-MEMS integration. As compared to the other technology (planar bonding), bump bonding is more interesting for its simplicity and applicability to wide range of device types. Towards higher density and reliability improvement, the authors have designed and fabricated a test structure. Our new contribution is to intentionally structuralize shapes of bumps to not only a flat head but concave as well as convex ones, expecting better occlusion and passive alignment. To experiment our new idea, the first test structure has been fabricated to assess alignment accuracy.**

Keywords— bump shape, flip chip bonding, test structure

I. Introduction

Bonding is an attractive technology to quickly realize a CMOS-MEMS system while avoiding all potential process incompatibilities [1] [2]. The technology is simple in principle: align and connect corresponding metallic parts. Au (gold) bump is one of the preferred component to make connections. In analogy to the Cu/SiO_2 bumpless bonding, the surface of Au bump is often planarized; however, protruded flat bump heads sometimes slip each other and it can produce an alignment error. Also, Au cannot chemical-mechanical polished (CMP) so it is not so evident to obtain chip-scale (as well as wafer-scale) uniformity of bump height. Non-uniformity of bump height will be the cause of areal poor bonding (open failure, Fig.1). To solve above problems, the authors are investigating to intentionally structuralize the bump shape. For example, if we can reliably produce concave and convex bump shapes, and face-to-face align them, they may nicely occlude at the registration (analogy to the occlusion of teeth), thereby will provide higher connection density and reliability (Fig. 2).

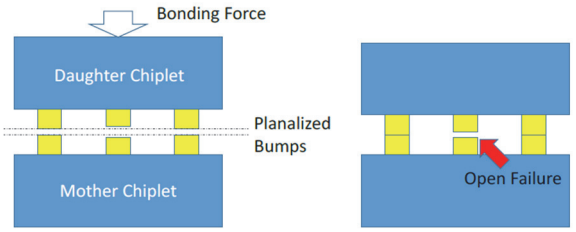


Fig.1 Non-uniformity of bump planarization induces open failure.

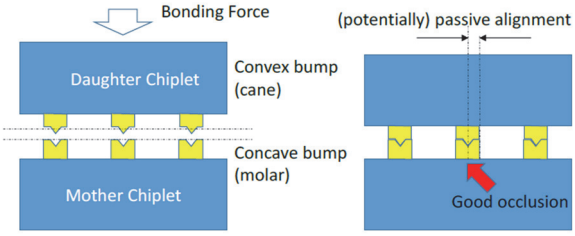


Fig.2 Our proposal: If we can control surface flatness to convex and concave we can improve bonding accuracy and reliability

II. Test Structure Discription

A. Bump shape test designs

The test structure (Fig.3) includes a serpentine daisy-chain wirings with a couple of voltage taps. On the chip, Vernier structure is introduced in order to optically observe the alignment error. The test structure is used for chip-level bonding. Intentional surface flatness modulation is designed by either opening or embedding micro structure in the SiO_2, which is supposed to be an isolation layer of CMOS / MEMS. As of today, five designs of three different types have been tested (Fig.5): (TD1) 5×5 μm^2 open square, (TD2) 3μm right angled triangle and (TD3) 3×3 μm^2 square. TD1 is intended to be a flat bump, and TD2 and TD3 are intended to be convex and concave bumps. Our intension has been to obtain concave bump by smaller SiO_2 openings (TD2-N and TD3-N) and convex one is by leaving protruding triangular (TD2-P) or smaller square (TD3-P) SiO_2 at the center of the open square.

B. Device fabrication

A test structure was fabricated on 4" fused silica wafer. For shorter process step, we applied a simplified process with bottom-up electroplating on flat Cr (50 nm) /Au (300 nm) seed layer (Fig.5). The seed layer and isolation of Cr (25 nm) / SiO_2 (1 μm) was batch sputtered. Twenty-five nanometer Cr worked as an adhesion layer for SiO_2. By using chemical amplification electron beam resist (OEBR-CAP112), openings were directly drawn by rapid electron beam writer (Advantest F5112+VD01), then was etched by high-density ICP plasma with CHF_3 gas. After cleaning, resist mold was made with negative-tone resist photolithography (ZPN-1150, 4 μm). Top Cr was removed and electroplating was made (Yamamoto-MS). Electrical routing

979-8-3503-2987-2/24/$31.00 ©

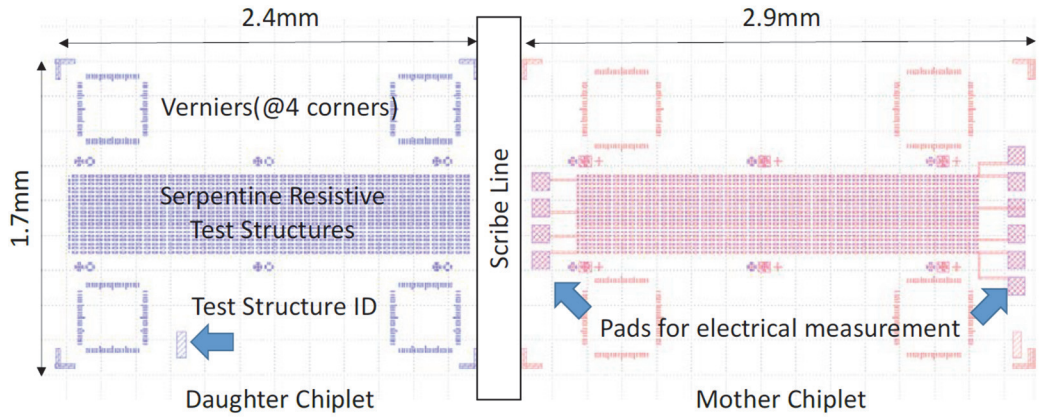


Fig.3 Test structure includes a serpentine resistive test structure for bump resistance measurement and Verniers for optical observation of alignment. Two types (mother and daughter) are made on the same wafer. Daughter chiplet of 2.4 mm-wide is supposed to be bonded over 2.9 mm-wide mother chiplet. One horizontal track contains 100 bumps. In total there are 2000 bumps per test chiplet.

was patterned with posi-resist (AZ P4620, 5 µm) and subsequent etching of SiO_2 / Cr / Au / Cr completed the process.

C. Obsevation of Bump Shape

Figures 6 to 8 show SEM images of fabricated test structure with bumps. As intended, TD1 5×5 μm^2 open square gave a (relatively) flat top (Fig.6). Test structures TD2 and TD3 did not perfectly gave intended shape; however we could indistinctly observe that 3 µm square hole (TD3-N) gave a center-convex (a molar-like) shape, which is our intention (Fig.7).

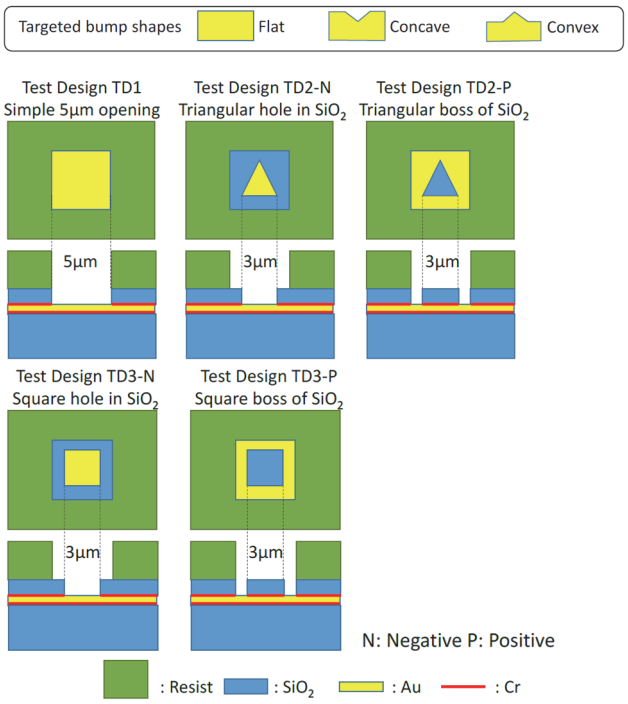


Fig.4 Five test designs were introduced in the test structure to obtain three target flat, convex and concave.

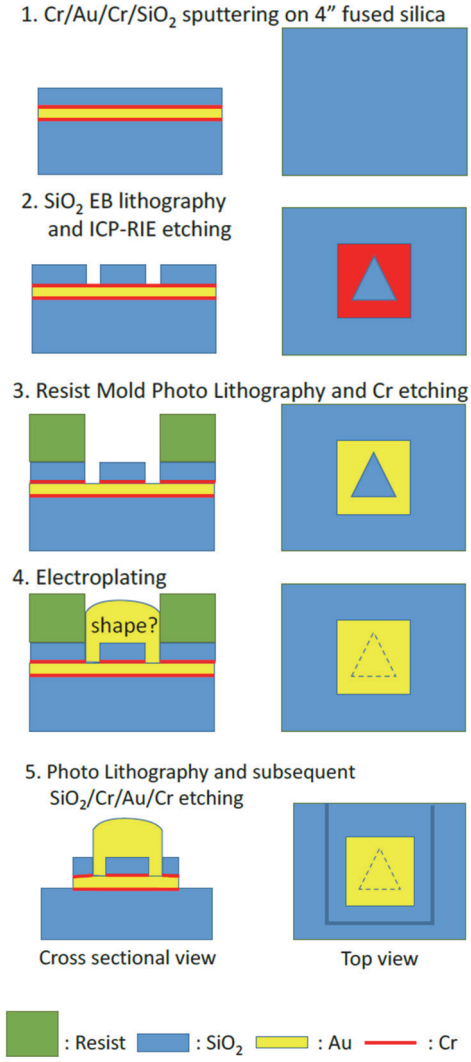


Fig.5 Simplified test structure fabrication process.

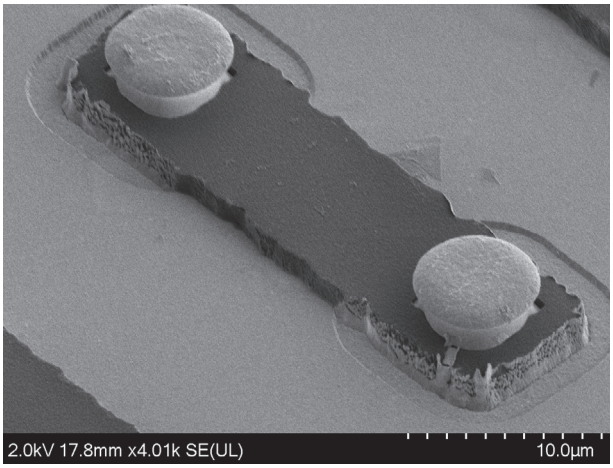


Fig.6 Test Design 1 (TD1) gave a flat surface as intended. Note that SEM photos were taken before seed layer removal.

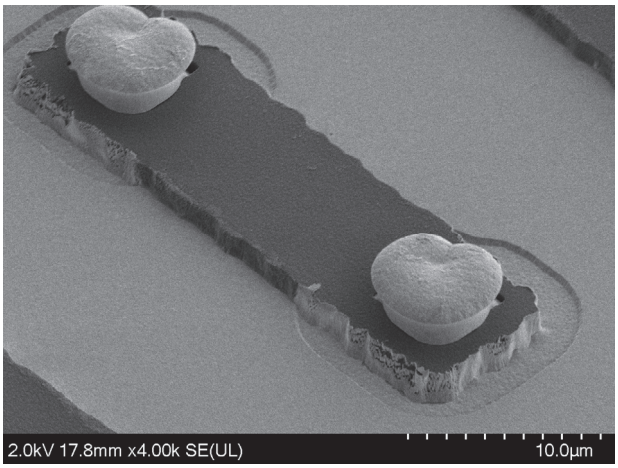


Fig.7 TD3-N effectively gave an intended concave (molar) structure but the surface was more flat than as we have expected.

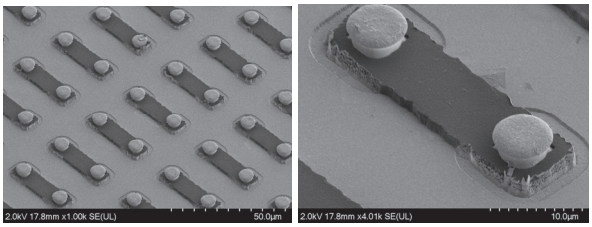
III. Bonding Test

A. *Chips Bonding condition*

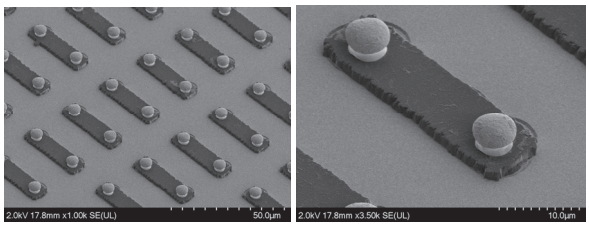
The chips thus obtained were bonded using a flip chip bonder (finetech, FINEPLACER Lambda2). For the test structure TD2 and TD3, P (hole) and N (boss) were bonded. Figure 9 shows the bonding profile, the applied force and temperature were 50N and 250°C, respectively.

B. *Results and Discussion*

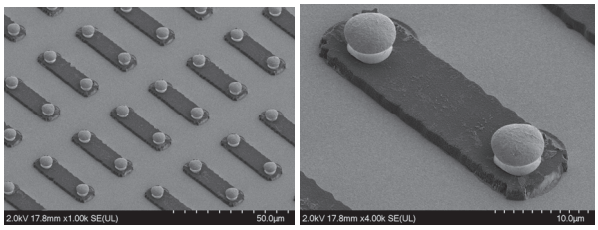
Because these test chips were fabricated on quartz substrate, they were easy to optically observe registration errors. In all TD1, 2, and 3, posisioning shift were observed during bonding process. TD1 and TD2 gave almost the same results: Bonding position was shifted almost same distance as the size of the bumps. TD3 was also shifted, but the shift was slightly smaller. This can be



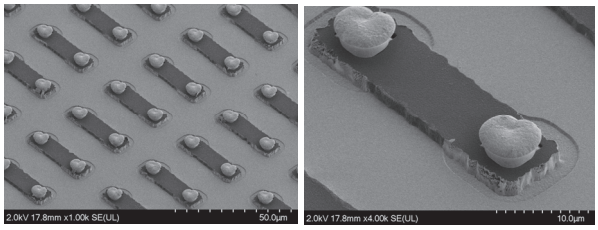
a) TD1 (5μm-Square)



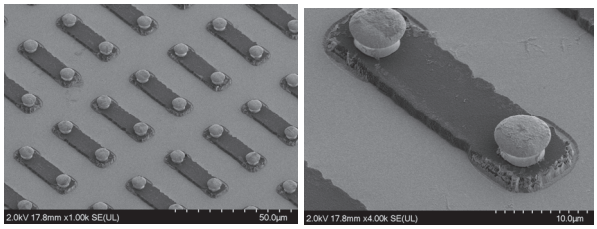
b) TD2-N (3μm-Trianglar Hole)



c) TD2-P (3μm-Trianglar Boss)



d) TD3-P (3μm-Square Hole)



e) TD3-P (3μm-Square Boss)

Fig.8 Through SEM observation, our present conclusion is it is easier to make convex bump structure, and the shape is more pronouncing with small square openings (TD3).

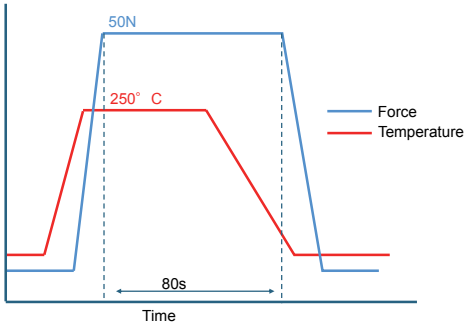
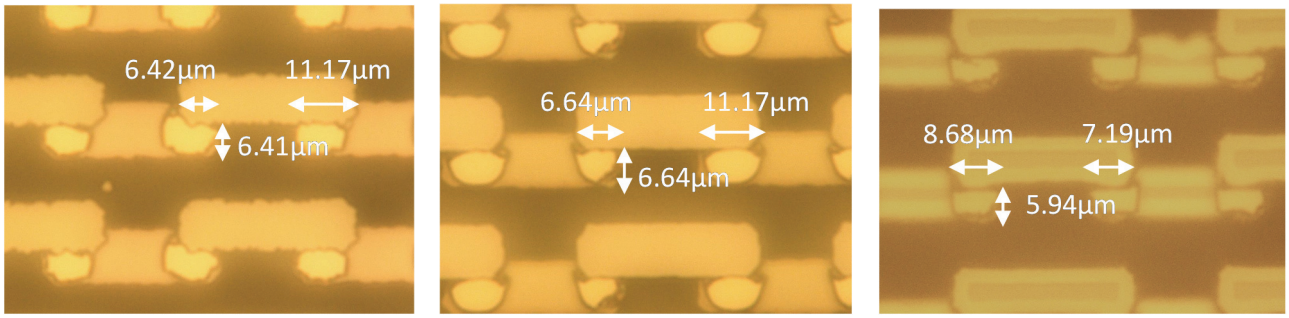


Fig.9 Bonding temperature and force profiles



a) TD1 b) TD2 c) TD3

Fig.10 By Optical observation after Flip Chip Bonding, the chips shifted about 6 μ m in vertical direction and horizontal overlap is not symmetrical on both sides.

understood that the concave shape bump works for suppressing the shift during bonding process.

Then we performed another flip-chip bonding operation using chips without bumps to identify where the position shift occurs or during the optical alignment or during the bonding (applying force). As the result, we found that bonding between chips without bumps caused almost no position shift. Therefore, the position shift can be attributed to the slip of bumps during bonding. The measurement result revealed that our fabricated bumps acted as both "convex"-"convex" heads, and resulted a slip of the total bump size distance, thereby suggesting us that convex-concave pairs dimension must be much larger (ideally in the same order of the bump size) than we have realized (which was less than 10% of the bump).

IV. Conclusions

We have demonstrated to fabricate and evaluate chips with Au bump test structure. Although we have obtained a clue to control surface morphology, some improvements are found to be necessary, especially to obtain convex (a canine-like) bumps. Further proces improvement may achieve the goal. Also, for bonding, the possibility was obtained that concave shape bump could suppress misalignment.

Acknowledgment

This work was jointly supported by ARIM project JPMXP1223UT1212, and "UTokyo Agile-X" X-Nics project JPJ011438, JST SICORP JPMJSC22H1, JST CREST JPMJCR20T2, all of the Ministry of Education, Culture, Sports, Science and Technology (MEXT). Fabrication and measurement were conducted in Takeda Cleanroom with help of Nanofabrication Platform Center of School of Engineering, the University of Tokyo, Japan.

References

[1] Yusuke Ebihara, Ayako Mizushima, Takashi Yoda, Kenji Hirakawa, Masayuki Iwase, Munehiro Ogasawara, Akio Higo, Yukinori Ochiai, and Yoshio Mita, "Two-pads per electrode in-situ test structure for micron-scale flip-chip bonding reliability of chip-on-chip device", *2022 IEEE 34th International Conference on Microelectronic Test Structures (ICMTS)*, online DOI: 10.1109/icmts50340.2022.9898248

[2] Jürgen M. Wolf, Gunter Engelmann, Lothar Dietrich, and Herbert Reichl, "Flip chip bumping technology—Status and update.", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 565.1 (2006): 290-295.